frontiers | Research Topics

Artificial intelligence solutions for global health and disaster response: Challenges and opportunities

Edited by
Tetyana Chumachenko, Plinio Pelegrini Morita,
Dmytro Chumachenko and Saman Ghaffarian

Published in
Frontiers in Public Health

ISSN 1664-8714
ISBN 978-2-8325-5393-0
DOI 10.3389/978-2-8325-5393-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Artificial intelligence solutions for global health and disaster response: Challenges and opportunities

Topic editors
Tetyana Chumachenko — Kharkiv National Medical University, Ukraine
Plinio Pelegrini Morita — University of Waterloo, Canada
Dmytro Chumachenko — National Aerospace University – Kharkiv Aviation Institute, Ukraine
Saman Ghaffarian — University College London, United Kingdom

Citation
Chumachenko, T., Morita, P. P., Chumachenko, D., Ghaffarian, S., eds. (2024). *Artificial intelligence solutions for global health and disaster response: Challenges and opportunities*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5393-0

Table of contents

frontiers | Frontiers in Public Health

TYPE Editorial
PUBLISHED 22 August 2024
DOI 10.3389/fpubh.2024.1439914

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Andreea Molnar,
Swinburne University of Technology, Australia

*CORRESPONDENCE
Dmytro Chumachenko
✉ dichumachenko@gmail.com

RECEIVED 28 May 2024
ACCEPTED 29 July 2024
PUBLISHED 22 August 2024

CITATION
Chumachenko D, Morita PP, Ghaffarian S and Chumachenko T (2024) Editorial: Artificial intelligence solutions for global health and disaster response: challenges and opportunities.
Front. Public Health 12:1439914.
doi: 10.3389/fpubh.2024.1439914

Editorial: Artificial intelligence solutions for global health and disaster response: challenges and opportunities

Dmytro Chumachenko[1]*, Plinio Pelegrini Morita[2], Saman Ghaffarian[3] and Tetyana Chumachenko[4]

[1]Mathematical Modelling and Artificial Intelligence Department, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine, [2]Ubiquitous Health Technologies Lab, University of Waterloo, Waterloo, ON, Canada, [3]Institute for Risk and Disaster Reduction, University College London, London, United Kingdom, [4]Epidemiology Department, Kharkiv National Medical University, Kharkiv, Ukraine

KEYWORDS

artificial intelligence, public health, global health, disaster response, risk assessment

Editorial on the Research Topic
Artificial intelligence solutions for global health and disaster response: challenges and opportunities

Artificial intelligence (AI) has emerged as a transformative tool in disaster preparedness and response, promising to reshape how we address some of the most pressing global health challenges. The rapid development of AI technologies offers unprecedented opportunities to enhance early warning systems, optimize resource allocation, and facilitate real-time decision-making processes. This editorial introduces a Research Topic of articles that explore the cutting-edge applications of AI across various domains of global health and disaster response, aiming to provide a comprehensive overview of both the potential and the hurdles associated with these technologies.

This Research Topic gathers insights from leading researchers, practitioners, and policymakers at the forefront of integrating AI into health crisis management. The scope of discussion spans multiple critical areas, including AI-powered surveillance systems, AI-supported decision-making in resource distribution, and the role of AI in managing mental health during crises. Each contribution advances our understanding of AI's capabilities and critically examines the ethical, legal, and social implications of deploying AI in sensitive and high-stakes environments. Through this Research Topic, we aim to highlight the innovative ways in which AI can promote health equity and improve healthcare delivery during disasters while also addressing the significant challenges that must be overcome to harness the full potential of AI in enhancing global health outcomes.

The Research Topic includes original research, systematic reviews, brief research reports, mini reviews, perspective, review, and methods papers. We received 66 submissions, 25 of which, after a careful review process, were accepted for publication in this Research Topic.

The paper by Liu P. et al. explores healthcare professionals' and patients' perceptions and experiences regarding mobile health (mHealth) apps in China. The study identifies the advantages and challenges of using mHealth apps using semistructured interviews. It highlights concerns such as the ease of use, effectiveness, and potential risks associated with these apps. The findings underscore the need for improved safety, regulatory standards, and user-friendly designs to serve older populations better and bridge the gap between healthcare providers and patients. The paper advocates for advancements in mHealth technology to enhance its utility and reliability in improving health outcomes.

The paper by Singh et al. delves into the burgeoning role of blockchain technology within the healthcare sector. It thoroughly examines how blockchain's core features—such as decentralization, immutability, and transparency—can address persistent challenges in healthcare, such as data fragmentation, security concerns, and the lack of interoperability among systems. By proposing a framework for the practical application of blockchain in various healthcare processes, the authors argue that blockchain technology can significantly enhance data security, patient privacy, and the efficiency of healthcare delivery systems. Moreover, the paper discusses the broader implications of blockchain integration into healthcare, including potential ethical and regulatory challenges.

Belachew et al.'s study investigates the adoption and utilization of telemedicine among patients with chronic diseases at the University of Gondar Comprehensive Specialized Hospital. The research uses a cross-sectional study design to capture the perceptions, willingness, and actual use of telemedicine services among these patients. Findings indicate a generally positive perception and willingness to engage with telemedicine despite low usage. This discrepancy highlights the potential for greater implementation of telemedicine, provided that patient awareness and infrastructural support are enhanced.

The paper by Shakhovska et al. investigates micro-stresses' impact on operators' performance within search systems. The study assesses operator behavior under stress by presenting test images to monitor reaction times and decision-making processes utilizing a human-machine interface model. Key findings include the development of a methodology to analyze and cluster operator responses based on their stress resistance. The study offers insights into optimizing the selection and training of operators by recognizing individual differences in stress handling.

The paper by Li et al. introduces a fractal multi-level distribution network (FMDN) model to optimize the delivery of essential supplies during disasters. This model considers factors like road damage and dynamic demand changes at disaster sites to minimize overall operational costs, which include construction, transportation, and penalty costs for unmet demand. The study confirms the model's effectiveness through numerical experiments using LINGO software, demonstrating that it can efficiently manage emergency resource distribution under variable and challenging conditions.

The paper by Pinto et al. examines the influence of online news on public health responses by focusing on the syphilis epidemic in Brazil. The study analyzes the volume and quality of news between 2015 and 2019 and its correlation with syphilis testing rates using text mining techniques. The findings reveal a positive association between high-quality news and increased syphilis testing, emphasizing the role of effective communication in enhancing public health policy actions.

The paper by Zhang et al. examines the influences on Chinese type 2 diabetes patients' readiness to adopt digital disease management apps. Utilizing the Technology Acceptance Model (TAM), Perceived Risk (PR) theory, and eHealth Literacy Theory (E-HLT), the study highlights significant predictors such as perceived usefulness, ease of use, and electronic health literacy. Results indicate these factors substantially impact patients' attitudes toward and intentions to use such applications, providing critical insights for developers and healthcare providers aiming to increase digital health tool adoption.

The paper by Sen et al. addresses the growing relevance of big data analytics in medical research, focusing specifically on text-mining techniques. It discusses various challenges encountered in this field, such as data quality and integration, privacy concerns, and the need for robust analytical tools. The authors propose several recommendations to overcome these hurdles, emphasizing the development of more sophisticated data processing algorithms and implementing stricter data privacy regulations. The paper aims to guide future research and development in efficiently utilizing medical big data for improved healthcare outcomes.

The paper by Towler et al. evaluates the efficacy of Machine-Assisted Topic Analysis (MATA) against traditional human-only thematic analysis. The study, conducted during the COVID-19 pandemic, used a dataset of 1,472 user responses from a behavioral intervention to compare the two methods. The results show that while human analysis provides depth, MATA significantly reduces the time required for data analysis. Both methods produced similar thematic outcomes, demonstrating that MATA is an effective tool for rapid qualitative analysis in public health emergencies.

The paper by Honchar et al. investigates the predictors of long-term health effects in COVID-19 patients post-discharge. The study assesses symptoms, performs echocardiography, and conducts a 6-min walk test both pre-discharge and 1-month post-discharge, utilizing a cohort of 221 hospitalized patients. The presence of post-COVID-19 syndrome (PCS) was then evaluated 3 months after discharge. The research identifies key pre-discharge predictors for PCS, including age, sex, inflammation levels, and oxygen needs. The study successfully developed a neural network-based classification model that predicts the development of PCS with high accuracy, demonstrating a significant step toward optimizing patient care post-COVID-19.

The study by Lihua et al. evaluates the effectiveness of a health self-management intervention tailored for individuals with metabolic syndrome (MS) who were bereaved by the Wenchuan earthquake. This randomized controlled trial utilized a detailed intervention program based on self-management principles, covering diet, exercise, medication adherence, and emotional management. Results showed significant improvements in MS management behaviors and some physiological indicators among the intervention group compared to controls, demonstrating the potential of structured self-management programs in enhancing disease outcomes in post-disaster settings.

The paper by Zhou et al. comprehensively evaluates various AI-based intraocular lens (IOL) power calculation formulas compared to traditional and newer vergence formulas. The study systematically analyzed 12 studies involving 2,430 eyes, utilizing methods such as mean absolute error (MAE), median absolute error (MedAE), and the percentage of eyes with a predictive error within specific diopters focused on highly myopic eyes. The top-performing AI-based formulas identified were XGBoost, Hill-RBF, and Kane, demonstrating superior accuracy over traditional methods for calculating IOL power for highly myopic patients. This paper highlights the potential of AI to enhance the precision of medical predictions in specialized settings.

The study by Ahun et al. investigates Turkish emergency physicians' ethical perspectives on using AI in epidemic triage. The study assesses the attitudes of 167 specialists toward AI's utility in emergency triage, particularly under pandemic conditions, using a detailed survey. The findings reveal a cautious optimism about AI's benefits for patient care and healthcare operations. However, significant concerns remain regarding responsibility for AI-driven decisions and the ethical handling of private patient data. Most respondents acknowledge AI's potential to enhance triage efficiency but emphasize the need for clear accountability and data privacy guidelines.

The paper by Long et al. explores the application of AI to optimize healthcare supply chain modes. It introduces a deep reinforcement learning algorithm to improve decision-making in selecting sustainable and efficient supply chain modes. Through simulation experiments, the study demonstrates that AI can effectively enhance the economic, social, and environmental benefits of healthcare supply chains. The findings suggest AI is a superior healthcare supply chain mode selection method, offering significant improvements over traditional methods.

The paper by Kryvenko et al. addresses the increasing need for effective compression techniques in storing and transferring large dental images. The authors explore lossy compression that retains the visual quality necessary for diagnostic purposes, emphasizing the role of discrete cosine transform-based encoders. They evaluate the performance of different encoders in achieving high compression ratios while ensuring the invisibility of compression artifacts through theoretical approaches and experimental validation with professional dentists. This study is crucial for advancing digital imaging in dental practice and improving telemedicine applications by enabling efficient image data management.

The paper by Zaidan provides a comprehensive analysis of how AI is being integrated into various aspects of global health, addressing significant challenges such as mental health, infectious diseases, cardiovascular diseases, and the impacts of aging. The review explores AI's role in enhancing disease surveillance, diagnosis, treatment modalities, and overall public health strategies. Additionally, it discusses the ethical considerations and the necessity for equitable AI integration in healthcare practices globally, highlighting AI's transformative potential while acknowledging the complexities and responsibilities accompanying its widespread adoption.

The paper by Valeanu et al. examines vaccine misinformation on Twitter, focusing on content written in Romanian. Using a dataset of 1,400 tweets, researchers manually classified each tweet as true, neutral, or fake information. They employed machine learning algorithms to predict the classification of tweets, finding that misinformation tends to be more frequently liked and shared. This study highlights the significant role of AI in distinguishing between valid and false health information online, aiming to support public health efforts by mitigating the spread of vaccine misinformation.

The paper by Yoon et al. examines the application of virtual reality (VR) and virtual try-on technologies to address privacy concerns in home-based fitness training. It introduces methods to anonymize participants' appearances and environments during video-based exercise sessions, enhancing user comfort and motivation. The study demonstrates the effectiveness of these technologies through a user study, highlighting their potential to enhance privacy, self-confidence, and coaching satisfaction. However, no significant differences in coaching satisfaction were noted. This innovative approach proposes a model for future remote fitness training that prioritizes user privacy and engagement.

The study by Roche et al. investigates the accuracy of AI in reading and interpreting HIV self-test results. The study, conducted in Kisumu, Kenya, involved participants using blood-based HIV self-tests in private pharmacies, with results interpreted by clients, pharmacy providers, and an AI algorithm. The AI algorithm demonstrated high sensitivity and specificity, showing promise as a quality assurance tool in HIV testing, comparing these interpretations to an expert panel's readings. The findings suggest that AI could enhance accuracy and reliability in interpreting HIV self-tests in real-world settings.

The paper by Tucker and Lorig explores the integration of agent-based social simulations (ABSS) with the everyday digital health perspective to inform policy during health crises like COVID-19. ABSS models the interactions of intelligent agents to simulate complex societal responses to health policies, serving as a virtual testbed for scenario testing without risking real-world consequences. The paper highlights challenges such as the need for valuable scenario definition and appropriate data availability, suggesting that incorporating everyday digital health insights can enhance ABSS's effectiveness in crisis management. This approach emphasizes the importance of understanding digital health technology adoption and its varied impacts across different population segments to improve the accuracy and relevance of health crisis simulations.

The paper by Benboujja et al. addresses the significant challenge of language barriers in pediatric healthcare. It presents a multilingual, AI-assisted curriculum designed to improve global healthcare education. The curriculum includes video modules in multiple languages tailored for diverse healthcare settings using generative language models, enhancing comprehension and accessibility for non-English speaking healthcare providers and caregivers. This innovative approach ensures that essential pediatric care knowledge is universally accessible, supporting the World Health Organization's advocacy for digitally enabled healthcare education.

The paper by Alotaibi et al. examines the application of AI, specifically ChatGPT, in the management of ocular cancer. It reviews existing literature to address the types of ocular cancer, the associated challenges, and how ChatGPT can assist in overcoming these barriers. The review underscores the limited awareness and healthcare access, financial constraints, and infrastructure deficiencies as significant hurdles in effective ocular cancer management. The paper discusses the prospective benefits of integrating ChatGPT to enhance diagnostic accuracy, patient education, and treatment planning while highlighting the necessity for further research to optimize AI applications in healthcare.

The paper by Moskalenko and Kharchenko discusses enhancing the robustness of medical AI systems against disturbances like adversarial attacks, fault injections, and data drift. It introduces a modified machine learning operations (MLOps) framework incorporating resilience optimization, predictive uncertainty calibration, and graceful degradation. The study demonstrates that adding these resilience mechanisms improves the system's ability to handle disruptions, particularly in medical image recognition, using datasets like DermaMNIST, BloodMNIST, and PathMNIST. The findings suggest that AI systems can be more reliable and trustworthy in critical healthcare applications by integrating these resilience-focused adaptations.

The paper by Miao et al. introduces a machine learning model for the early identification of elevated arterial stiffness (EAS) using easily accessible clinical and questionnaire data from 77,134 participants. The study demonstrates the effectiveness of the XGBoost algorithm, which outperformed other models in accuracy and predictive capability, utilizing advanced feature selection and model training methods. The research emphasizes the potential of this cost-effective model to significantly enhance the screening processes for arterial aging, making it accessible for broad clinical application.

The paper by Liu X.-d. et al. focuses on enhancing the predictive accuracy of epidemic time series data by incorporating Gated Recurrent Units (GRU) into Graph Neural Networks (GNN). This integration, referred to as GRGNN, is evaluated using datasets including COVID-19 cases from African and European countries and a chickenpox dataset from Hungarian regions. The study demonstrates that GRGNN consistently outperforms traditional models in prediction accuracy, validating its effectiveness in epidemic forecasting and highlighting its potential for early warning systems in public health.

The Research Topic of articles within this Research Topic underscores AI's vast potential and multifaceted challenges in enhancing disaster preparedness and response. From improving early warning systems to ensuring health equity in crises, the insights provided by these papers highlight both the transformative capabilities and the ethical considerations necessary for the responsible deployment of AI technologies. As we move forward, the continued collaboration between researchers, practitioners, and policymakers will be crucial in harnessing AI's power to respond to, anticipate, and mitigate global health emergencies.

Author contributions

DC: Writing – original draft, Writing – review & editing. PM: Writing – original draft, Writing – review & editing. SG: Writing – original draft, Writing – review & editing. TC: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We are grateful to all the authors and reviewers contributing to this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Frontiers in Public Health
TYPE Original Research
PUBLISHED 07 August 2023
DOI 10.3389/fpubh.2023.1234436

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Abdullah Alkattan,
Ministry of Health, Saudi Arabia
Bokolo Anthony Jr.,
Institute for Energy Technology, Norway

*CORRESPONDENCE
Ashenafi Kibret Sendekie
ashukib02@yahoo.com;
ashenafi.kibret@uog.edu.et

RECEIVED 04 June 2023
ACCEPTED 24 July 2023
PUBLISHED 07 August 2023

CITATION
Belachew EA, Getachew D, Netere AK, Gizachew E and Sendekie AK (2023) Perception, willingness, and practices of telemedicine in patients with chronic diseases: implication of digital health in patients' perspective at a tertiary care hospital in Ethiopia. *Front. Public Health* 11:1234436. doi: 10.3389/fpubh.2023.1234436

Perception, willingness, and practices of telemedicine in patients with chronic diseases: implication of digital health in patients' perspective at a tertiary care hospital in Ethiopia

Eyayaw Ashete Belachew [1], Demis Getachew[2], Adeladlew Kassie Netere[1], Eshetie Gizachew[3] and Ashenafi Kibret Sendekie [1*]

[1]Department of Clinical Pharmacy, School of Pharmacy, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, [2]Department of Pharmacology, School of Pharmacy, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, [3]Department of Information System, College of Informatics, University of Gondar, Gondar, Ethiopia

Background: Technology-based healthcare services have important implications for the diagnosis, prevention, and treatment of diseases, as well as providing access to high-quality care that both the patient and the healthcare practitioner can benefit from. To access medical information, patients have also searched for methods of technology-based healthcare services like telemedicine (TM). However, little is known regarding the perceptions, willingness, and practices of TM among Ethiopian patients, especially in the study setting.

Objective: This study assessed the perceptions, willingness, and practice of TM among patients with chronic disease at the University of Gondar Comprehensive Specialized Hospital (UoGCSH), Northwest Ethiopia.

Methods: A cross-sectional study was conducted from June 1 to July 30, 2022, among patients with chronic diseases who were on follow-up at the UoGCSH. Eligible participants were included in the study using a systematic random sampling technique. A structured questionnaire was used and recorded in the Kobo data collection tool. The collected data were managed and analyzed using the Statistical Package for Social Science (SPSS) version 26.

Results: Out of 422 patients approached, 384 (91% response rate) were included in the final analysis. The mean (±SD) age of the participants was 48.07 ± 16.17 years. The overall perceptions mean (±SD) score of the respondents was 3.92 ± 1.06. Generally, near to three-fourths (71.1%) of the participants had a positive perception of TM services, and around two-thirds (63.3%) had a willingness to be involved in the TM service. However, only around one-fourth (24.5%) of the participants were perceived to have a high level of TM practice currently.

Conclusion: The findings suggest that although the level of perception and willingness of TM services among patients with chronic diseases was positive, their level of practice was low. Therefore, creating awareness and suitable conditions to improve their utilization of TM could be important.

KEYWORDS
attitude, chronic diseases, perception, practice, telemedicine, willingness

Introduction

Digital-based intervention has been popular in the self-management of patients with chronic diseases, particularly during the COVID-19 pandemic, when patients faced difficulties accessing face-to-face health services with healthcare providers (1–4). Telemedicine (TM) has been used for communication between physicians, nurses, paramedical personnel, and patients in need of medical care (5, 6). The origin of TM, that is, the support of health care services through communication and technology, can be traced back to the time during which electronic devices were invented (7). TM has been used to describe the direct provision of healthcare services, and its applications were demonstrated as teleconsultation, remote psychotherapy, remote imaging, telepathology, tele dermatology, patient care at the first level of health care, specialized counseling systems, remote surgery, remote anesthesiology, remote cardiology, remote radiology, remote oncology, home monitoring, patient education, continuing medical education, and remote home medical care (8, 9). Globally, different studies were conducted to show the adoption, use, and utility of TM during COVID-19 (10–13).

Chronic diseases have a significant impact on people's lives in the following ways: impaired mobility, reduced quality of life, negative emotions, increased economic burden, and a higher mortality rate. Despite this devastating potential, the incidence of chronic diseases is increasing (14, 15). According to the World Health Organization's 2015 report, chronic diseases killed 38 million people each year and have become a public health concern (15). Deaths account for roughly 80% of all deaths in low- and middle-income countries. Morbidity and mortality from chronic diseases are increasing faster in sub-Saharan Africa than at any other time in history. Ethiopia is no exception to the burden imposed by such chronic diseases (16). The prevalence of chronic non-communicable disease (NCD) in Ethiopia has ranged from 29 to 35%: diabetes at 5%, cardiovascular disorders at 13.4%, and respiratory conditions at 1–18% (1–18%) (17). Aside from its high prevalence, chronic disease management is difficult because it necessitates long-term follow-up by health care providers and patients, as well as a large amount of resources. This results in a scarcity of health-care resources available to patients and their families and serves as a driving force for telehealth innovation (18). TM enables remote monitoring of vital signs in patients with chronic conditions, and by doing so, it decreases mortality and hospitalization while increasing quality of life (4, 19–21). It eliminates the need to go to the doctor's office and sit side by side with others, which could lead to the transmission of communicable diseases (22). It also reduces rates of exacerbation and hospitalization in patients with asthma and chronic obstructive pulmonary disease (COPD) by allowing patients to receive better medical attention at the convenience of both the doctor and themselves (22). A review also disclosed that the use of digital-based messages helps in improving hypertension, diabetes and asthma (23).

Abbreviations: COPD, Chronic Obstructive Pulmonary Diseases; CI, Confidence Interval; DM, Diabetes Meletus; HTN, Hypertension; TM, Telemedicine; UoGCSH, University of Gondar Comprehensive Specialized Hospital; NGO, Non-Governmental organization.

TABLE 1 Sociodemographic characteristics of chronic disease patients toward telemedicine at ambulatory clinic of UOGCSH, Ethiopia 2022 (*N* = 384).

Variables	Category	Frequency (%)
Sex	Male	205 (53.4)
	Female	179 (46.6)
Age of participants	Mean (±SD)	48.07 (16.17)
	<31 years	72 (18.8)
	32–50 years	149 (38.8)
	50–60 years	73 (19)
	≥61 years	90 (23.4)
Residence	Urban	232 (60.4)
	Rural	152 (39.6)
Marital status	Single	67 (17.4)
	Married	274 (71.4)
	Divorced	22 (5.7)
	Widow	19 (4.9)
Educational status	No formal education	109 (28.4)
	Primary (1–8)	93 (24.2)
	Secondary (9–12)	104 (27.1)
	College and above	78 (20.)
Occupation	Government employee	60 (15.6)
	Business/self-employee	237 (61.7)
	Unemployed	61 (15.9)
	Freelance	26 (6.8)
Monthly income	≤860	28 (7.3)
	861–1,500	37 (9.6)
	1,501–3,000	195 (50.8)
	3,001–4,999	33 (8.6)
	≥5,000	91 (23.7)
Health cost coverage	Out pocket	101 (26.3)
	Health insurance	207 (70.3)
	NGO	7 (1.8)
	Other	6 (1.6)
Internet and /or phone cell use	Yes, on my Owen	214 (55.7)
	On the help of other	18 (4.7)
	No	151 (39.3)
Smart phone use	Yes	132 (34.4)
	My family use it	13 (3.4)
	No	239 (62.2)
Current living situation	Living alone	26 (6.8)
	Living with family	351 (91.4)
	Other	7 (1.8)

(Continued)

TABLE 1 (Continued)

Variables	Category	Frequency (%)
Hospital traveling service	Walk	9 (2.3)
	Private car	4 (1)
	Taxi/public transportation	371 (96.6)

TABLE 2 Clinical characteristics of chronic patients at the ambulatory clinic of UOGCSH, Ethiopia 2022 (*N* = 384).

Variable	Chronic disease type	Frequency (%)
Type chronic disease	Cardiovascular disorders	84 (21.9)
	Endocrine disorders	69 (18.0)
	Respiratory disorders	65 (16.9)
	Nerve system disorders	54 (14.1)
	Reproductive disorders	44 (11.5)
	Circulatory disorders	84 (21.9)
	Musculoskeletal disorders	30 (7.8)
	Integumentary disorder	18 (4.7)
	Urinary disorders	17 (4.4)
	Digestive disorders	3 (0.8)
Duration of disease	<5 years	202 (52.6)
	≥5 years	182 (47.4)
Fallen on the way to our hospital	Yes	61 (15.9)
	No	323 (84.1)
Admission history to the hospital with in 6 months	Yes	219 (57)
	No	165 (43)
Waiting time in the hospital is long	Vey often	143 (37.2)
	Yes	135 (35.2)
	Not so much	53 (13.8)
	No	53 (18.8)
Follow-up time	Monthly	120 (31.3)
	Once every 2 months	169 (44)
	Once every 3 months	79 (20.6)
	Semi-annually	12 (3.1)
	Annually	4 (1)

Although interest in telehealth is growing, its implementation is not common practice yet. Initiatives for e-health applications are not similar and consistent throughout the world. For example, in the Netherlands, the implementation of e-health is developed through the collaboration of caregivers, caretakers, and health insurance (14). People living in low-income countries, such as African countries, face more health problems than those in developed countries, and they have limited access to health innovations (24). When it comes to Ethiopia, which is the second-most populous country in Africa, with over 117 million people in 2021, in order to meet the healthcare needs of such a large population, the implementation of TM is recommended to expand (25). Ethiopia's Ministry of Health (MoH) recently launched a digital and innovation center where experts can synthesize, promote best practices, and scale up innovation tools. According to the April 30th, 2021 report, ∼54.7 million people were telecommunications subscribers. There were ∼54.7 million mobile voice subscribers, 25 million internet subscribers, 349,000 fixed broadband subscribers, and 923 fixed service subscribers. The telecom coverage was reported to be 85.4–95%, and the density was 50% (25).

Knowing the global and national implications of chronic disease management, developed countries advocated for a new technology, TM that supports the health system. It aids in lowering health-care costs, increasing access to care, reducing the need for a larger workforce, and reducing time waste. Although its implementation is in its early stages, Ethiopia needs to implement TM after learning about its benefits. However, to the best of authors knowledge and broader search, assessment of patients with chronic diseases perceptions, willingness, and practice of TM has not been explored yet. Even though various literatures are conducted during COVID-19 world concerning to the use of TM and its implication, there is scarcity of information in the study area.

Therefore, the main research questions are: (I) Do patients with chronic diseases perceive TM as good or bad? (II) Are patients with chronic diseases willing to use TM or not? (III) Do patients practice TM or not? To answer these research questions, this study assessed perceptions, willingness, and practice of TM among patients with chronic diseases at the UoGCSH, Northwest Ethiopia. The findings will serve as a baseline for identifying opportunities and barriers from the patients' perspective to ensure the delivery of better healthcare through the implementation of TM.

Methods and materials

Study design and setting

An institutional-based cross-sectional study with patient interviews and medical record review was carried out at UOGCSH ambulatory chronic care from June 20 to July 20, 2022. The UoGCSH is a comprehensive public referral health facility in northern Ethiopia that serves as a teaching hospital for students from the University of Gondar. The hospital is ∼738 km from Ethiopia's capital city. The UoGCSH statistics and information office revealed that the hospital's ambulatory care follows up with patients every Monday to Friday and serves 15,000 patients with chronic conditions such as HTN, diabetes, asthma, COPD, heart failure, epilepsy, and other cardiovascular disorders as of the November 2019 report.

TABLE 3 Perception toward telemedicine in patients with chronic disease at the UoGCSH, Ethiopia 2022 ($N = 384$).

Statements	Response level	Frequency (%)
Anxious about using a tablet or mobile device	Very much	81 (21.1)
	Yes	114 (29.7)
	Not so much	30 (7.8)
	No	159 (41.4)
Opinion about the usefulness of the eHealth application for making decision for own health	Not useful at all	37 (9.6)
	Not useful	61 (15.9)
	Unsure	69 (18)
	Useful	189 (49.2)
	Very useful	28 (7.3)

Perception statements	Level of response, *n* (%)					
	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean (SD)
Telemedical service may be necessary for chronic patient care	52 (13.5)	229 (59.6)	64 (16.7)	19 (4.9)	20 (5.2)	3.72 (0.94)
Providing a telemedicine service helps faster medical care	62 (16.1)	179 (46.6)	68 (17.2)	19 (4.9)	56 (14.6)	3.74 (1.20)
Providing telemedicine is important for medical care to remote and underserved areas of healthcare.	34 (8.9)	180 (46.9)	104 (27.1)	33 (8.6)	33 (8.6)	3.96 (0.95)
Providing a telemedicine service saves effort.	38 (9.9)	175 (45.6)	105 (27.3)	33 (8.6)	33 (8.6)	3.95 (0.97)
Providing a telemedicine service saves money.	33 (8)	200 (52.1)	85 (22.1)	33 (8.6)	33 (8.6)	3.96 (0.95)
Providing a telemedicine service saves transportation cost.	36 (9.4)	231 (60.1)	45 (11.7)	39 (9.6)	44 (9.8)	3.95 (0.96)
Providing a telemedicine service reduces waiting lists in medical centers.	43 (11.2)	225 (58.6)	52 (13.5)	31 (8.3)	32 (8.5)	3.9 (1.00)
Providing a telemedicine service can improve communication between patients and their doctor or nurse.	62 (16.	179 (46.6)	68 (17.7)	19 (4.9)	56 (14.6)	3.84 (1.120)
Providing a telemedicine service can help in providing appropriate instructions in emergencies.	36 (9.4)	200 (52.1)	54 (14.1)	19 (4.9)	75 (19.5)	4.15 (1.13)
Using eHealth applications to access healthcare offers greater security.	43 (11.2)	201 (52.3)	46 (12)	19 (4.9)	75 (19)	4.15 (1.16)
Easy to share medical information on eHealth applications	39 (10.2)	210 (54.7)	40 (10.4)	20 (5.2)	75 (19.5)	4.15 (1.14)
Using eHealth applications to access healthcare is convenient.	40 (10.4)	195 (50.8)	54 (14.1)	20 (5.2)	75 (19.5)	3.15 (1.15)
Using eHealth application is good to consulting physician	29 (7.6)	200 (52.1)	60 (15.6)	20 (5.2)	75 (19.5)	4.18 (1.10)
eHealth application makes appointment online to see physician	36 (9.4)	216 (56.3)	36 (9.4)	21 (5.5)	75 (19.5)	4.16 (1.13)
Over all mean (± SD) of the respondents						3.92 (1.06)
Overall perception of TM service	Good	273 (71.1%)				
	Poor	111 (28.9%)				

10.3389/fpubh.2023.1234436

TABLE 4 Thoughts of participants about using an application to get, keep, and update health information ($N = 384$).

Statements	Level of response, *n* (%)				
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I would be worried about the privacy of my health information to use TM	40 (10.4)	35 (9.1)	26 (6.8)	96 (25)	187 (48.7)
I don't need TM to handle my health needs because it does not keep privacy	40 (10.6)	56 (14.6)	57 (14.8)	85 (22.1)	146 (38)
I don't like using computers, mobiles or the internet because it can lead to disclosing medical information to people who are not authorized to do so	96 (24.9)	59 (15.4)	57 (14.8)	80 (20.8)	92 (24)

Study participants

All patients with common chronic diseases who had been followed for their medical condition at the ambulatory chronic care unit of the UoGCSH during the study period were included as study participants. To be included, they could be adults (age higher and equal to 18), diagnosed and followed for at least one chronic disease, and agreed to participate in the study. Patients who were severely ill and unable to provide an interview, those who were with sever neurological and psychiatric problems, and patients with hearing problems were all excluded.

Sample size and sampling technique

The sample size was calculated using the single mean population proportion formula with a 95% confidence level ($Z = 1.96$), a proportion of the outcome (i.e., perceptions, willingness, and practice of TM) of 50% ($P = 0.5$) to obtain an adequate sample size and a relative precision of 5% ($D = 0.05$).

$$n = \frac{\left(\frac{Z\partial}{2}\right)^2 P(1-P)}{D^2} = \frac{(1.96)^2 *0.5*0.5}{(0.05)^2}$$

$n = 384.16 \approx 384$

Where; $Z\alpha/2$ = confidence level at 95% (standard value of 1.96); P = estimated prevalence or proportions of the outcome; d = range of CI or margin of error.

Finally, considering a 10% contingency for possible non-response, the calculated sample size was to be 422.

A systematic random sampling method was used to include study participants. The total number of chronic patients who visit this hospital per month is about 15,000; from this, 1,500 patients who fulfilled the inclusion criteria visited the hospital every month, and 300 patients who fulfilled the inclusion criteria were found to have an appointment during the data collection period. Thus, we had 1,800 visitors in a month (taking into consideration that the sample was collected within a month). As a result, the sampling fraction (k-interval) is 1,800/422 = 4 (approximately). As a result, the first participant in the study was chosen at random, and each subsequent participant was chosen every four persons. The study subjects who met the inclusion criteria were considered, and if one was deemed ineligible, the next one was chosen, and the same approach was used throughout the data collection procedure.

Data collection tools and techniques

Data on patients' perceptions, willingness, and practice of TM were collected using validated data collection instruments. The data were collected following the adoption of questionnaires from various studies (14, 22) and were prepared using the Kobo data collection tool. The data collection format is divided into four sections. Part I: focusing on the study population's socio-demographic characteristics (age and gender, marital status, educational level, residence, work status, internet or mobile phone usage status, religion, mode of transportation, and health service coverage) and clinical characteristics questionnaires (type of chronic disease, duration of disease, frequency of visit, perception of waiting for health service, and e-health usage were assessed using structured questionnaires); Part II: Perception and/or opinion toward TM (in this part, patients' past and present perceptions and opinions about its benefits in terms of cost reduction, time saving, quick access, and access to remote areas were assessed). This perception toward TM use was measured by 14 questions on a five-level Likert (5 = strongly agree, 4 =agree, 3 = neutral, 2 = disagree, 1 = strongly disagree) scale questionnaire, and the overall level of perception was determined and categorized by using the overall mean score of the questionnaires. Participants with scores above the mean have a positive perception of TM, while those with scores below the mean have a negative perception of TM. Part III was composed of patients present as well as their future willingness to use TM. The questionnaire consists of 16 questions: three questions with a score of 4 (0–3), one question with a score of 3 (1–3), and 12 yes or no questions (1 = yes and 0 = no). Therefore, the overall score of the questionnaire was 24 points. To rate the overall willingness of the participants to engage in TM, we have categorized those scores ≥ 13 as having good willingness and those scores below 18 as having poor willingness to engage in TM. Part IV: Focuses on their current practice in the TM (their rate or frequency of usage and the type of information they extract from an e-health service). To measure the practice of respondents, we developed 12 yes-or-no questions. The overall level of practice was determined by the score of the questions. Those participants who answered yes to questions six and above were considered to have good practice of TM.

TABLE 5 Willingness to involve and practice of Telemedicine among chronic patients at the ambulatory clinic of UOGCSH, Ethiopia 2022 (*N* = 384).

Willingness and practice statement	Category	Frequency (%)
Confident toward your own health management	Very confident	87 (22.7)
	Somewhat confident	153 (39.8)
	Not too confident	63 (16.4)
	Not at all confident	81 (21.1)
Are you familiar with telemedicine	Yes	133 (34.6)
	Yes, I have heard, but I don't know the details	150 (39.1)
	No	101 (26.3)
Interest track information about a chronic illness	Yes	160 (41.7)
	No	224 (58.3)
Interest track information to track your diet and calories	Yes	146 (38.0)
	No	238 (62)
Interest information track your exercise	Yes	150 (39.1)
	No	234 (50.9)
Interest information track you remind you when to take prescriptions	Yes	227 (59.1)
	No	157 (40.9)
Interest information track your remind you when you need tests	Yes	175 (2.6)
	No	374 (97.4)
Would you be interested in using this type of website if it were from your doctor	Yes	191 (49.7)
	No	193(50.3)
Would you be interested in using this type of website if it were from the hospital you use	Yes	171 (44.5)
	No	213 (55.5)
Would you be interested in using this type of website if it were from pharmacist	Yes	150 (39.1)
	No	334 (60.9)
Would you be interested in using this type of website if it were from your health insurance plan	Yes	216 (56.3)
	No	168 (43.7)
Would you be interested in using this type of website if it were from a government group like Medicare	Yes	203 (52.9)
	No	184 (47.1)
Would you be interested in using this type of website if it were from nurse	Yes	141 (36.7)
	No	243 (63.3)
Would you be interested in using this type of website if it were from a company like Google or Apple	Yes	138 (35.9)
	No	246 (64.1)
In general, if your health information were online, how worried would you about the privacy and confidentiality of your information?	Very worried	60 (15.6)
	Somewhat worried	142 (37)
	Not too worried	99 (25.8)
	Not at all worried	83 (21.6)
Currently, how much you are interest to use TM application	Very interested	43 (11.2)
	Interested	127 (33.1)
	A little bit interested	73 (19)
	Not at all interested	67 (36.7)
Overall perceived willingness to use TM:	Willing to involve	242 (63.3)
	Not willing to involve	67 (36.7)

Cronbach's alpha was employed to determine the internal consistency of the data collection tools. Perception toward TM ($\alpha = 0.78$), thoughts to update their information toward TM ($\alpha = 0.82$), willingness to use TM ($\alpha = 0.79$), practice of TM ($\alpha = 0.83$), and frequency of using TM ($\alpha = 0.88$) all indicated that the tools have an acceptable range of reliability.

Data quality control technique

Before collecting data, the literature on the questionnaires was reviewed. A pretest was conducted on 21 patients with chronic diseases (5% of the samples) having follow-ups at Debark general hospital. Then, after the pretest feedback, the actual data was collected. Following data collection, proper categorization and coding of the data were performed, and the collected data were reviewed for completeness and accuracy by checking the recorded data. After entering the data into SPSS, it was double-checked for accuracy.

Data entry, management, and analysis

Data was collected by the Kobo data collection tool and then downloaded to XLS and analyzed using Statistical Package

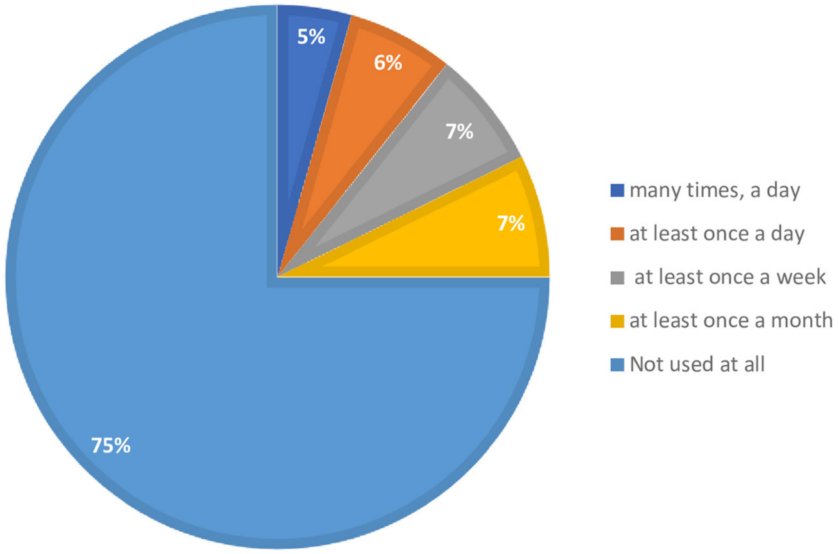


FIGURE 1
Participants' frequency of using TM applications.

for Social Sciences (SPSS) version 20 statistical software. Both descriptive and analytic statistics were utilized. The normal distribution of the data was examined using Q-Q plot and histogram. For descriptive analysis, results were expressed as numbers, percentages, and means (±SD). To estimate the potential relationship between patients' TM practice and other variables, Pearson correlation and cross tabulation were used. A *P*-value of <0.05 was considered statistically significant.

Operational definition

Chronic disease

In this study, a chronic disease, as defined by the U.S. National Center for Health Statistics (US-NCHS), is a disease lasting 1 year or longer (26) and it is a permanent; it leaves residual disability, it is caused by non-reversible pathological alteration, it requires special training of the patient for rehabilitation; or the patient may be expected to require a long period of supervision, observation, or care (15); it includes chronic respiratory diseases, including asthma, COPD, chronic circulatory diseases include HTN, chronic heart failure, chronic endocrine diseases include diabetes, thyroid disorders, chronic musculoskeletal disorders (including rheumatoid diseases), chronic digestive diseases (including pancreatitis), and liver diseases. This study also included the commonest chronic diseases proportionally in the study setting.

Results

Sociodemographic characteristics

The study included 384 (91%) of the 422 patients who were approached; other participants (9%) were not involved in the study due to being unable to continue the study. Approximately 232 (60.4%) of the participants were city dwellers. Around one-third (28.4%) of the respondents lacked formal education. Approximately, the monthly income of half (192 or 50.8%) of the participants was between 1,500 and 3,000 Ethiopian birr per month. More than half (214, 55.7%) of the patients used a cellphone, but only around one-third (132, 34.4%) of the participants used a smart phone. Most respondents (371 or 96.6%) used taxis or public transportation to get around. Every 2 months, 169 (44% of the participants) visit the hospital (Table 1).

Clinical characteristics of the participants

The most common type of chronic disease among the participants was cardiovascular disorder (84.9%), followed by endocrine disorder (69.0%) and respiratory disorder (65.9%). More than half of the patients (202, 52.6%) had diseases that had lasted <5 years More than half 219 (57%) of the participants had a hospital admission history within the previous 6 months. In the hospital, 143 (37.2%) of participants thought the wait was very long, while 135 (35.2%) thought it was long (Table 2).

Perceptions of participants toward telemedicine

The overall perceptions mean (±SD) score of the respondents was 3.92 (±1.06). About half (189 or 49.2%) of the participants perceived that the TM application was useful for making health decisions for their own health. The majority of the participants (281 or 73.1%) perceived the necessity of TM services for chronic patient care. Similar, the majority of the participants had a positive perception of TM service in terms of faster medical care delivery, saving effort and money, reducing transportation costs, reducing

TABLE 6 Practice of different methods of telemedicine application among chronic diseases patients in ambulatory clinic of UOGCSH, 2022 ($N = 384$).

Statement	Category	Frequency (%)
Have you ever searched online for information about a disease or medical problem	Yes	31 (8.9)
	No	350 (91.1)
Have you ever searched online for information about a doctor	Yes	16 (4.2)
	No	368 (95.8)
Have you ever typed information on an application about your dietary, physical exercise, and overall lifestyle modifications	Yes	16 (4.2)
	No	368 (95.8)
Have you ever typed information on an application about a chronic illness you have	Yes	18 (4.7)
	No	366 (95.3)
Have you ever renewed a prescription online	Yes	13 (3.4)
	No	371 (96.6)
Have you ever consulted your doctor using TM	Yes	11 (2.9)
	No	373 (97.1)
Have you ever used a personal health record for your health	Yes	7 (1.8)
	No	377 (98.2)
Have you ever looked at a test result online	Yes	13 (3.4)
	No	371 (96.6)
Have you ever used a device that measures health information (like blood pressure; blood glucose levels) that connects to your mobile/website application	Yes	5 (1.3)
	No	379 (98.7)
Have you ever posted anything online about your health or health care	Yes	5 (1.3)
	No	379 (98.7)
Have you ever Joined an online group that is for a health issue that you or your family member has	Yes	13 (3.4)
	No	371 (96.6)
Have you ever booked appointment with doctors using TM	Yes	29 (7.6)
	No	355 (92.4)
Overall perceived level of TM practice	High	94 (24.5)
	Low	290 (75.5)

waiting lists in medical care centers, improving communication between the patient and the doctor, and assisting in providing appropriate instructions in emergencies. Overall, close to three-fourths (273 or 71.1%) of the participants had a positive perception of telemedicine services (Table 3).

Perceptions toward privacy issue vs. telemedicine

Around half (48.7%) of the participants strongly believed that they would be worried about the privacy of their health information if it is online, and one quarter (24%) of the respondents strongly perceived that using computers, mobiles, or the internet disclosed their medical information for third person (Table 4).

Willingness to involve and practice of telemedicine

Around three-fourths (283 or 73.7%) of participants were at least familiar with or had heard about TM. More than one-third 160 (41.7%) of the participants were interested in tracking information about their chronic illness, their diet and calories (146 or 38%), exercise (150 or 39.1%), and being reminded when to take prescriptions (227 or 59.1%). Around half (49.7%) of participants were interested in using the TM website if it came from their doctor, while 216 (56.3%) and 203 (52.9%) were interested in using it if it came from the health insurance plan and the government, respectively. Overall, around two-thirds (63.3%) of participants were willing to participate in the TM service (Table 5).

Practice toward telemedicine

A lower proportion of participants practiced TM for their medical and health-related issues. In addition, three-fourths (75%) of the respondents reported that they did not use the TM application at all for their medical conditions (Figure 1). Overall, only about one-fourth (24.5%) of patients with chronic diseases were perceived to have a high practice of TM application for their medical condition (Table 6).

Association between patient demographic character and inclination to use TM

Cell phone or internet use, use of a smartphone, healthcare insurance coverage, anxious to use tablet or device, interest to use eHealth application, frequency of using health application, and frequency of follow-up time had a statistically significant association with TM practice (Table 7).

Discussion

The current study sought to determine the level of perception, willingness, and extent of TM practice among patients with chronic diseases. This study showed that most participants had a positive perception of and willingness to use TM services. However, their perceived practice level was low. The study's main finding is that familiarity with TM influences all attitudes, willingness, and practice of TM.

The current study showed that around three-fourths of participants had a positive perception of the TM service. This finding is similar to that of studies conducted in Saudi Arabia (27) and Bangladesh (28), which found that the majority of patients had a positive attitude toward TM. However, the current finding is contrary to a study in the Netherlands (14), which showed that the majority of respondents had a negative attitude toward the TM service. The discrepancy might be because of higher awareness regarding TM services use in the current time, especially related to COVID-19 epidemic, but the Netherlands' study was conducted before the epidemic, and the impact of TM may not have been sufficiently recognized earlier. The other possible reason might be that the eHealth implementation in the earlier study was collaborative with patients and carers. This may result in a lower attitude because of the service costs. Furthermore, technology-based applications might also be a reason for higher perception.

The majority of participants were willing to participate in the TM service. This finding is in line with earlier studies across the world regarding patients' willingness to use digital-based health services (26, 28–30). A previous study among diabetes patients in Ethiopia showed that more than two-thirds of participants were inclined to use mobile-based health services (30). A study conducted in Australia also revealed that the majority of participants were satisfied and willing to utilize TM if it reduced costs and waiting times (29). In addition, another study conducted in Pakistan showed that more than half of the participants were willing to rely on text messages to communicate with healthcare providers (31). Furthermore, according to a study conducted in Bangladesh, a high proportion of participants were eager to learn about TM. They believed that TM could benefit the ongoing advancement of healthcare (28). A Cochrane review also showed that a digital-based text message intervention improved patients' self-management practices and made patients more involved in the TM service (23). These findings may suggest that digital-based health interventions could be popular and advancement in the healthcare system from the patient's perspective.

However, in terms of TM practice, this study revealed that around three-fourths of the participants had overall low TM practice. This finding is consistent with a study conducted in the Netherlands, which found that most applications were practiced low (14). On the contrary, a study in Poland found that the majority of participants had used TM in various ways (32). In addition, according to an Australian study, close to half of the respondents used TM applications (29). Furthermore, studies in Pakistan (31) and Saudi Arabia (27) also showed that a higher proportion of respondents utilized TM services in different ways. The discrepancy could be because of differences in digital-enabled infrastructure and healthcare systems across settings. Internet access, socioeconomic status, and TM implementation differences might affect the application of the service among chronic disease patients. The other possible difference might be awareness of privacy issues and digital-related health applications. In this study, most participants worried and believed that the digital-based service jeopardized their privacy. This study suggests that patients can be attracted to a digital-based system by increasing awareness, promoting pushing factors and encouraging driving factors. The study revealed that the use of the internet, having a smartphone, and interest in and frequency of using health applications were among the significant factors associated with the practice of TM services among patients. This is because their experience with the practices and techniques helps them engage easily. As a result, better opportunities could be provided to fully implement the TM service in Ethiopia.

Implication of this study

Assessing the perception, willingness, and practice of long-term patients in TM implementation has a paramount role in establishing and implementing TM. Identifying barriers to its implementation is an initial step to providing possible solutions for better control of chronic diseases through TM. There is substantial evidence to support the benefits of TM in the control of chronic diseases and the general advancement of the health care delivery system. This study focused on assessing chronic disease patients' perception, willingness, and practice toward telemedicine. A well-developed perception, willingness, and practice of TM will provide a baseline for the identification of barriers and fill the gaps to ensure better health care through TM implementation. This will help design digital-based interventions for managing chronic disease patients. Furthermore, it will extend a body of knowledge to patients, carers, practitioners, the community, stakeholders, and the healthcare system community in general.

TABLE 7 Comparison of patients' background characteristics between patients in terms of practicing TM.

Characteristics	Category	TM practice		*P*-value
		High	Low	
Age	Mean (± SD)	46.4 (17.2)	48.58 (15.7)	0.12
Sex	Male	57 (27.8)	148 (72.2)	0.154
	Female	37 (20.7)	142 (79.3)	
Use of internet or cell phone	Yes, on my own	49 (22.9)	165 (77.1)	0.006*
	Yes, with the help of others	10 (55.6)	8 (44.4)	
	No	33 (21.9)	118 (78.1)	
Use of smart phone	Yes	38 (28.8)	94 (71.2)	0.001*
	My family use it	9 (69.2)	4 (30.8)	
	No	45 (18.8)	194 (81.2)	
Health coverage	Payment	31 (30.7)	70 (69.3)	0.007*
	Insurance	54 (20)	216 (80)	
	NGO	5 (42.3)	8 (57.8)	
Living condition	Living alone	10 (38.5)	16 (61.5)	0.092
	Living with family	79 (22.5)	272 (77.5)	
	Other	3 (42.9)	4 (57.1)	
Admission to the hospital	Yes	48 (21.9)	171 (78.1)	0.118
	No	46 (27.9)	119 (72.1)	
Fall in the way to hospital	Yes	13 (21.3)	48 (78.7)	0.237
	No	81 (25.1)	242 (74.9)	
Anxious to use tablet or device	Very much	14 (17.3)	67 (82.7)	0.021*
	Yes	24 (21.1)	90 (78.9)	
	No so much	12 (40)	18 (60)	
	No	44 (27.7)	115 (72.3)	
Interest to use eHealth application for your health information	Very interested	19 (41.5)	24 (58.5)	0.033*
	Somewhat interested	32 (24.8)	97 (75.2)	
	Not too interested	14 (19.2)	59 (80.8)	
	Not all interested	29 (20.6)	112 (79.4)	
Frequency of using health application	Many times, a day	10 (58.8)	7 (41.2)	<0.001*
	At least once a day	9 (37.5)	15 (62.5)	
	At least once a week	17 (63)	10 (37)	
	At least once a month	17 (60.7)	11(39.3)	
	Not used at all	39 (13.5)	249 (86.5)	
Frequency of follow-up time	Monthly	20 (17.7)	100 (82.3)	<0.001*
	Once every 2 months	37 (21.9)	132 (78.1)	
	Every 3 month	28 (35.4)	51 (64.6)	
	Semi annually	3 (25)	9 (75)	
	Annually	3 (75)	1 (25)	

*Indicates variables with $P < 0.05$.

The findings of this study will also serve as a baseline for future research in the area. In addition, these findings will also help to scale up and integrate with other digital health policies and initiatives.

Strengths and limitation of the study

The results of this study should be interpreted while remembering the following limitations: First, due to the nature of the study design, a cause-and-effect relationship could not be explained. The finding relies on self-reported outcomes from participants, which could depend on their faith and honesty. In addition, the sample size calculation did not take into account the prevalence of chronic diseases in Northwest Ethiopia. This may affect the sample size calculation results. Despite the limitations, our findings highlighted the attitude, willingness, and extent of TM practice among patients with chronic disease. It could be a baseline for future research in this area. In the future, it would be welcome to explore the opportunities and challenges of practicing a digital-based healthcare system in Ethiopia using a prospective national-based study.

Conclusion

This study concluded that although patients with chronic diseases perceived and willingly accepted TM services, their perceived practice level was low. Experience with health applications and the availability of the internet and cell phones were among the factors associated with practicing TM. In general, this study provides baseline information about perceptions, willingness, and actual practice of digital-based healthcare among Ethiopian chronic disease patients. Creating awareness of privacy issues vs. digital-based health applications and creating an opportunity may result in practicing the available digital health system among patients and even for an entire population.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Research proposal was reviewed and approved by the Research and Ethics Review Committees of the School of Pharmacy at the University of Gondar with a reference number of UOG-SOP-257/2022. The patients/participants provided their written informed consent to participate in this study.

Author contributions

EAB and AKS participated in the conception, design, administration, and supervising of the study, while DG, AKN, and EG contributed in the data curation, methodology, and interpreted the data. All authors drafted the initial manuscript, read, approved the final manuscript, and contributed the critical review and the content.

Acknowledgments

We want to express our gratitude to the clinical directors of UoGCSH for letting us carry out this study. We especially thank the study participants for their willing involvement.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Blandford A, Wesson J, Amalberti R, AlHazme R, Allwihan R. Opportunities and challenges for telehealth within, and beyond, a pandemic. *Lancet Glob Health.* (2020) 8:e1364–5. doi: 10.1016/S2214-109X(20)30362-4

2. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. *J Allergy Clin Immunol Pract.* (2020) 8:1489–91. doi: 10.1016/j.jaip.2020.03.008

3. Novara G, Checcucci E, Crestani A, Abrate A, Esperto F, Pavan N, et al. Telehealth in urology: a systematic review of the literature. How much can telemedicine be useful during and after the COVID19 pandemic? *Eur Urol.* (2020) 78:786–811. doi: 10.1016/j.eururo.2020.06.025

4. Sim I. Mobile devices and health. *N Engl J Med.* (2019) 381:956–68. doi: 10.1056/NEJMra1806949

5. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. *J Med Int Res.* (2020) 22:e16407. doi: 10.2196/16407

6. Calim S, Ulaş SC, Demirci H, Tayhan EB. Effects of high fidelity simulation model on midwives' shoulder dystocia management skills: an educational Intervention study. *Nigerian J Clin Pract.* (2022) 25:773. doi: 10.4103/njcp.njcp_1393_21

7. Breen G-M, Matusitz J. An evolutionary examination of telemedicine: a health and computer-mediated communication perspective. *Soc Work Public Health.* (2010) 25:59–71. doi: 10.1080/19371910902911206

8. Gu D, Humbatova G, Xie Y, Yang X, Zolotarev O, Zhang G. Different roles of telehealth and telemedicine on medical tourism: an empirical study from Azerbaijan. *InHealthcare.* (2021) 9:1073. doi: 10.3390/healthcare9081073

9. Lustig TA. *The Role of Telehealth in an Evolving Health Care Environment: Workshop Summary.* Washington, DC: National Academies Press (2012). Available online at: https://nap.nationalacademies.org/catalog/13466/the-role-of-telehealth-in-an-evolving-health-care-environment

10. Assaye BT, worku Shimie A. Telemedicine use during COVID-19 pandemics and associated factors among health professionals working in health facilities at resource-limited setting 2021. *Inform Med Unlocked.* (2022) 33:101085. doi: 10.1016/j.imu.2022.101085

11. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. *Irish J Med Sci.* (2021) 190:1–10. doi: 10.1007/s11845-020-02299-z

12. Chitungo I, Mhango M, Mbunge E, Dzobo M, Musuka G, Dzinamarira T. Utility of telemedicine in sub-Saharan Africa during the COVID-19 pandemic. A rapid review. *Hum Behav Emerg Technol.* (2021) 3:843–53. doi: 10.1002/hbe2.297

13. Jnr BA. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. *J Med Syst.* (2020) 44:132. doi: 10.1007/s10916-020-01596-5

14. Hofstede J, De Bie J, Van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. *Int J Med Inform.* (2014) 83:967–74. doi: 10.1016/j.ijmedinf.2014.08.011

15. Shirwar AK, Bharati AS. An ayurvedic management of kitibha kushta (plaque psoriasis)-a case study. *J Ayurveda Integr Med Sci.* (2022) 7:194–9. Available online at: https://jaims.in/jaims/article/view/1874

16. Yosef T. Prevalence and associated factors of chronic non-communicable diseases among cross-country truck drivers in Ethiopia. *BMC Public Health.* (2020) 20:7. doi: 10.1186/s12889-020-09646-w

17. Tesfay FH, Zorbas C, Alston L, Backholer K, Bowe SJ, Bennett CM. Prevalence of chronic non-communicable diseases in Ethiopia: a systematic review and meta-analysis of evidence. *J Frontiers in Public Health.* (2022) 10:936482. doi: 10.3389/fpubh.2022.936482

18. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. *Sens Int.* (2021) 2:100117. doi: 10.1016/j.sintl.2021.100117

19. Fraiche AM, Eapen ZJ, McClellan MB. Moving beyond the walls of the clinic: opportunities and challenges to the future of telehealth in heart failure. *JACC Heart Fail.* (2017) 5:297–304. doi: 10.1016/j.jchf.2016.11.013

20. Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* (2015) 9:CD002098. doi: 10.1002/14651858.CD002098.pub2

21. Mao Y, Lin W, Wen J, Chen G. Impact and efficacy of mobile health intervention in the management of diabetes and hypertension: a systematic review and meta-analysis. *BMJ Open Diab Res Care.* (2020) 8:e001225. doi: 10.1136/bmjdrc-2020-001225

22. Alboraie M, Allam MA, Youssef N, Abdalgaber M, El-Raey F, Abdeen N, et al. Knowledge, applicability, and barriers of telemedicine in Egypt: a national survey. *Int J Telemed Appl.* (2021) 2021:5565652. doi: 10.1155/2021/5565652

23. Shi C. Mobile phone messaging for facilitating self-management of long-term illnesses: summaries of nursing care-related systematic reviews from the cochrane library. *Int J Evid Based Healthc.* (2013) 11:344–5. doi: 10.1111/1744-1609.12041

24. Kola L, Kohrt BA, Hanlon C, Naslund JA, Sikander S, Balaji M, et al. COVID-19 mental health impact and responses in low-income and middle-income countries: reimagining global mental health. *Lancet Psychiatry.* (2021) 8:535–50. doi: 10.1016/S2215-0366(21)00025-0

25. Manyazewal T, Woldeamanuel Y, Blumberg HM, Fekadu A, Marconi VC. The potential use of digital health technologies in the African context: a systematic review of evidence from Ethiopia. *NPJ Digital Med.* (2021) 4:125. doi: 10.1038/s41746-021-00487-4

26. Bernell S, Howard SW. Use your words carefully: what is a chronic disease? *Front Public Health.* (2016) 4:159. doi: 10.3389/fpubh.2016.00159

27. Alajwari HA, Alfayez A, Alsalman D, Alanezi F, Alhodaib H, Al-Rayes S, et al. Knowledge and attitude of Saudi Arabian citizens towards telemedicine during the COVID-19 pandemic. *Int Health.* (2022) 14:604–9. doi: 10.1093/inthealth/ihab082

28. Haque MMA, Jahan Y, Khair Z, Moriyama M, Rahman MM, Sarker MH, et al. Perceptions about Telemedicine among Populations with Chronic Diseases amid COVID-19: data from a Cross-Sectional Survey. *Int J Environ Res Public Health.* (2022) 19:4250. doi: 10.3390/ijerph19074250

29. Shiferaw F, Zolfo M. The role of information communication technology (ICT) towards universal health coverage: the first steps of a telemedicine project in Ethiopia. *Glob Health Action.* (2012) 5:15638. doi: 10.3402/gha.v5i0.15638

30. Jemere AT, Yeneneh YE, Tilahun B, Fritz F, Alemu S, Kebede M. Access to mobile phone and willingness to receive mHealth services among patients with diabetes in Northwest Ethiopia: a cross-sectional study. *BMJ Open.* (2019) 9:e021766. doi: 10.1136/bmjopen-2018-021766

31. Iftikhar S, Saqib A, Sarwar MR, Sarfraz M, Arafat M, Shoaib Q-u-a. Capacity and willingness to use information technology for managing chronic diseases among patients: a cross-sectional study in Lahore, Pakistan. *PLoS ONE.* (2019) 14:e0209654. doi: 10.1371/journal.pone.0209654

32. Prevett M. Chronic non-communicable diseases in ethiopia-a hidden burden. *Ethiop J Health Sci.* (2012) 22(S):1–2. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3542741/

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 12 September 2023
DOI 10.3389/fpubh.2023.1245415

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Jose Luis Martinez-Flores,
Popular Autonomous University of the State of Puebla, Mexico
R. Uthayakumar,
Gandhigram Rural Institute, India

*CORRESPONDENCE
Zhe Zhou
✉ wg6170@126.com

RECEIVED 23 June 2023
ACCEPTED 22 August 2023
PUBLISHED 12 September 2023

CITATION
Li H, Zhou Z, Hu C, Ning Y, Wang Z, Huang H and Tao K (2023) Optimization of emergency allocation of necessities of life based on fractal perspective.
Front. Public Health 11:1245415.
doi: 10.3389/fpubh.2023.1245415

Optimization of emergency allocation of necessities of life based on fractal perspective

Hong Li[1,2], Zhe Zhou[3]*, Chuan Hu[2], Yumei Ning[2], Zihao Wang[2], Hua Huang[4] and Kangsheng Tao[2]

[1]Department of Economics and Social Development, Gansu Normal College for Nationalities, Hezuo, Gansu, China, [2]School of Business Administration, Zhongnan University of Economics and Law, Wuhan, Hubei, China, [3]School of Public Health and Management, Wenzhou Medical University, Wenzhou, Zhejiang, China, [4]Business School of Hbue, Hubei University of Economics, Wuhan, Hubei, China

The supply and reserves of emergency necessities of life are important for emergency management in disaster events. The scope of the necessities of life changes with social development, and their reserves and allocation in sudden disaster events continually face new challenges. Timely distribution of the necessities of life during disasters is critical to saving lives and maintaining social order. Therefore, this study proposes a fractal multi-level distribution network (FMDN) optimization model with multiple warehouse points, multiple emergency distribution centers, and multiple disaster points from the perspective of fractal theory. The FMDN model considers the influence of road damage on vehicle speed and the dynamic change in demand at the affected points. The FMDN model aims to minimize the operating costs of a distribution network, including the cost of building emergency reserve points, transportation costs, and penalty costs for lack of demand. Numerical experiments verify the feasibility of the model. The FMDN model is solved using LINGO software programming, and an optimal distribution path and quantity are obtained. Analyzing the numerical example results shows that the model is suitable for emergencies and has good applicability.

KEYWORDS

necessities of life, emergency allocation, optimization model, demand disturbance, fractal perspective

1. Introduction

Since the twenty-first century, the frequent occurrence of large-scale sudden disasters has caused huge economic costs to global society. Notably, China is one of the countries most affected by frequent disasters. The United Nations *Global Assessment Report on Disaster Risk Reduction* points out that global disaster risk events are on the rise, especially natural disaster losses caused by climate change and environmental damage, causing the world $2.5 trillion in direct losses this century. Earthquakes and tropical storms cause direct losses of more than 180 billion US dollars a year globally. According to the *Economists War-Game Pandemic Threat to Global Growth* analysis report, the outbreak of the public health emergency of COVID-19, which created a pandemic, may ultimately destroy 1% of global GDP and cost global GDP more than $1.1 trillion (1).

In recent years, the number of disasters worldwide has been increasing. For example, on the morning of 7 February 2021, a glacier in the Jamori district of the Indian state of

Uttarakhand collapsed, triggering massive flooding that left more than 207 people dead or missing and forcing the emergency evacuation of thousands of people from surrounding areas. In June–July 2021, the United States and Canada experienced historic heat and drought. The heat and drought resulted in more than 1,319 deaths in the U.S. and Canada, thousands of people seeking emergency medical treatment for heat-related discomfort and illness, and more than 58 million people affected by the heat and drought. On 16 December 2021, Typhoon Rey struck the Philippines, killing 378 people, leaving 60 missing and 742 injured, and affecting 3.95 million people.

Major natural disasters in China in recent years include the following: On 12 May 2008, a massive earthquake measuring 8.0 on the Richter scale struck Wenchuan County, Sichuan Province, killing 69,227 people, injuring 374,643, and leaving 17,923 missing. On 14 April 2010, an earthquake measuring 7.1 on the Richter scale struck Yushu, Qinghai, killing 2,698 people and leaving 270 missing. On 7 August 2010, a mudslide in Zhouqu County, Gannan Prefecture, killed 1,463 people, left 302 missing, and injured 72. On 3 August 2014, an earthquake measuring 6.5 on the Richter scale struck Ludian County, Yunnan Province, affecting 1,088,400 people and killing 617. With the development of urbanization, China's main population distribution is one of the spatial aggregations characterized by urban agglomerations and super-large urban agglomerations. During large-scale emergencies, the demand for rescue materials increases geometrically in the short term. The demand structures for medical materials, necessities, emergency rescue materials, and restoration and reconstruction materials in disaster areas are complex, and the types and quantities of material demand at different stages show large fluctuations and dynamic uncertainties. This makes it difficult for emergency management to accurately predict material needs at affected sites and often leads to bottleneck problems, such as configuration delays.

The main research content of fractal theory is the scale distribution characteristics of fractal systems, that is, using mathematical methods to reveal the intrinsic scale invariance of irregular shapes in nature and to analyze the original systems' characteristics and intrinsic evolution laws. The application of fractal theory in the distribution of necessities of life helps the regional fractal distribution units integrate and utilize limited resources, make independent decisions, adapt and optimize themselves, and complete the superior distribution tasks on time.

This study focuses on life necessities' connotation, category, demand, and distribution characteristics. Considering the uncertainty of demand quantity, the fractal theory was adopted to efficiently transport various life necessities from warehouses to the disaster area to minimize the total delivery time. It is vital to meet the basic needs of people in disaster-stricken areas concerning the necessities of life and to maintain social order there.

2. Literature review

2.1. Literature review on necessities of life

The necessities of life are defined in both a broad and a narrow sense. In the broad sense, they refer to goods that meet people's basic living needs, and the broad sense is increasing in scope: In a developing social economy, with advances in social production technology and an increase in product use, non-life necessities are gradually transformed into life necessities over a long period of time. In the narrow sense, the necessities of life comprise four categories: clothing, food (including firewood, rice, oil, salt, and sanitation), housing (temporary rental housing available for replacement), and transportation (e.g., public transportation). These constitute the basic means of subsistence, the basic requirements for human beings to live their lives.

Pantazis et al. (2–5) found that British people's ideas on what life necessities comprised were usually broader and more multidimensional than those assessed by experts. Their social activities, roles, and relationships were key factors that determined the content of their necessities of life. Gordon et al. (6) believed that public identification with a range of social activities, roles, and relationships was an important factor in determining the necessities of Guernsey residents. Necessities perceived by the public depend on the interaction between the market availability and the social development structure on which the current living environment depends. Fahmy et al. (7) argued that the extent to which there is a clear public consensus on the necessities of life needs to be further explored and suggested that additional methods be applied to determine public perceptions.

Research on the necessities of life at home and abroad focuses on the establishment and improvement of emergency reserve systems, the improvement of relevant emergency planning systems, and the integrated optimization modeling of the location-routing problem (LRP) of emergency distribution centers. Teimoury et al. (8) studied the vegetable import quota problem in Tehran city through system dynamics model simulation. Brady (9) simulated the coordination and response of information flow and medical, police, and fire resources to help emergency managers quickly develop robust emergency plans to deal with potential threats.

2.2. Literature review on fractal theory

The fractal theory has been applied to image processing, universe exploration, financial analysis, medical diagnosis, earthquake research, and logistics. Guo et al. (10) and Bocewicz et al. (11) studied the scale characteristics of public transport networks (PTNs) in L-space through fractal analysis and considered the effect of the real bus routes, providing new perspectives and tools for human migration in spatial networks. Saad and Bahadori (12) proposed a new information fractal framework to study the improvement in the sustainability of an entire food distribution chain using two variables: the greenfield service limit and the minimum-on-board-vehicle-weight fill level. Webber and Dunbar (13) studied the fractal structure of the distribution of communities of practice and analyzed the implications for the business management structure. Zhang and Li (14) studied the multifractality of traffic flow at the spatial scale and used it to quantify the uniformity of the flow's spatial distribution. Li et al. (15) combined a fractal method with a passenger flow allocation model to establish a fractal quantification method to measure the temporal and spatial distribution and the characteristics of passenger flow in rail transit networks. Man and Chen (16) studied the fractal and fractal dimension characteristics of Shenzhen, China. Shao et al. (17) studied the multifractality of three major cryptocurrencies using multifractal detrended fluctuation analysis (MFDFA).

Fractal theory for the study of finance has also been studied by many scholars (18–21). Jiang et al. (22) used multiple fractals to quantify financial market inefficiencies in the context of risk management. Fractal theory has also found application in the field of logistics distribution. Jingwen (23) studied the feasibility of using fractal theory to design the spatial organization of logistics companies. Ryu et al. (24) proposed a fractal-based inventory management (fVMI) model to minimize inventory costs and satisfy customer demand. Yue et al. (25) proposed a fractal hierarchical honeycomb structure for vehicle routing problem (VRP) solving. Bi et al. (26) constructed a correlation model of regional logistics dynamics using multifractal theory; they studied the coupling relationship between regional logistics dynamics and multifractal dynamics using empirical analysis.

A number of scholars have studied the optimization of material distributions under uncertain demand. Wu and Peng (27) proposed a chance-constrained model with uncertain customer demand and center-setting costs, constructing and solving the model using uncertainty theory. Wen et al. (28) studied the facility location-allocation (FLA) problem in a random environment, established an uncertain expected value model based on uncertain measures, and used a hybrid intelligence algorithm to solve numerical examples. Liu et al. (29) studied the problem of locating a multi-product logistics distribution center in an uncertain environment and constructed a total annual cost model to minimize construction, management, inventory, and transportation costs; they validated the model with numerical examples using Lingo software. Xu and Qi (30) established a dual-objective mixed-integer linear programming model to solve the problem of multi-point gasoline-emergency-distribution vehicle-path optimization with vehicle sharing and time window coordination. Cui et al. (31) proposed an objective function to minimize operating costs under uncertain demand and established a logistics center location optimization model with a three-node expansion mechanism. Xing et al. (32) introduced the fractal theory into the emergency organization construction of coastal cities in China. They proposed a virtual network fractal emergency organization structure with self-organization, self-optimization, and self-affinity characteristics. Using Typhoon "Lichma" as an example, they verified the efficient, dynamic, and open performance of virtual network fractal emergency organizations in adapting to changes in disaster scenarios. Li et al. (33) applied computational mathematical organization theory to construct a fractal emergency organization optimization model. Li et al. (34) studied the emergency response collaborative organization model and proposed an unconventional emergency response organization model based on fractal theory.

The current research on emergency necessities is inadequate. The public's perception of the content of the necessities of life needs to be redefined using a new methodology. With the development of technology and social progress, the scope of necessities is expanding, and the distribution of emergency necessities in sudden disasters consequently faces new challenges. This study, therefore, applies fractal theory to an emergency distribution network model of the necessities of life to examine these challenges.

3. Fractal-based emergency allocation framework for life necessities

The American mathematician Benoit B Mandelbrot (35) defined fractals as follows: "Fractals are geometric figures or natural shapes composed of parts that resemble the whole in some way." Fractals have the following characteristics: (a) Parts of a figure or object in a fractal are the same as the whole formal structure, but different in size, and may be slightly deformed. (b) Regardless of the level of examination, a fractal's form is either extremely irregular, highly discontinuous, or fragmented. (c) A fractal contains some "special elements" that vary greatly in rank and cover an extremely wide range.

Figure 1 shows the framework for establishing the emergency allocation of life necessities. A fractal distribution center framework is proposed to solve the problem of emergency configuration of life necessities; its core purpose is to obtain a function of self-organization, self-similarity, and self-optimization of sub-fractal units. Both the fractal distribution center layer and the sub-distribution center layer adopt a goal-driven mechanism: On the one hand, they are subject to the top-level distribution center goal, and on the other hand, they can optimize their internal processes and dynamically adapt to changes in the external environment and changes in the top-level goals. Fractal elements are autonomous units with autonomous decision-making, autonomous adaptation, and the execution of top-level distribution tasks.

Fractal elements are similar to the whole in some form, and a fractal emergency configuration center for necessities includes a top-level distribution center layer, a fractal distribution center layer, and a sub-distribution center layer. The top-level distribution center layer is the decision-making command center for emerging fractal configurations of the necessities of life. Generally, government agencies play this role and issue specific orders for the location and quantity of specific warehouses for the transportation of necessities during emergencies. The fractal distribution center layer is a combination of distribution platform centers that include various entities (e.g., governments, enterprises, logistics, and alliances) in a region. The fractal distribution center layer is composed of a combination of different distribution modes (e.g., third-party logistics and fourth-party logistics), whose common goal is to accomplish the task of emergency distribution of necessities. The bottom layer is the sub-distribution center layer, which is responsible for the execution of distribution tasks. The sub-distribution center layer is a combination of various fractal enterprise distribution platform centers, government fractal distribution platform centers, and fractal logistics alliance centers, providing emergency allocation of necessities of life.

Unlike previous studies, the fractal emergency distribution framework of the necessities of life can decompose the distribution tasks and alleviate the distribution pressure in a short time. Under the goal-driven mechanism, the sub-fractal units make independent decisions, adapt, and organize themselves independently, which can timely and reliably achieve the distribution goal of the top-level decomposition. The key feature is that, on the one hand, fractal distribution units increase the potential scope of collaborative units. On the other hand, the fractal distribution unit makes independent decisions and optimizes independently, which saves resources and improves distribution efficiency while achieving goals.

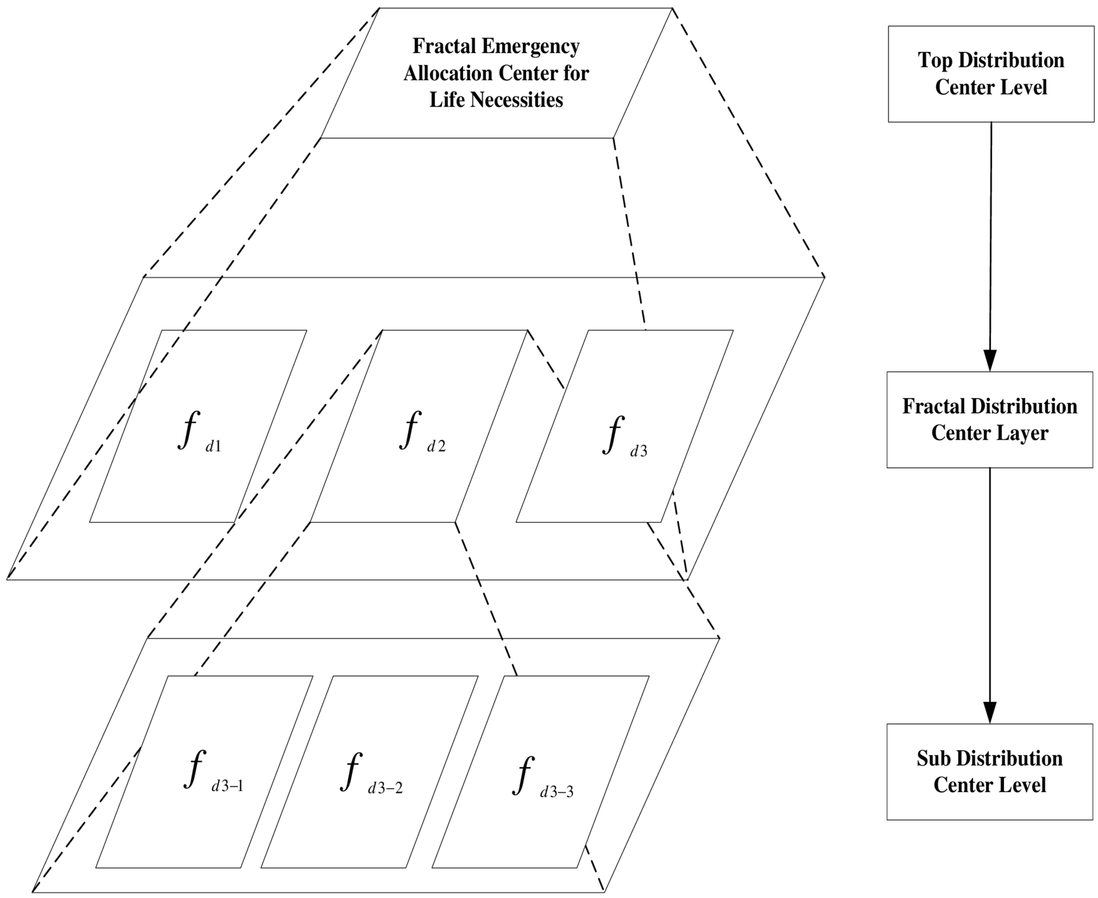


FIGURE 1
Fractal emergency allocation framework for life necessities.

4. Optimization model for the emergency allocation of life necessities under demand disturbance

4.1. Problem description

In large-scale emergencies, due to the limited supply of life necessities near the disaster site, it is necessary to send them from the emergency reserve point to the necessities' distribution center as soon as possible. Emergency necessities first arrive at the emergency distribution center and are then distributed to each emergency demand point. The problem is to construct a fractal emergency distribution network optimization model that encompasses multiple emergency storage points, multiple emergency distribution centers, and multiple disaster demand points. The model must take into account the construction costs of emergency reserve points, emergency distribution center costs, transportation costs, and demand shortage penalty costs. The road condition coefficient and the penalty cost of lack of demand are introduced into the model. Finally, to minimize the operating costs of the emergency distribution network, an emergency allocation optimization model is constructed; this model is for necessities with uncertain demand and takes into account road access conditions. The fractal emergency distribution network of necessities of life is shown in Figure 2.

4.2. Model assumptions

(1) During the emergency rescue period, the carrying capacity of vehicles at each stage is limited, and the vehicles adopt a round-trip distribution mode.

(2) In the case of a road network interruption, there is at least one other path to reach the disaster point. The road condition coefficient influences the delivery time and speed. Basic road condition information (travel time, travel cost, and road condition coefficient) is known.

(3) Different kinds of emergency necessities are packaged and transported by independent vehicles.

(4) The end of the rescue mission is marked by the receipt of no less than the required amount of necessities at each emergency demand point.

4.3. Variable description

R is the set of alternative nodes for emergency reserve points ($r \in R$).

D is the set of alternative nodes for emergency distribution centers ($d \in D$).

Q is the set of emergency demand points for life necessities ($q \in Q$).

M is the set of emergency necessity categories ($m \in M$).

K is the set of fractal distribution stages for emergency life necessities ($k \in K$).

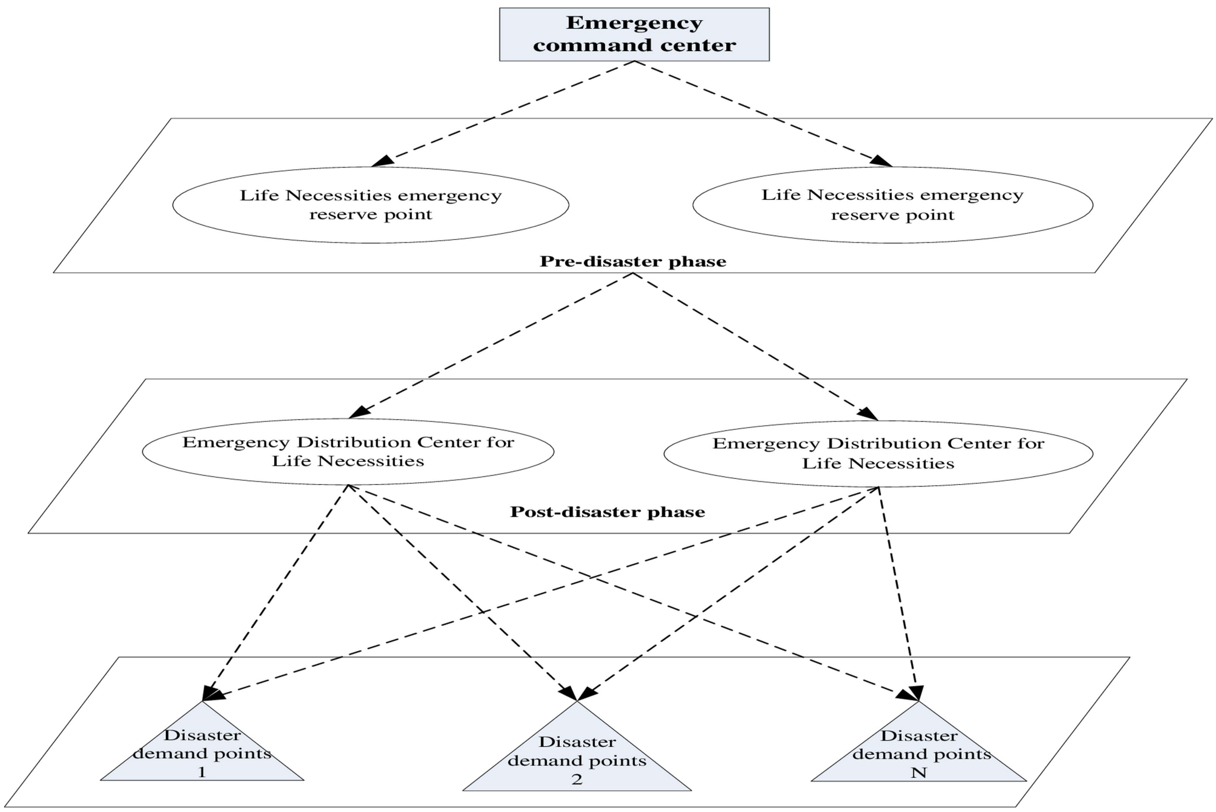


FIGURE 2
Fractal emergency distribution network for life necessities.

C_{trm}^{k} is the fixed cost of establishing the *r*th emergency reserve point for Category *m* of life necessities in stage *k*.

C_{tdm}^{k} is the fixed cost of establishing the *d*th emergency distribution center for Category *m* of life necessities in stage *k*.

Y_{rdm}^{k} is the unit mileage transportation cost of distributing life necessities *m* from the emergency reserve point *r* to the emergency distribution center *d* in stage *k*.

Y_{dqm}^{k} is the unit mileage transportation cost of distributing life necessities *m* from the emergency distribution center *d* to the emergency demand point *q* in stage *k*.

S_{rm}^{k} is the supply of life necessities *m* from the emergency reserve point *r* in stage *k*.

d_{qm}^{k} is the actual demand for life necessities *m* at the emergency demand point *q* in stage *k*.

L_{rd} is the shortest distribution distance from the emergency reserve point *r* to the emergency distribution center *d*.

ξ_{qm}^{k} is the penalty price per unit of shortage of emergency life necessities *m* at the emergency demand point *q* in stage *k*.

X_{rdm}^{k} is the distribution volume of emergency life necessities *m* from the emergency reserve point *r* to the emergency distribution center *d* in stage *k*.

X_{dqm}^{k} is the distribution volume of emergency life necessities *m* from the emergency distribution center *d* to the emergency demand point *q* in stage *k*.

V_{dqm}^{K} is the delivery speed of life necessities *m* from the emergency distribution center *d* to the emergency demand point *q* in stage *k*. As the existing traffic routes are often damaged after emergencies, the normal speed of the distribution vehicles for life necessities arriving in the disaster area will inevitably be affected. In this study, the average speed of the vehicles is represented by $\overline{V_{dqm}^{k}}$.The formula is given as follows:

$$V_{dqm}^{k} = \beta_{dq}^{k} * \overline{V_{dqm}^{k}} \tag{1}$$

in Constraint (1), β_{dq}^{k} is the road condition coefficient for the emergency distribution center *d* to emergency demand *q* in stage *k*. $\beta_{dq}^{k} \in [0,1]$; the larger the value, the better the traffic condition of the road.

t_{dqm}^{k} is the delivery time of life necessities *m* from the emergency distribution center *d* to emergency demand *q* in stage *k*.

$\hat{E}_{qm}^{k}$ is the shortage of life necessities *m* at emergency demand *q* in stage *k*.

f_{rm}^{k} is the binary variable. When life necessities *m* are selected at the emergency reserve point *r* in stage *k*, it equals 1; if not, it equals 0.

f_{dm}^{k} is the binary variable. When life necessities *m* are selected at the emergency distribution center *d* in stage *k*, it equals 1; if not, it equals 0.

4.4. Model building

The occurrence of a sudden catastrophic event can result in the destruction of transportation routes, and the speed of material distribution vehicles can be seriously affected. This model uses a road condition coefficient and demand disturbance parameters to solve for the optimum distribution route and optimum distribution quantity of an emergency distribution network.

A fractal multi-stage emergency distribution network model that considers road traffic conditions and constant demand. Hence, we have the resulting model:

$$\begin{aligned}\min Z = &\sum_{r\in R}\sum_{m\in M}\sum_{k\in K} C_{trm}^{k} f_{rm}^{k} + \sum_{r\in R}\sum_{d\in D}\sum_{m\in M}\sum_{k\in K} Y_{rdm}^{k} L_{rd} X_{rdm}^{k} \\ &+ \sum_{d\in D}\sum_{m\in M}\sum_{k\in K} C_{tdm}^{k} f_{dm}^{k} \\ &+ \sum_{d\in D}\sum_{q\in Q}\sum_{m\in M}\sum_{k\in K} Y_{dqm}^{k} \beta_{dq}^{k} t_{dqm}^{k} \overline{v_{dqm}^{k}} X_{dqm}^{k} \\ &+ \sum_{q\in Q}\sum_{m\in M}\sum_{k\in K} \hat{E}_{qm}^{k} \xi_{qm}^{k}\end{aligned} \quad (2)$$

$$\sum_{d\in D} X_{dqm}^{k} \geq \mathrm{d}_{qm}^{k} + \hat{E}_{qm}^{k}, q\in Q, \quad m\in M, k\in K \quad (3)$$

$$\sum_{r\in R} X_{rdm}^{k} \leq S_{rm}^{k} f_{rm}^{k}, \quad d\in D,\ m\in M, k\in K \quad (4)$$

$$\sum_{r\in R} X_{rdm}^{k} = \sum_{d\in D} X_{dqm}^{k}, \quad q\in Q,\ m\in M,\ k\in K \quad (5)$$

$$\sum_{r\in R} f_{rm}^{k} \geq 1, \quad m\in M,\ k\in K \quad (6)$$

$$\sum_{d\in D} f_{dm}^{k} \geq 1, \quad m\in M,\ k\in K \quad (7)$$

$$f_{rm}^{k} \in \{0,1\}, \quad r\in R, m\in M, k\in K \quad (8)$$

$$f_{dm}^{k} \in \{0,1\} \quad d\in D,\ m\in M,\ k\in K \quad (9)$$

$$X_{rdm}^{k} \geq 0, r\in R, d\in D, m\in M, k\in K \quad (10)$$

$$X_{dqm}^{k} \geq 0, d\in D, q\in Q, m\in M, k\in K \quad (11)$$

$$\hat{E}_{qm}^{k} \geq 0, \quad q\in Q,\ m\in M,\ k\in K \quad (12)$$

Constraint (2) is the objective function. The formula aimed to minimize the total cost of the emergency distribution network by adding volume-shortage parameters and its penalty cost. The total cost includes the fixed construction cost of the emergency reserve point, the transportation cost from the emergency reserve point to the emergency distribution center, the fixed construction cost of the emergency distribution center, the transportation cost from the emergency distribution center to the emergency demand point, and the penalty cost of the demand shortage at the emergency demand point. Constraint (3) states that the number of necessities allocated to each emergency demand point should satisfy the demand as much as possible in stage k. Constraint (4) is that in stage k, the distribution quantity of the necessities of life m from the emergency reserve point r to the emergency distribution center d should be less than or equal to the supply quantity of the necessities of life m supplied by the emergency reserve point r, and only the selected emergency reserve point can distribute. Constraint (5) is the sum of the emergency life necessities supplied from the emergency reserve point, equal to the sum of the delivery volume of the emergency distribution center. Constraint (6) established at least one emergency reserve point for life necessities in the emergency distribution network in stage k. Constraint (7) established at least one emergency distribution center for life necessities in the emergency distribution network in stage k. Constraints (8) and (9) are decision variable constraints. Constraints (10), (11), and (12) are non-negative constraints.

In reality, large-scale emergencies lead to dynamic changes in the number of emergency life necessities demanded in disaster areas. In the objective function (2), it is difficult to accurately simulate the number of necessities of life under disaster by studying the emergency distribution network model with determined demand. Since the solution to constraint (2) of the objective function is easy to find, constraint (2) is not solved further.

To accurately simulate the changes in demand for life necessities during a disaster event, the following analysis is performed:

In this study, the following interval estimates are given for the demand for different categories of life necessities m in stage k:

$d_{qm}^{k} \in [\overline{d_{qm}^{k}},\ \widehat{d_{qm}^{k}}]$ ($q\in Q$, $m\in M$, $k\in K$), where $\overline{d_{qm}^{k}}$ is the nominal demand value of life necessities m at demand point q in stage k. $\widehat{d_{qm}^{k}}$ is the maximum disturbance value that deviates from the nominal value. $\widehat{d_{qm}^{k}} / \overline{d_{qm}^{k}}$ is the disturbance coefficient.

In emergencies, the probability of the number of necessities of life reaching the limit is extremely low. That is, when $d_{qm}^{k} = \overline{d_{qm}^{k}}$, the demand forecast is insufficient. When $d_{qm}^{k} = \overline{d_{qm}^{k}} + \widehat{d_{qm}^{k}}$, there is a surplus of emergency life necessities. In this study, a control parameter γ_{qm}^{k} is introduced to control the disturbance of the demand for life necessities at each emergency demand point in the interval $\left[\overline{d_{qm}^{k}}, \widehat{d_{qm}^{k}}\right]$:

$$H_{qm}^{k} = \frac{d_{qm}^{k} - \overline{d_{qm}^{k}}}{\widehat{d_{qm}^{k}}} \leq \gamma_{qm}^{k} \left(q\in Q, m\in M, k\in K\right) \quad (13)$$

By introducing the control parameter γ_{qm}^k into the model, the solution infeasibility that occurs when demand d_{qm}^k varies within an interval is avoided. If γ_{qm}^k=0, then d_{qm}^k=$\overline{d_{qm}^k}$ is a deterministic problem with demand equal to the lower bound of the interval. if γ_{qm}^k=1, then d_{qm}^k= $\overline{d_{qm}^k} + \widehat{d_{qm}^k}$ is an absolutely robust correspondence problem. Therefore, the range of demand disturbance can be controlled by adjusting the value of γ_{qm}^k between the interval $[0,1]$. This provides a model solution for demand disturbance. Thus, constraint (3) can be transformed into the following:

$$\sum_{d\in D} X_{dqm}^k \ge \overline{d_{qm}^k} + \gamma_{qm}^k \hat{d}_{qm}^k + \hat{E}_{qm}^k \left(q\in Q, m\in M, k\in K\right) \quad (14)$$

Therefore, a fractal multi-stage emergency allocation optimization model of life necessities is built considering road traffic conditions under demand disturbance:

$$\begin{aligned} \min Z = & \sum_{r\in R}\sum_{m\in M}\sum_{k\in K} C_{trm}^k f_{rm}^k + \sum_{r\in R}\sum_{d\in D}\sum_{m\in M}\sum_{k\in K} Y_{rdm}^k L_{rd} X_{rdm}^k \\ & + \sum_{deD}\sum_{meM}\sum_{keK} C_{trm}^k f_{rm}^k + \sum_{d\in D}\sum_{q\in Q}\sum_{m\in M}\sum_{k\in K} Y_{dqm}^k \beta_{dq}^k t_{dqm}^k \\ & \overline{v_{dqm}^k}\left(\overline{d_{dqm}^k} + \gamma_{qm}^k \widehat{d_{qm}^k} + \hat{E}_{qm}^k\right) \\ & + \sum_{q\in Q}\sum_{m\in M}\sum_{k\in K} \hat{E}_{qm}^k \xi_{qm}^k \end{aligned} \quad (15)$$

Subject to Constraint (4) ~ Constraint (14).

5. Numerical experiments

In the 7.1 Richter scale earthquake that struck Yushu City, Qinghai Province, China, the emergency distribution of necessities, such as food, medicine, and tents, was urgently needed by the victims. Let there are six life necessities reserve points, four regional emergency distribution centers, and six emergency demand points in the provincial, prefectural, and municipal emergency distribution network that require the supply of life necessities, as shown in Figure 3. The minimum number of emergency distribution centers in the emergency allocation network is 2 (Tables 1–9). The relevant data are as follows:

LINGO is a software developed by LINDO that can solve large-scale linear and non-linear programming problems and deal with optimization problems. In this example, to verify the model's effectiveness, the penalty cost of the lack of quantity at the demand point in an emergency is considered, as is road damage. The model was verified and solved using LINGO software to minimize the operating cost of a three-level fractal emergency distribution network for a given class of life necessities. We used LINGO11.0 software, 8 G memory, and a 1.80 GHz AMD processor to solve the model. The model obtained optimal transportation volumes, as shown in Tables 10, 11.

The emergency reserve point obtained from the model solution is $r4$, and emergency distribution centers were selected as $d4$ and $d3$. Since the goal of the solution is to achieve the lowest total cost of the system, only one distribution center is used in the actual configuration. The total operating cost of the distribution network was 367.31 million yuan. The distribution path is $(r4, d3)$. Emergency life necessities are first distributed from the $r4$ emergency reserve

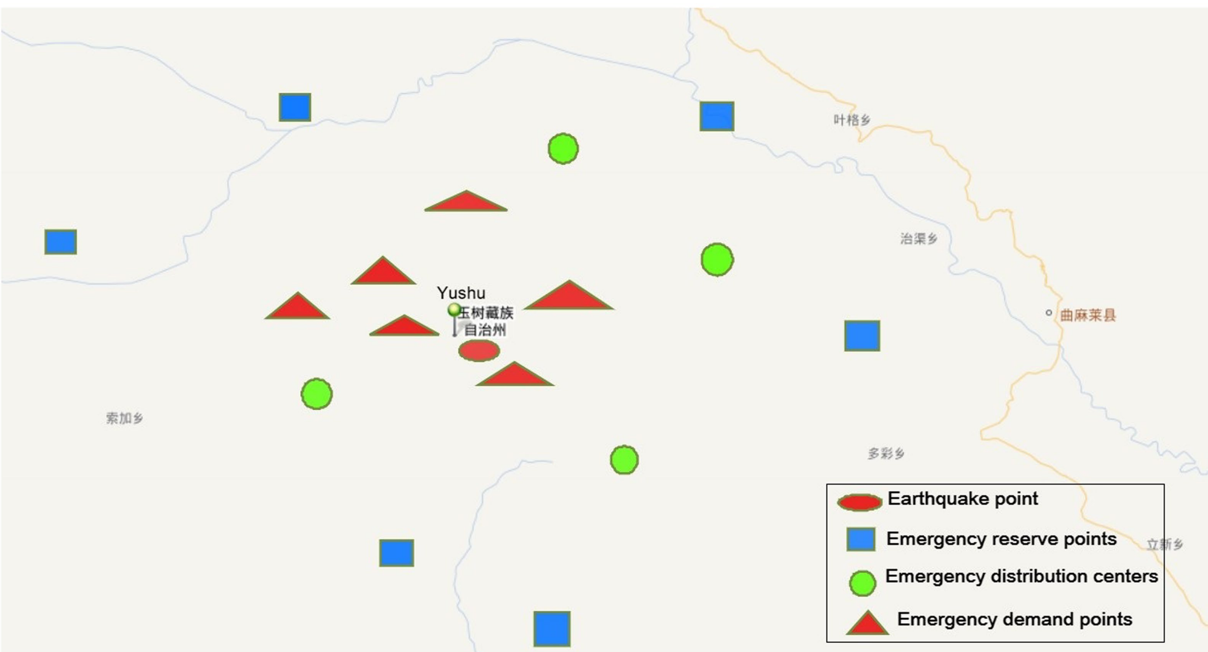


FIGURE 3
Regional disaster map.

TABLE 1 Fixed cost of emergency reserve points (unit: 10,000 CNY) and quantity of life necessities supplied at emergency reserve points (unit: tons).

R	$r1$	$r2$	$r3$	$r4$	$r5$	$r6$
(C_{trm})	1,000	600	3,000	1,000	2000	800
(S_{rm})	3,000	2000	5,000	3,000	4,000	1,500

TABLE 2 Fixed cost of the emergency distribution center (unit: 10,000 CNY).

(d)	$d1$	$d2$	$d3$	$d4$
(C_{tdm})	7,000	8,000	6,000	6,000

TABLE 3 Distribution distance from the emergency reserve point to the emergency distribution center (km).

R \ D	$d1$	$d2$	$d3$	$d4$
$r1$	30	23	25	26
$r2$	28	21	13	9
$r3$	17	14	15	16
$r4$	13	34	11	27
$r5$	24	12	18	33
$r6$	50	46	71	68

TABLE 4 Relevant values of emergency demand points.

(q)	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
d_{qm}	200	280	300	320	300	350
$\hat{E}_{qm}$	80	90	60	30	22	27
ξ_{qm}	0.0025	0.0021	0.0022	0.0029	0.0026	0.0023

TABLE 5 Transportation costs from the emergency reserve point to the emergency distribution center (10,000 CNY/per ton/km).

R \ D	$d1$	$d2$	$d3$	$d4$
$r1$	6	3	6	2
$r2$	2	5	8	9
$r3$	7	4	5	6
$r4$	3	4	1	7
$r5$	4	2	8	3
$r6$	3	8	2	9

TABLE 6 Transportation costs from the emergency distribution center to the emergency demand point (10,000 CNY/per ton/km).

D \ Q	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
$d1$	0.09	0.058	0.08	1.10	0.079	0.04
$d2$	0.08	1.18	0.02	0.097	1	0.087
$d3$	1.12	0.056	1.14	0.089	0.05	0.07
$d4$	0.085	0.09	0.079	1.08	0.03	0.06

TABLE 7 Road condition coefficient from the emergency distribution center to the emergency demand point.

D \ Q	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
$d1$	0.56	0.80	0.60	0.45	0.70	0.90
$d2$	0.60	0.37	0.96	0.52	0.48	0.62
$d3$	0.42	0.81	0.39	0.63	0.85	0.78
$d4$	0.65	0.56	0.70	0.46	0.91	0.82

TABLE 8 Travel time of vehicles from the emergency distribution center to the emergency demand point (h).

D \ Q	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
$d1$	0.5	0.7	0.6	0.4	0.7	0.8
$d2$	0.6	0.3	0.9	0.5	0.4	0.6
$d3$	0.4	0.9	0.3	0.6	0.7	0.7
$d4$	0.6	0.5	0.7	0.4	0.8	0.6

TABLE 9 Average speed of vehicles from the emergency distribution center to the emergency demand point (km/h).

D \ Q	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
$d1$	25	70	54	36	61	79
$d2$	52	18	88	44	38	56
$d3$	30	72	28	58	76	68
$d4$	59	48	61	37	82	74

TABLE 10 Model results: volume of the emergency reserve point to the emergency distribution center (unit: tons).

R \ D	$d1$	$d2$	$d3$	$d4$
$r1$	0	0	0	0
$r2$	0	0	0	0
$r3$	0	0	0	0
$r4$	0	0	2059	0
$r5$	0	0	0	0
$r6$	0	0	0	0

point to the $d3$ emergency distribution center and then distributed to six emergency demand points through the $d3$ emergency distribution center. The specific distribution routes are shown in Table 12.

To better verify the model's performance, given a value of 0.5 for the control parameter, measure the cost and location of the emergency distribution network for necessities with a 5, 10, and 30% disturbance coefficient. Comparing the target optimal values of the stochastic demand model, the relative robust model, and the absolute robust model, the model results are shown in Table 13.

Figure 4 compares the total target cost values with 5, 10, and 30% disturbance coefficients. As can be seen from the comparison of the total target cost values of the model in Figure 4, as the value of the disturbance coefficient increases, the total target cost value increases, and in which the absolute robust model objective value is maximized. The deviation range of objective function values between the relatively robust and stochastic demand models is less than 5%. Therefore, it is shown that the relatively robust model achieves the optimality of the target value under uncertain demand. The robustness goal of minimum cost of the emergency distribution network under demand disturbances is achieved.

TABLE 11 Model results: transportation volume from the emergency distribution center to the emergency demand point (unit: tons).

D \ Q	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
$d1$	0	0	0	0	0	0
$d2$	0	0	0	0	0	0
$d3$	280	370	360	350	322	377
$d4$	0	0	0	0	0	0

TABLE 12 Model results: emergency distribution routes.

Distribution nodes	Distribution routes					
	$d1$	$d2$	$d3$	$d4$		
$R \to D$	–	–	$r4 \to d3$	–		
$D \to Q$	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$
	$d3 \to q1$	$d3 \to q2$	$d3 \to q3$	$d3 \to q4$	$d3 \to q5$	$d3 \to q6$

TABLE 13 Model results.

Disturbance coefficient	Absolute robust model		Stochastic demand model		Relative robust model	
	Cost	Location	Cost	Location	Cost	Location
5%	37987.28	$(r4, d3)$	37417.41	$(r4, d3)$	37359.14	$(r4, d3)$
10%	39243.55	$(r4, d3)$	38603.81	$(r4, d3)$	37987.28	$(r4, d3)$
30%	44268.65	$(r4, d3)$	41649.43	$(r4, d3)$	40499.83	$(r4, d3)$

6. Conclusion

In the world, there is a high demand for necessities of life, with a wide variety of categories. With the development in social productivity and progress in science and technology, the scope of the necessities of life in the narrow sense is expanding. This study applies fractal theory to an emergency distribution network analysis framework and constructs a fractal-multi-stage-distribution network (FMDN) model under demand disturbance. The FMDN model includes multiple emergency reserve points, multiple emergency distribution centers, and multiple disaster points. Then, considering the actual situation of the emergency, the road condition coefficient and demand disturbance parameters are added to the FMDN model. The goal is to minimize the delivery cost of the FMDN model and solve it using LINGO software programming. Finally, a numerical experiment was solved and analyzed, and an optimal distribution path and a distribution volume were obtained. The feasibility of the model was verified.

It is important for the national and local emergency management departments to ensure the basic living and production needs of people in disaster areas are met. It is also important that the smooth implementation of emergency relief work through an efficient emergency distribution network be obtained to allocate the necessities of life to emergency demand points during large-scale emergencies. Whether life necessities can be allocated to disaster areas in time directly impacts the lives and property of those affected as well as the stability of the disaster area. We suggest that the central and local governments establish emergency reserve points based on the types of necessities of life. The government should consider the introduction of a minimum-quantity reserved system for emergency distribution centers, as well as formulating effective and feasible emergency distribution plans for essential commodities. Consideration should also be given to strengthening the existing road infrastructure of emergency distribution networks; establishing emergency linkage mechanisms across provinces, autonomous regions, and local departments to achieve inter-regional emergency linkage and collaboration; and improving the layout of emergency distribution multimodal transportation facilities to ensure the timeliness and stability of emergency distribution of life necessities in large-scale emergencies. This study provides a reference for government departments to deal with the distribution of life necessities during emergencies.

Only the allocation of emergency necessities with uncertain demand and road traffic conditions was considered in our model. We did not, therefore, take into account the perishability of life necessities, cold chain transportation, and other related factors. This should be considered in future research.

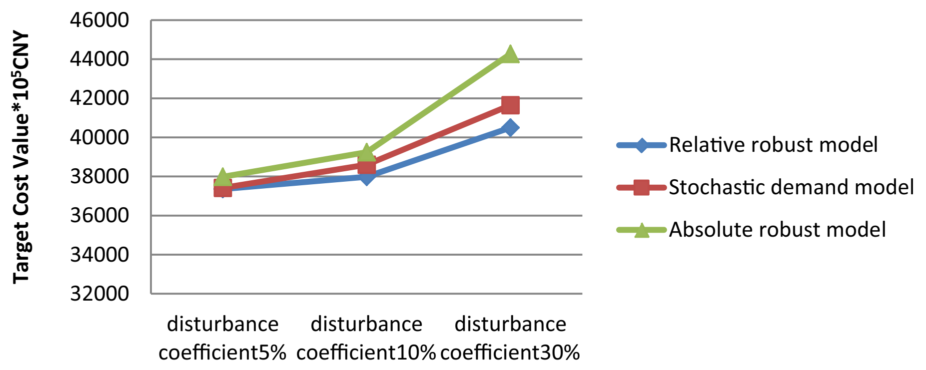


FIGURE 4
Total target cost comparison.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

CH: supervision. HL: conceptualization, formal analysis, methodology, writing the original draft, and writing, reviewing, and editing. ZZ: reviewing and editing. ZW and YN: data curation. HH and ZZ: validation. KT: visualization. HL and CH: funding acquisition. All authors contributed to the manuscript and approved the submitted version.

Funding

This study was supported by the Phased Research Results of the Gansu Provincial Philosophy and Social Science Planning Project (2022YB122). The study was supported by the Gansu Provincial Department of Education Innovation Fund Project (2023A-150) and Research Platform for the Knowledge Mechanism of China's Key Core Technological Innovation Ability Improvement on the Background of Dual Circulation (31513110812/108).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Yang N. Innovation and exploration of global economic governance mechanism: taking the construction of RCEP as an example. *Int Econ Tra Res.* (2020) 36:67. doi: 10.13687/j.cnki.gjjmts.2020.12.005

2. Pantazis C, Townsend P, Gordon D. *The necessities of life in Britain: poverty and social exclusion survey* (1999).

3. Pantazis C, Townsend P, Gordon D. *The necessities of life in Britain*. PSE working paper Bristol University (2000).

4. Pantazis C, Gordon D, Townsend P. The necessities of life. Working paper 1 of the poverty and social exclusion survey. *BMJ Brit Med J.* (2001) 1:185.

5. Pantazis C, Gordon D, Townsend P. (2006). *The necessities of life: poverty and social exclusion in Britain the Millennium Survey.*

6. Gordon D, Pantazis C, Townsend P. *Changing necessities of life, 1983–1999* Townsend Centre for International Poverty Research, University of Bristol (1999).

7. Fahmy E, Sutton E, Pemberton S. Are we all agreed? Consensual methods and the 'necessities of life' in the UK today. *J Soc Policy.* (2015) 44:591–610. doi: 10.1017/S0047279415000033

8. Teimoury E, Nedaei H, Ansari S, Sabbaghi M. A multi-objective analysis for import quota policy making in a perishable fruit and vegetable supply chain: a system dynamics approach. *Comput Electron Agric.* (2013) 93:37–45. doi: 10.1016/j.compag.2013.01.010

9. Brady TF. Emergency management: capability analysis of critical incident response. *Winter Simul Conf.* (2003) 2:1863–7.

10. Guo L, Zhu Y, Luo Z, Li W. The scaling of several public transport networks in China. *Fractals.* (2013) 21:350010. doi: 10.1142/S0218348X13500102

11. Bocewicz G, Banaszak Z, Nielsen I. Delivery-flow routing and scheduling subject to constraints imposed by vehicle flows in fractal-like networks. *Arch Cont Sci.* (2017) 27:135–50. doi: 10.1515/acsc-2017-0008

12. Saad SM, Bahadori R. Development of a dynamic information fractal framework to monitor and optimise sustainability in food distribution network. *Int J Food Eng.* (2020) 16:20180061. doi: 10.1515/ijfe-2018-0061

13. Webber E, Dunbar R. The fractal structure of communities of practice: implications for business organization. *PLoS One.* (2020) 15:e0232204. doi: 10.1371/journal.pone.0232204

14. Zhang Y, Li S-G. Multi-fractal analysis for vehicle distribution based on cellular automation model. *Mod Phys Lett B.* (2015) 29:550153. doi: 10.1142/S0217984915501535

15. Li X, Chen P, Chen F, Wang Z. Passenger flow analysis of Beijing urban rail transit network using fractal approach. *Mod Phys Lett B.* (2018) 32:1850001. doi: 10.1142/S021798491850001X

16. Man X, Chen Y. Fractal-based modeling and spatial analysis of urban form and growth: a case study of Shenzhen in China. *Int J Geoinf.* (2020):9. doi: 10.3390/ijgi9110672

17. Shao YH, Xu H, Liu YL, Xu HC. Multifractal behavior of cryptocurrencies before and during COVID-19. *Fractals*. (2021) 29:2150132. doi: 10.1142/S0218348X21501322

18. Mandelbrot B. The Pareto-levy law and the distribution of income. *Int Econ Rev*. (1960) 1:79–106. doi: 10.2307/2525289

19. Mandelbrot B. *Fractales, hasard et finance: 1959–1997* Flammarion (2009).

20. Ke JC, Duan Y, Xu C, Zhang Y. Cross-correlation multifractal analysis of technological innovation, financial market and real economy indices. *Fract Fract*. (2023) 7:267. doi: 10.3390/fractalfract7030267

21. Mandelbrot B, Goldenfeld N. Fractals and scaling in finance: discontinuity, concentration, risk. *Phys Today*. (1998) 51:59. doi: 10.1063/1.882342

22. Jiang Z, Xie W, Zhou W, Sornette D. Multifractal analysis of financial markets: a review. *Rep Prog Phys*. (2018) 82:12. doi: 10.1088/1361-6633/ab42fb

23. Jingwen L. Fractal modeling of logistics enterprise spatial organization. 2011 International Conference on E-business and E-government (ICEE), Shanghai.

24. Ryu K, Moon I, Oh S, Jung M. A fractal echelon approach for inventory management in supply chain networks. *Int J Prod Econ*. (2013) 143:316–26. doi: 10.1016/j.ijpe.2012.01.002

25. Yue YX, Zhang T, Yue QX. Improved fractal space filling curves hybrid optimization algorithm for vehicle routing problem. *Comput Intell Neurosci*. (2015) 2015:9. doi: 10.1155/2015/375163

26. Bi Y, Yuan HQ, Chang SH. Dynamic correlation analysis of regional logistics from the perspective of multifractal feature. *Fractals Compl Geom Patt Scal Nat Soc*. (2020) 8:2040015. doi: 10.1142/S0218348X20400150

27. Wu Z, Peng J. A chance-constrained model of logistics distribution center location under uncertain environment. *Adv Infor Sci Ser Sci*. (2014)

28. Wen ML, Yang Y, Qin ZF, Kang R. The capacitated facility location-allocation problem under uncertain environment. *J Intell Fuzz Syst*. (2015) 29:2217–26. doi: 10.3233/IFS-151697

29. Liu X, Peng J, Chen L. Uncertain programming model for location problem of multi-product logistics distribution centers. *Appl Math Sci*. (2015) 131:6543. doi: 10.12988/ams.2015.59597

30. Xu GC, Qi GLY. Vehicle routing problem for collaborative multidepot petrol replenishment under emergency conditions. *J Adv Transp*. (2021) 2021:20. doi: 10.1155/2021/5531500

31. Cui HX, Chen XY, Guo M, Jiao Y, Cao JD, Qiu JL. A distribution center location optimization model based on minimizing operating costs under uncertain demand with logistics node capacity scalability. *Phys A*. (2023) 610:13. doi: 10.1016/j.physa.2022.128392

32. Xing H, Jieru H, Jiang Z. How to build the fractal emergency organization under the disaster response for coastal cities of China? *Int J Disaster Risk Reduct*. (2023) 89:16. doi: 10.1016/j.ijdrr.2023.103628

33. Li A, Deng XQ, Zhao QH. Unconventional emergencies emergency coordination organization from the fractal perspective. *Syst Eng Theory Pract*. (2017) 37:937–48.

34. Li A, Deng XQ, Zhao QH. Dynamic reconstruction of emergency response organization for unconventional emergencies from a fractal perspective: taking the august 12 Tianjin port explosion as an example. *Manag Rev*. (2018) 28:193–206.

35. Mandelbrot BB. The fractal geometry of nature. *Am Phys*. (1998) 51:468. doi: 10.1119/1.13295

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 14 September 2023
DOI 10.3389/fpubh.2023.1220160

Healthcare professionals' and patients' assessments of listed mobile health apps in China: a qualitative study

PeiYu Liu[1], XueYun Li[1] and Xiao Man Zhang[2]*

[1]Department of Anesthesia and Operation Room, Nanjing Drum Tower Hospital, Nanjing, China,
[2]Department of Nursing, Xuzhou Medical University, Xuzhou, China

Check for updates

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Rüdiger Christoph Pryss,
Julius Maximilian University of Würzburg, Germany
Dinesh Gunasekeran,
National University of Singapore, Singapore

*CORRESPONDENCE
Xiao Man Zhang
✉ 229320852@qq.com

RECEIVED 10 May 2023
ACCEPTED 30 August 2023
PUBLISHED 14 September 2023

CITATION
Liu P, Li X and Zhang XM (2023) Healthcare professionals' and patients' assessments of listed mobile health apps in China: a qualitative study.
Front. Public Health 11:1220160.
doi: 10.3389/fpubh.2023.1220160

Background: In recent years, mobile health (mHealth) has gradually developed in China, and intelligent medicine has become an important research topic. However, there are still significant problems in mHealth applications (apps). Although healthcare professionals and patients are the main users, few studies have focused on their perceptions of the quality of mHealth apps.

Objective: This study aimed to (1) understand the respective perceptions of healthcare professionals and patients regarding mHealth apps, (2) assess what barriers exist that influence the user experience, and (3) explore how to improve the quality of mHealth apps and the development of the mHealth market in China. The study aims to promote the standardization of mHealth apps and provide effective information for the improvement and development of mHealth apps in the future.

Methods: Semistructured interviews with 9 patients and 14 healthcare professionals were conducted from January 2022 to April 2022 in the Affiliated Hospital of Xuzhou Medical University. The participants used mHealth apps for more than 3 months, including the "Good Mood" and "Peace and Safe Doctors" apps and apps developed by the hospital that were popular in China. Interview transcripts were analysed using thematic analysis.

Results: The following five themes were extracted: different concerns, hidden medical dangers, distance and insecurity, barriers for older people, and having positive perceptions of mHealth apps. Healthcare professionals prioritized simplicity in regard to mHealth apps, whereas patients rated effectiveness as the most crucial factor. The study also revealed several problems with mHealth apps, including insufficient information about physician qualifications, inaccurate medical content, nonstandard treatment processes, and unclear accountability, which led to a sense of distance and insecurity among participants. Older individuals faced additional obstacles when using mHealth apps. Despite these issues, the participants remained optimistic about the future of mHealth app development.

Conclusion: The utilization, advantages, and obstacles of mHealth applications for healthcare professionals and patients were explored through semistructured interviews. Despite the promising prospects for mHealth apps in China, numerous issues still need to be addressed. Enhancing the safety monitoring system and developing user-friendly mHealth apps for older adult patients are essential steps to bridge the gap between healthcare providers and patients.

KEYWORDS

patients, healthcare professionals, mobile health, quality, qualitative

Introduction

Background

In China, medical human resources are limited, and medical development is uneven. Mobile health (mHealth) is popular among Chinese residents as a convenient way to utilize equal medical resources (1). According to China's iiMedia Group, the number of users of mHealth apps in China rose from 151 million to 298 million in 2015 ~ 2016, an increase of 50.6% over the previous year (2). The COVID-19 pandemic of recent years has also facilitated further advancements in mHealth apps. Mhealth apps provide equal medical resources to patients in remote areas and have good cost-effectiveness (3).

As smart medical devices, mHealth applications (apps) involve smart sensors, display screens, database storage and other multifunctions. These apps can monitor a user's mental health, behavior, activity trajectory, and clinical data and remind users of correct behavior. Mhealth apps have been researched and developed in terms of pregnancy, smoking, asthma, pain, cancer, mental health, spirituality, and vision (4, 5).

Due to the continuous development of mHealth apps, the validity, acceptability, reliability, and quality of these apps need to be understood. To date, the evaluation of mHealth apps is mainly divided into clinical effectiveness (6–8), acceptability (9, 10), usability (9, 10), compliance (9), user experience (11, 12), and information quality (13, 14). Evaluation methods are also divided into three major categories: scales and questionnaires (15, 16), comparative experiments (6, 7), and qualitative interviews (9, 10). According to different evaluation methods, information and results from different perspectives can be obtained. Researchers often use comparative experiments to determine whether mHealth apps have improved regarding certain clinical indicators or cognition. Scales and questionnaires include the Mobile App Rating Scale (MARS) (17), the user version of the Mobile App Rating Scale (uMARS) (18), the System Usability Scale (SUS) (19), and self-designed questionnaires based on the Technology Acceptance Model (TAM) (20). The MARS and uMARS can be divided into four subscales (engagement, functionality, aesthetics, and information quality) and a qualitative scale that aims to examine users' feelings during the use process (17, 18). The MARS and SUS have strong reliability and good internal and external validity, as recognized by most researchers (19, 21–23).

Qualitative interviews are used by researchers to understand interviewees' internal and in-depth ideas. To investigate users' evaluations of mHealth apps, Anderson (12) conducted qualitative interviews with individuals residing in university towns in Australia to gain insights into their experiences with and expectations of self-management apps. Serafica and colleagues (11) studied the experience of mHealth apps among people in rural Hawaii. In addition, in previous studies, qualitative interviews have been conducted to explore British COPD patients' opinions on self-management apps (24), as well as to examine community health workers' perspectives and experiences regarding mHealth apps in Brazil (25). However, it is unclear what experiences people in mainland China, which has a large population and many ethnicities, have with popular mHealth apps. In an era of rapid internet development, what do these individuals think of mHealth apps in China? What are the shortcomings that need to be improved, and what important information has been overlooked by developers? These issues are worth discussing.

In recent years, research on mHealth apps has gradually increased, and the types of mHealth apps for various diseases have also increased. Chinese researchers have utilized the MARS and Silberg Scale to assess the quality of various types of mobile health apps, including psychological (26), cardiovascular (27), sleep management (28), and postpartum depression apps (29), in terms of usability, effectiveness, and acceptability. Due to poor supervision and other issues, the quality of mHealth apps can be inconsistent, which can lead to decreased usage by users (30). Additionally, China's vast population and diverse cultural differences among provinces and cities in different regions can make it difficult for mHealth apps to stand out in this crowded medical market (1).

In 1989, Davis proposed the TAM to explain users' acceptance of information technology. The TAM suggests that the use of a system is determined by behavioral intentions, and behavioral intentions are determined by perceived usefulness and perceived ease of use (31). The diffusion of innovations theory discusses how new ideas, new things, and new products are accepted by the public (32). This theory divides the diffusion process into five stages (awareness, persuasion, decision, application, and determination) and emphasizes that mass communication can effectively provide new information. However, interpersonal communication is more effective in changing people's attitudes and behaviors. According to these two theories, the experiences of healthcare professionals and patients with mHealth apps, as the main users, is particularly important. However, there are still gaps in the research on the experiences of healthcare professionals and patients in relation to mHealth apps in the Chinese healthcare market. To fill these knowledge gaps and address future software iterations, we conducted semistructured interviews with patients and healthcare professionals recruited from a tertiary grade A hospital. The purpose of this study was to (1) understand healthcare professionals' and patients' perceptions of the quality of mHealth apps; (2) evaluate the obstacles that affect users' quality experiences; and (3) explore how to improve the quality of mHealth apps. The results are intended to promote the standardization of mHealth apps and provide effective information for the improvement and development of mHealth apps in the future.

Methods

Design and setting

In this study, semistructured interviews were conducted at the Affiliated Hospital of Xuzhou Medical University, which is a general hospital in Xuzhou, Jiangsu Province, China. The Affiliated Hospital of Xuzhou Medical University was founded in 1897 and has a history of 123 years. In the first half of 2021, nearly 1.4 million patients received outpatient and emergency services at this hospital, with nearly 90,000 discharged patients. Although Xuzhou belongs to Jiangsu Province, as an important transportation hub city, it has close ties with Shandong, Anhui, Henan and other provinces. The hospital treats many patients. Xuzhou also has a long history of more than 6,000 years of civilization and 2,600 years of history, and the population is more than 8 million. The hospital has developed a series of medical treatment and management systems. Healthcare professionals working in this hospital have access to mHealth apps and use them frequently.

Abbreviations: mHealth, Mobile Health; App, Application; MARS, Mobile App Rating Scale; uMARS, User Version of the Mobile App Rating Scale; SUS, System Usability Scale.

Participants

By using purposive sampling and the principle of maximum differentiation, patients who visited this hospital and clinical frontline healthcare professionals who worked in this hospital from January to April 2022 were selected as participants for this study. The selection criteria for patients were as follows: patients who know what are mHealth apps; those who had used at least one mHealth app; those who had used the apps for at least 3 months; those aged ≥18 years; those with clear awareness and thinking that allowed good communication; and those who volunteered to participate in this study. The selection criteria for healthcare professionals were as follows: healthcare professionals who had worked on the clinical front line for ≥3 years; those who had used the apps for at least 3 months; those with good communication skills; and those who volunteered to participate in this study.

According to a literature analysis (33–35), the evaluation period of the majority of mHealth app intervention studies can be roughly divided into pre-and postintervention periods, with four treatment cycles (12 weeks) before and after the intervention. Based on the 21-day effect of behavioral development and a literature analysis, this study set a target of the use of mHealth apps for a minimum of 3 months to ensure that participants had a thorough understanding of the mHealth apps they used. The sample size was determined based on data saturation, which occurs when the information provided by interviewees becomes repetitive and no new topics emerge.

Interview outline

The study adopted a descriptive research method, utilizing interviews to gain a comprehensive understanding of the quality evaluation of mHealth apps from the perspectives of both patients and healthcare professionals. In accordance with the study's objectives, two nurses working in the pain department were chosen for preinterviews after a thorough review of the relevant literature and consultations with two members of the research team. According to the preinterview situation, an expert with rich research experience was consulted, and discussions and revisions were conducted to determine the final interview outline (Table 1).

Data collection

Researchers introduced the purpose and methods of this study to patients at the outpatient department. After obtaining their consent to participate, the researchers asked the patients if they knew what mHealth apps were. If the patients could name one or more apps, the researchers proceeded to the next question. If not, the patients were terminated from the study and did not proceed with the enrolment process. This approach ensured that only patients who had some knowledge of mHealth apps were included in the study. Then, researchers inquired about whether they had previously used a mHealth app, what the name of the app was, how long they had been using it, and how often they used it. The selection of healthcare professionals was based on communication with department managers, who selected individuals based on their work experience. Then, researchers asked questions about whether they used mHealth apps in their work, the amount of time spent using them, and the frequency of use per day. Based on these questions, individuals were deemed eligible for inclusion in the study if they met the inclusion criteria; otherwise, they were excluded.

Due to the long wait times of approximately 1–2 h, interviews with patients were conducted during the patient's wait time. After consulting with the department head nurse and doctors, the interviews with healthcare professionals were scheduled in the afternoon of a workday, with each healthcare professional invited and allotted 1 h of sufficient time for the interview. All interviews took place in a quiet classroom in the hospital. One-on-one, face-to-face interviews were conducted by a qualitative research-trained researcher. The entire interview process was recorded with a smartphone, and the key content and nonverbal content were recorded on paper. The interviews lasted approximately 30 min. When the researcher did not understand an answer, she repeated it to the participant to determine whether the meaning of the expression was accurate.

Data analysis

Within 24 h of the interview, the researcher transcribed the interview contents into written materials, and then another team member checked and ensured the accuracy of the transcript. After the transcription was completed, the interview subjects determined whether the transcription content accurately expressed their meaning or whether it needed to be modified. We used the standard method of thematic analysis to analyse the transcripts. Two researchers analysed the data independently. First, the researchers read the transcript in depth several times until they felt that they understood what the participants had said. Second, open coding was initially started line by line. Significant fragments were extracted, which were identified as "quotes," and assigned "codes." Third, the meaning was formulated for the codes, which were then organized into themes. Finally, the themes were further integrated into descriptions.

Ethical considerations

Ethical approval was obtained from the Affiliated Hospital of Xuzhou Medical University Ethics Review Board (Reference Number: XYFY2019-KL018). Information about the study was explained to all participants before the interviews, including information on research confidentiality and the participants were advised regarding their ability to withdraw from the study at any time during the interview. Participants were asked if they agreed to record the entire interview; if they did not want the interview recorded, a change was made immediately to manual recording. Finally, the participants signed written informed consent forms and volunteered to participate.

Results

Participant characteristics

A total of 14 healthcare professionals and 9 patients were interviewed. Their ages ranged from 24 to 57 years. The healthcare professionals had all worked in clinical first-line therapy for more than 3 years (Table 2). The patients were all from the area surrounding Xuzhou and were engaged in different industries, such as decoration

TABLE 1 Qualitative semi structured evaluation guide for healthcare professionals and patients.

Serial number	Healthcare Patients
①	Have you used any mHealth apps (if the mHealth app has a patient version)? Would you recommend it to your friends?
②	What do you think about the development of mHealth apps in the future?
③	For the mHealth app that you have used, what do you think are its shortcomings? What else do you think needs to be improved?
④	If you were to evaluate an mHealth app, what aspects/perspectives would you use? What indicators do you think are the most important for evaluation?

TABLE 2 Characteristics of the interviewed healthcare professionals.

Sociodemographics	Statistics
Profession, *n*	
Nurse	10
Physician	4
Age (years)	
Range	25
Mean	34.57
SD	7.07
Sex	
Male	5
Female	9
Professional experience (years)	
<5 years	2
5–10 years	7
10–15 years	3
>15 years	2
Frequency of mHealth app usage	
<3 times/day	2
>3 times/day	12

design, teaching, and homemaking, and different departments (Table 3). The mHealth apps used by the participants are presented in Appendix 1. These mHealth apps are frequently used in China and are popular based on the number of downloads.

Themes and subthemes

A total of five themes were identified: (1) different concerns; (2) hidden medical dangers; (3) distance and insecurity; (4) barriers for older people; and (5) having positive perceptions of mHealth apps.

Theme 1: Different concerns

In the interviews, 7 (7/14, 50%) healthcare professionals considered easy operation to be the focus of the evaluation (Table 4). A nurse shared her feelings about mHealth during her work:

> *Now with the electrocardiograph machine, it is a relatively easy operation. One key or two keys can be used; it is not complex. For example, for the current electrocardiogram, you press a key. If an app requires data to be filled in, it will be a challenge for the patients.* (M3)

Another nurse shared that her workload was increased by cumbersome mHealth operations:

> *A lot of patients ask us to help them, which indirectly increases our workload. We always have miscellaneous tasks every day.* (M4)

Furthermore, 6 (6/9, 67%) patients agreed that effectiveness was the most important part of their evaluation of mHealth apps. According to one patient,

> *The most important thing, I think, is that it's helpful, effective, and beneficial regarding my illness. If the software is not helping me, why do I want to use it?* (P4)

Theme 2: Hidden medical dangers

Subtheme 1: Insufficient information about qualifications

Four (4/9, 44%) interviewed patients said that the qualifications of the healthcare professionals were not included on the platform, and they could not objectively judge the professionals' medical levels. A patient described the awkwardness of using the app to select an obstetrics and gynaecology professional because of the unknown qualifications of the healthcare professionals in the mHealth app:

> *I selected a gynaecology professional on it; maybe I didn't see it clearly. I thought the doctor was a woman when I looked at the doctor's name at first. Then, I looked again—a man. I can't see a male doctor. I didn't want to…I didn't want him to check.* (P1)

Another patient believed that the authenticity of mHealth apps would be confusing if they did not detail the qualifications of healthcare professionals:

> *Every doctor's skill is really important. Now there are only two things to say about the software, such as what the doctor is, but not exactly. Of course, doctors with high grades are certainly relatively experienced. Users can also evaluate these doctors, not just show good evaluations. Some bad evaluations are blocked in some software, so we do not trust the authenticity of the software.* (P9)

Subtheme 2: Inaccurate medical content

Five (5/9, 56%) interviewed patients said that the health information provided by the mHealth apps did not have a detailed source description, and they could not judge the scientific nature and authenticity of the information. A mother shared her doubts about the information provided on a mother-to-child mHealth app during her pregnancy:

> *It's a bizarre statement that pregnant women can't eat this and that on this app. In the past, pregnant women could eat anything that a normal person could eat. I'm generally selective about what's on this app.* (P2)

A patient who came for a medical examination said that the low authenticity of the content provided by the mHealth app would affect his compliance:

> *If the authenticity is low, I don't believe what it is providing, so my compliance is low, and even if it's right, I' m still reluctant to follow it.* (P4)

Furthermore, a patient said that due to poor authenticity, he selectively used mHealth apps; if he had problems, he still sought a doctor for the first time:

> *Some of the online content is right or wrong … even though I know that some health care knowledge in the app is good, I still mainly seek a doctor's advice.* (P9)

TABLE 3 Characteristics of the interviewed patients.

Sociodemographics	Statistics
Age (years)	
Range	33
Mean	35.88
SD	10.57
Sex	
Male	3
Female	6
Educational Level	
Junior high school	2
High school diploma or equivalent	3
College	4
Residence	
Rural	5
Urban	4
Frequency of mHealth app usage	
<3 times/day	7
>3 times/day	2

Subtheme 3: Nonstandard treatment processes

A standardized medical treatment process can effectively reduce the incidence of medical accidents and disputes (36). Standardized processes not only reduce human injury for patients but also protect healthcare professionals. Two doctors (2/14, 14%) reported that the medical procedures of an mHealth app were not standardized, which could lead to medical errors:

> *The medical software needs to be more rigorous to meet the basic requirements of the National Health Commission and to ensure safety. All processes and operations should be scientific, regular and secure. It's a big taboo to prescribe prescription drugs and diagnose patients. We can' t afford the consequences of misdiagnosis. We can't tell if there's a problem. We can only give patients some medical advice or health measures they can take.* (M7)
>
> *I personally believe that the regulation of smartphone medical software is relatively loose. The entire process is not perfect.* (M8)

Subtheme 4: Unclear accountability

According to two participants (2/23, 9%), the responsibility of mHealth apps is unclear, there are hidden dangers, and it is difficult to determine responsibility.

TABLE 4 Focus of mHealth app quality assessments by healthcare professionals and patients.

	Healthcare professionals		Patients	
Encoding	Number of persons	Percentage	Number of persons	Percentage
Easy operation	7	50%	2	22%
Effectiveness	2	14%	6	67%
Function	2	14%	2	22%
Accessibility	3	21%	1	11%
Security	6	43%	5	56%
Payment	2	14%	2	22%

> *I don't know who I can look for if there's something wrong with pills.* (P5)

> *Once the mistake and the inevitable consequences have taken place, who should be responsible for them? This cannot be determined clearly. There are also problems in the hospital that are unclear. Now there are too many doctor–patient disputes.* (M7)

Theme 3: Distance and insecurity

Subtheme 1: Blunt expression

The counseling function of mHealth apps is mainly in the form of text input and output (37), and healthcare professionals cannot observe the body language of patients. In addition, patients are prone to doubt the responses of healthcare professionals, which produces a sense of distance and insecurity. Four participants (4/23, 17%) said mHealth apps display too much text, less vivid dynamic graphics and less videos and are difficult to understand. A nurse with rich clinical experience believed that presentation can not only close the distance between nurses and patients and improve patient compliance but also meet the needs of patients at all cultural levels:

> *From the patients' point of view, if you recommend mHealth apps well, they'll see more. Like the rich form I've just said: if every health education app has video education, it is convenient for the patient. You'll be able to close the distance from the patient, and improve their compliance, cooperation with treatment or rehabilitation. In addition, there are also cultural levels and regional differences. For example, a person with a high level of education, you go to tell them the information is not as good as their own. However, patients with a low educational level or patients from some rural regions want you to say it, write it down, rather than taking their own initiative to learn.* (M4)

In addition, some nurses suggested that more dynamic videos could be made:

> There can be some health education videos … dynamic. (M12)

A patient wanted to be provided with video calls and face-to-face communication:

> *How did the doctor judge which situation I was in? It would be best to have a video to communicate. In addition, I wasn't very good at typing. I feel a little bit relieved to see the doctor himself in the video, at least feel zero distance.* (P9)

Subtheme 2: Generating medical virtuality

With the development of networks and technology, intelligent virtual agents have permeated various fields, such as medicine, teaching, work and life, promoting the development of society (38). However, because of the particularity of the target population of mHealth apps, they often cause insecurity for users:

> *Like apps, many times the consulting doctor is not local. If a problem emerges, I do not where I can go to find them.* (P4)

Another patient said the medical software led to a sense of distance:

> *I have a distant feeling in the app that I'm not talking to real people and that sometimes it's the robot that automatically replies. It's not as good as talking to a doctor in person. I can also understand the doctor's level and the medical level of the hospital.* (P2)

Subtheme 3: Lack of humanistic care

Mhealth apps can expand the medical network, connect patients from multiple fields, and share medical resources equally. Although communication can increase peer support, humanistic care and cross-cultural differences are ignored. Users receive cold and mechanized responses. A patient believed he could receive warm care from paramedics in the hospital, but the app only responded automatically:

> *I received polite replies (on the app), but they were cold and official. Although doctors also have a bad attitude in the hospital, there are doctors and nurses in the hospital every day to check in and ask you how you slept last night. After you leave the hospital, the nurse will follow up on your recent situation. However, it's over when you're done consulting on the app, and no one cares about your follow-up.* (P5)

> *I don't feel like anyone cares about me. A smart device can't do what doctors and nurses do, and no one cares what danger will happen to you. In addition, the speed of the response of apps is too slow, which made me feel I am not important ….* (P7)

Theme 4: Barriers for older people

Subtheme 1: Terminology and professionalization

The main difference between mHealth apps and other apps lies in the purpose of mHealth app use and the particularity of the population (25). Medical terminology and medical professionalism are also unique to mHealth apps. Due to degenerative changes in the brain and organs of older people, their ability to perceive, think, remember and understand is weakened. They may feel a certain degree of psychological resistance before using medical apps, thus reducing their desire to use these apps (39). A doctor shared his experience in recommending medical software to older patients:

> *When I recommended the 'Good Mood' app to the patients, I found that the older patients rejected it directly and said that they did not understand it when I told them they would have to use a smartphone.* (M7)

Because of late contact with smart devices and poor understanding in China, older people tend to resist mHealth apps:

> *Due to the particularity of the work, we connect the older patients or children. Older patients have a poor understanding, and their generation has a low level of education and medical knowledge, even with oral explanations. They may not understand, let alone use the app.* (M3)

Subtheme 2: Unfriendly technology and design

A small font size and cumbersome operations seriously affect older patients' experiences of mHealth apps, which are also common problems with smartphone apps. A 56-year-old patient had several operational problems during the use of registered services:

> *Apps have to be downloaded and constantly remind you to update, which is especially complex. I have to register a service repeatedly. My patience wears out. I have no confidence in myself.* (P9)

Healthcare professionals and individuals who cohabitate with older relatives frequently encounter senior citizens who lack proficiency in utilizing mHealth apps. Older people come to seek their help. However, they said that this indirectly increases their burden:

> *It doesn't feel good to use it because when an older patient comes, the process of linking a card and a bank card can be slightly more cumbersome. He will not do it, so we help him.* (M2)
>
> *I have two old people in my family. Some software is said to be used for elderly individuals, but older people cannot see clearly and do not know how to use it. Well, we as their children will help them, but we also have our own things that keep us busy.* (P2)

Theme 5: Having positive perceptions of mHealth apps

In the interviews, the healthcare professionals and patients had a positive attitude as the main users of mHealth apps. While participants criticized their drawbacks, they looked forward to future mHealth apps to bring new changes to society. For example, although 23 participants noted that the current mHealth apps have many quality problems that seriously affect their use experience, they answered positively when asked whether they would recommend these apps to others. They believed that the future medical model would reach a new platform if developers could effectively improve the quality of mHealth apps:

> *As young people grow older, they will prefer to use smartphone apps rather than go to the community to do these things. So I suggest that when the community can promote this, and then young people use these apps for a long time, your medical information is equivalent to filing in this community, which is actually good for the later development of the community and can reduce the waste of medical resources in big hospitals.* (M7)

Discussion

Principal findings

This study was conducted during the implementation of lockdown regulations in China, a time when the use of mHealth or telemedicine provided a novel approach to patient care. The purpose of this study was to examine the experiences of healthcare professionals and patients who used mHealth apps and to identify the challenges they faced while using these apps. The primary findings indicated that there are several issues related to the quality of mHealth apps, although users still recognize the benefits of these apps. This may be attributed to the limited access to hospital-based medical care during the COVID-19 lockdown period in China, where patients relied on mHealth apps to receive timely and effective medical attention from healthcare professionals.

The participants were drawn from a range of clinical areas, including pain management, oncology, psychiatry, thymic surgery, general surgery, obstetrics and gynaecology, and physical examination centres. Based on their own usage experiences, the participants demonstrated the most frequently utilized features, such as appointment scheduling and registration, the dissemination of health information, and online consultations. These findings are in line with the results reported by Anderson (12) and Wu (40). MHealth apps can provide various benefits to patients, such as saving time, enhancing the efficiency of medical treatment, and reducing discomfort and anxiety. During the period when lockdown regulations were implemented in China to combat the COVID-19 pandemic, patients demonstrated a preference for using mHealth apps for appointment scheduling or consultation purposes, thereby reducing the time and opportunities spent in social contact with others.

Although the nine patients had different types of diseases, the interviews did not reveal any differences in their experiences using mHealth apps based on their specific medical conditions. It is possible that the participants had relatively mild or subhealth conditions, which may have influenced the results. However, according to Wang's systematic review (41), the effectiveness of mHealth functions can vary depending on the specific type of disease being treated. A meta-analysis (42) showed that mHealth apps had no significant effect on drug compliance for patients with various chronic diseases. Further research is needed to determine whether different types of chronic diseases have varying effects on each function and module of mHealth apps.

It was found that Chinese patients prioritized the effectiveness of mHealth apps in managing their illnesses, whereas Chinese healthcare professionals favored apps with user-friendly interfaces and optimized designs. However, Jiang et al. found that healthcare professionals were primarily interested in the economic benefits of mHealth apps in reducing healthcare costs (43). Xiang's study found that Chinese healthcare professionals prioritized the ability of artificial intelligence to reduce repetitive workload, and nonhealthcare personnel valued efficiency in diagnosis and treatment. Both groups of participants reported that artificial intelligence does not pose any safety risks (44). In contrast, this study found that both Chinese healthcare professionals and patients perceive mHealth as having safety issues. Furthermore, we propose that the government needs to establish strict procedural guidelines for the development and regulation of mHealth. These findings suggest that there is a need for more comprehensive research on the impact of mHealth on patient safety, as well as for the development of appropriate regulatory frameworks to ensure the safe and effective use of this technology (45). Generally, this study highlighted the importance of considering the perspectives of different stakeholders when designing mHealth apps.

During the interviews, healthcare professionals expressed concerns about the burden placed on their work by older patients who are unable to use mHealth apps. With the implementation of China's three-child policy, the number of female clinical doctors and nurses on maternity leave has increased, resulting in healthcare professionals having to assist patients with basic tasks and teach them how to use mHealth apps. This has undoubtedly increased the workload of healthcare professionals. These findings highlight the need for medical institutions to employ dedicated internet staff or information nurses to address these issues. This study provides valuable insights into the underresearched area of user perspectives.

Healthcare professionals also faced challenges when assisting older patients who lacked necessary information such as call numbers and personal identification information required for verification. This issue may be related to the lower education levels of contemporary older individuals (46) or the larger number of hospitalized patients in nearby rural areas. These findings are consistent with those of Raghunathan's study (47). Older individuals aged over 80 years were less willing to use mHealth apps, and those with higher levels of education were more accepting of these apps than those with lower levels of education (48). Therefore, mHealth app designs should not only cater to young and middle-aged individuals but also focus on the needs of older users.

It was found that the interviewed patients perceived mHealth apps as lacking in care, which could potentially strain the doctor–patient relationship. Furthermore, healthcare professionals in China face challenges due to heavy workloads and low wages (49, 50). Simpkin et al. (51) explored the perspectives of medical students on intelligent healthcare and gained insights into how to integrate technology effectively and consciously into the medical field to enhance empathy and compassionate care. To address the concerns raised by patients, a feedback function could be added to the platform, mHealth apps could be developed with a patient-centred approach (48), and humanistic theory could be applied to intelligent healthcare.

The quality issues of mHealth have also been emphasized by international organizations and researchers from other countries. To assist developers in evaluating the quality of apps, several tools have been developed (52). A systematic review of 87 studies published through 2018 revealed that researchers developed 48 different rating scales for evaluating the availability and quality of mHealth apps (52). These rating scales primarily focused on content quality and usability. This study uses open-ended questions to overcome these rating frameworks. The findings indicated that there were quality issues in mHealth apps from various aspects, including security risks, technical issues, design, and humanistic concerns. International organizations considered the entire quality regulatory issue from a larger perspective, accounting for the interests of multiple parties (53). They believed that improving the efficiency of mHealth regulation should emphasize process transparency rather than imposing accountability (54). This study concluded that healthcare professionals and patients considered the need for more standardized accountability to clarify their own work responsibilities or better protect their individual rights.

Implications for future research and clinical practice

The participants of this study identified various barriers in the areas of process, regulation, technology and humanistic. Additionally, they found that mHealth apps increase workloads for both healthcare professionals and individuals who cohabitate with older relatives to a certain extent. Drawing on Shachak's research findings (55), future research directions in health information technology implementation should not only focus on users' social psychological issues and usage behavior but also consider technology, processes, contexts, and users as a dynamic interactive system. In a multicultural, multireligious, and multiethnic Asian country with a predominantly older population, medical humanities can help patients feel respected and cared for. China is a country that values family-oriented traditions, and the loneliness experienced by older individuals who live alone can easily lead to mental illness. MHealth apps can provide these individuals with more medical and social support. Future studies should focus on continuously improving the quality of mHealth apps while considering how to better integrate humanistic care with scientific and technological advancements.

Strengths and limitations

The primary strength of our study lies in the use of qualitative methods to explore the experiences of healthcare professionals and patients with mHealth apps, as well as the factors influencing the continued use of mHealth apps by users in the Chinese healthcare market. However, a major limitation of our study is the small sample size, limited to individuals from XuZhou, the results are not representative of the whole of China, and the fact that only the views of healthcare professionals and young patients with older patients around them were considered. Although the selected mHealth apps have a high download frequency in the market, these apps could not represent all mHealth apps in China. To address this limitation in future research, we can expand the sample size and increase the diversity of the sample by including more technicians and related personnel in drug enterprises. Additionally, importantly, the participants in our study were homogeneous in terms of ethnicity; therefore, our findings may not accurately represent minority groups outside this demographic background.

Conclusion

This study aimed to explore the perceptions of individuals regarding mHealth apps in China. Healthcare professionals and patients expressed optimism about the potential of mHealth apps but also highlighted several concerns and potential risks associated with their quality. Thus, it is crucial to enhance the safety and regulatory frameworks surrounding mHealth apps and to develop user-friendly mHealth apps specifically designed for older patients, which will help bridge the gap between healthcare professionals and patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

PL contributed to the study concept and design. PL and XL contributed to the acquisition of data, data analysis, interpretation of data, and drafting of the manuscript. XZ contributed to the interpretation of data, and critical revision of the manuscript and final approval of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This study was funded by the University Research Board of Xuzhou Medical University (grant number 2018KJ07) and the Xuzhou Science and Technology Bureau (grant number KC19223).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1220160/full#supplementary-material

References

1. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using Mobile apps for health management: a new health care mode in China. *JMIR Mhealth Uhealth.* (2019) 7:e10299. doi: 10.2196/10299

2. iiMedia. (2017). *IiMedia group*. Available at: www.iimedia.cn/c400/49397.html.

3. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. *Telemed J E Health.* (2015) 21:81–5. doi: 10.1089/tmj.2014.0053

4. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. *J Telemed Telecare.* (2018) 24:22–30. doi: 10.1177/1357633X16673538

5. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a Meta-analysis. *JAMA Pediatr.* (2017) 171:461–9. doi: 10.1001/jamapediatrics.2017.0042

6. Desveaux L, Agarwal P, Shaw J, Hensel JM, Mukerji G, Onabajo N, et al. A randomized wait-list control trial to evaluate the impact of a mobile application to improve self-management of individuals with type 2 diabetes: a study protocol. *BMC Med Inform Decis Mak.* (2016) 16:144. doi: 10.1186/s12911-016-0381-5

7. Elbers S, Pool J, Wittink H, Köke A, Scheffer E, Smeets R. Mobile health app (AGRIPPA) to prevent relapse after successful interdisciplinary treatment for patients with chronic pain: protocol for a randomized controlled trial. *JMIR Res Protoc.* (2020) 9:e18632. doi: 10.2196/18632

8. Jiwani R, Wang J, Berndt A, Ramaswamy P, Mathew Joseph N, Du Y, et al. Changes in patient-reported outcome measures with a technology-supported behavioral lifestyle intervention among patients with type 2 diabetes: pilot randomized controlled clinical trial. *JMIR Diabet.* (2020) 5:e19268. doi: 10.2196/19268

9. Giunti G, Rivera-Romero O, Kool J, Bansi J, Sevillano JL, Granja-Dominguez A, et al. Evaluation of more stamina, a Mobile app for fatigue Management in Persons with multiple sclerosis: protocol for a feasibility, acceptability, and usability study. *JMIR Res Protoc.* (2020) 9:e18196. doi: 10.2196/18196

10. Langlet B, Maramis C, Diou C, Maglaveras N, Fagerberg P, Heimeier R, et al. Formative evaluation of a smartphone app for monitoring daily meal distribution and food selection in adolescents: acceptability and usability study. *JMIR Mhealth Uhealth.* (2020) 8:e14778. doi: 10.2196/14778

11. Serafica R, Inouye J, Lukkahatai N, Braginsky N, Pacheco M, Daub KF. The use of mobile health to assist self-management and access to services in a rural community. *Comput Inform Nurs.* (2019) 37:62–72. doi: 10.1097/CIN.0000000000000494

12. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. *PLoS One.* (2016) 11:e0156164. doi: 10.1371/journal.pone.0156164

13. Jeon E, Park HA, Min YH, Kim HY. Analysis of the information quality of korean obesity-management smartphone applications. *Healthc Inform Res.* (2014) 20:23–9. doi: 10.4258/hir.2014.20.1.23

14. Xie B, Su Z, Zhang W, Cai R. Chinese cardiovascular disease Mobile Apps' information types, information quality, and interactive functions for self-management: systematic review. *JMIR Mhealth Uhealth.* (2017) 5:e195. doi: 10.2196/mhealth.8549

15. Erfannia L, Barman MP, Hussain S, Barati R, Arji G. How mobile health affects primary healthcare? Questionnaire design and attitude assessment. *Digit Health.* (2020) 6:2055207620942357. doi: 10.1177/2055207620942357

16. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. *Disabil Rehabil Assist Technol.* (2021) 16:712–21. doi: 10.1080/17483107.2019.1701103

17. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth.* (2015) 3:e27. doi: 10.2196/mhealth.3422

18. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile application rating scale (uMARS). *JMIR Mhealth Uhealth.* (2016) 4:e72. doi: 10.2196/mhealth.5849

19. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. *J Med Internet Res.* (2014) 16:e89. doi: 10.2196/jmir.2951

20. Shemesh T, Barnoy S. Assessment of the intention to use Mobile health applications using a technology acceptance model in an Israeli adult population. *Telemed J E Health.* (2020) 26:1141–49. doi: 10.1089/tmj.2019.0144

21. LeBeau K, Huey LG, Hart M. Assessing the quality of Mobile apps used by occupational therapists: evaluation using the user version of the Mobile application rating scale. *JMIR Mhealth Uhealth.* (2019) 7:e13019. doi: 10.2196/13019

22. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of Mobile phone apps for weight management: user-centered study with employees from a Lebanese University. *JMIR Mhealth Uhealth*. (2019) 7:e9836. doi: 10.2196/mhealth.9836

23. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable sensor-based exercise biofeedback for Orthopaedic rehabilitation: a mixed methods user evaluation of a prototype system. *Sensors*. (2019) 19:432. doi: 10.3390/s19020432

24. Williams V, Price J, Hardinge M, Tarassenko L, Farmer A. Using a mobile health application to support self-management in COPD: a qualitative study. *Br J Gen Pract*. (2014) 64:e392–400. doi: 10.3399/bjgp14X680473

25. Schoen J, Mallett JW, Grossman-Kahn R, Brentani A, Kaselitz E, Heisler M. Perspectives and experiences of community health workers in Brazilian primary care centers using m-health tools in home visits with community members. *Hum Resour Health*. (2017) 15:71. doi: 10.1186/s12960-017-0245-9

26. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. *JMIR Mhealth Uhealth*. (2019) 7:e13236. doi: 10.2196/13236

27. Xiao Q, Wang Y, Sun L, Lu S, Wu Y. Current status and quality assessment of cardiovascular diseases related smartphone apps in China. *Stud Health Technol Inform*. (2016) 225:1030–1. doi: 10.3233/978-1-61499-658-3-1030

28. Xu ZF, Luo X, Shi J, Lai Y. Quality analysis of smart phone sleep apps in China: can apps be used to conveniently screen for obstructive sleep apnea at home? *BMC Med Inform Decis Mak*. (2019) 19:224. doi: 10.1186/s12911-019-0916-7

29. Li Y, Zhao Q, Cross WM, Chen J, Qin C, Sun M. Assessing the quality of mobile applications targeting postpartum depression in China. *Int J Ment Health Nurs*. (2020) 29:772–85. doi: 10.1111/inm.12713

30. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). *Implement Sci*. (2016) 11:146. doi: 10.1186/s13012-016-0510-7

31. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. *Appl Clin Inform*. (2018) 9:604–34. doi: 10.1055/s-0038-1668091

32. Rogers EM. *Diffusion of innovations. 4th* ed. New York, NY: The Free Press (1995).

33. Berry JD, Paganoni S, Carlson K, Burke K, Weber H, Staples P, et al. Design and results of a smartphone-based digital phenotyping study to quantify ALS progression. *Ann Clin Transl Neurol*. (2019) 6:873–81. doi: 10.1002/acn3.770

34. Lapierre MA, Zhao P, Custer BE. Short-term longitudinal relationships between smartphone use/dependency and psychological well-being among late adolescents. *J Adolesc Health*. (2019) 65:607–12. doi: 10.1016/j.jadohealth.2019.06.001

35. Recio-Rodríguez JI, Lugones-Sanchez C, Agudo-Conde C, González-Sánchez J, Tamayo-Morales O, Gonzalez-Sanchez S, et al. Combined use of smartphone and smartband technology in the improvement of lifestyles in the adult population over 65 years: study protocol for a randomized clinical trial (EVIDENT-age study). *BMC Geriatr*. (2019) 19:19. doi: 10.1186/s12877-019-1037-y

36. Parks-Savage A, Archer L, Newton H, Wheeler E, Huband SR. Prevention of medical errors and malpractice: is creating resilience in physicians part of the answer? *Int J Law Psychiatry*. (2018) 60:35–9. doi: 10.1016/j.ijlp.2018.07.003

37. Alwashmi MF, Fitzpatrick B, Davis E, Gamble JM, Farrell J, Hawboldt J. Perceptions of health care providers regarding a Mobile health intervention to manage chronic obstructive pulmonary disease: qualitative study. *JMIR Mhealth Uhealth*. (2019) 7:e13950. doi: 10.2196/13950

38. Bhavnani SP, Sitapati AM. Virtual care 2.0-a vision for the future of data-driven technology-enabled healthcare. *Curr Treat Options Cardiovasc Med*. (2019) 21:21. doi: 10.1007/s11936-019-0727-2

39. Galambos C, Rantz M, Craver A, Bongiorno M, Pelts M, Holik AJ, et al. Living with intelligent sensors: older adult and family member perceptions. *Comput Inform Nurs*. (2019) 37:615–27. doi: 10.1097/CIN.0000000000000555

40. Wu X, Chen J, Yun D, Yuan M, Liu N, Yan P, et al. Effectiveness of an ophthalmic hospital-based virtual service during the COVID-19 pandemic. *Ophthalmology*. (2021) 128:942–5. doi: 10.1016/j.ophtha.2020.10.012

41. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. *Telemed J E Health*. (2014) 20:570–83. doi: 10.1089/tmj.2013.0243

42. Peng Y, Wang H, Fang Q, Xie L, Shu L, Sun W, et al. Effectiveness of Mobile applications on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. *J Manag Care Spec Pharm*. (2020) 26:550–61. doi: 10.18553/jmcp.2020.26.4.550

43. Jiang Y, Sun P, Chen Z, Guo J, Wang S, Liu F, et al. Patients' and healthcare providers' perceptions and experiences of telehealth use and online health information use in chronic disease management for older patients with chronic obstructive pulmonary disease: a qualitative study. *BMC Geriatr*. (2022) 22:9. doi: 10.1186/s12877-021-02702-z

44. Xiang Y, Zhao L, Liu Z, Wu X, Chen J, Long E, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. *Artif Intell Med*. (2020) 102:101780. doi: 10.1016/j.artmed.2019.101780

45. Food and drug administration, *Policy for device software functions and Mobile medical applications* (2022). Available at: www.fda.gov/media/80958/download

46. Jiao J, Wang Y, Zhu C, Li F, Zhu M, Wen X, et al. Prevalence and associated factors for frailty among elder patients in China: a multicentre cross-sectional study. *BMC Geriatr*. (2020) 20:100. doi: 10.1186/s12877-020-1496-1

47. Raghunathan NJ, Korenstein D, Li QS, Tonorezos ES, Mao JJ. Determinants of mobile technology use and smartphone application interest in cancer patients. *Cancer Med*. (2018) 7:5812–9. doi: 10.1002/cam4.1660

48. Zhang L, Babu SV, Jindal M, Williams JE, Gimbel RW. A patient-centered Mobile phone app (iHeartU) with a virtual human assistant for self-Management of Heart Failure: protocol for a usability assessment study. *JMIR Res Protoc*. (2019) 8:e13502. doi: 10.2196/13502

49. Ma S, Xu X, Trigo V, Ramalho NJ. Doctor-patient relationships (DPR) in China. *J Health Organ Manag*. (2017) 31:110–24. doi: 10.1108/JHOM-09-2016-0165

50. Yu D, Ma Y, Sun Q, Lu G, Xu P. A nursing care classification system for assessing workload and determining optimal nurse staffing in a teaching hospital in China: a pre-post intervention study. *Int J Nurs Pract*. (2015) 21:339–49. doi: 10.1111/ijn.12295

51. Simpkin AL, Dinardo PB, Pine E, Gaufberg E. Reconciling technology and humanistic care: lessons from the next generation of physicians. *Med Teach*. (2017) 39:430–5. doi: 10.1080/0142159X.2017.1270434

52. Grundy Q. A review of the quality and impact of Mobile health apps. *Annu Rev Public Health*. (2022) 43:117–34. doi: 10.1146/annurev-publhealth-052020-103738

53. Grundy Q, Held F, Bero L. A social network analysis of the financial links backing health and fitness apps. *Am J Public Health*. (2017) 107:1783–8. doi: 10.2105/AJPH.2017.303995

54. Marelli L, Lievevrouw E, Van Hoyweghen I. Fit for purpose? The GDPR and the governance of European digital health. *Policy Stud*. (2020) 41:447–67. doi: 10.1080/01442872.2020.1724929

55. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. *J Biomed Inform*. (2019) 100:103315. doi: 10.1016/j.jbi.2019.103315

56. Kell K, Aymerich MA, Horn V. FDI-Unilever Brush day & Night partnership: 12 years of improving behaviour for better oral health. *Int Dent J*. (2018) 68:3–6. doi: 10.1111/idj.12404

Appendix 1

MHealth apps used by participants and their functions and basic information.

frontiers | Frontiers in Public Health

TYPE Systematic Review
PUBLISHED 15 September 2023
DOI 10.3389/fpubh.2023.1229386

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Patience Usip,
University of Uyo, Nigeria
Yan Wang,
Amazon, United States
Mounir Hemam,
University of Khenchela, Algeria

*CORRESPONDENCE
Olena Vovk
✉ olena.b.vovk@lpnu.ua

RECEIVED 26 May 2023
ACCEPTED 23 August 2023
PUBLISHED 15 September 2023

CITATION
Singh Y, Jabbar MA, Kumar Shandilya S, Vovk O and Hnatiuk Y (2023) Exploring applications of blockchain in healthcare: road map and future directions. *Front. Public Health* 11:1229386. doi: 10.3389/fpubh.2023.1229386

Exploring applications of blockchain in healthcare: road map and future directions

Yuvraj Singh[1], M. A. Jabbar[2], Shishir Kumar Shandilya[1], Olena Vovk[3]* and Yaroslav Hnatiuk[3]

[1]School of Computing Science and Engineering, VIT Bhopal University, Bhopal, Madhya Pradesh, India, [2]Department of Computer Science and Engineering (AI&ML), Vardhaman College of Engineering, Hyderabad, Telangana, India, [3]Department of Artificial Intelligence Systems, Lviv Polytechnic National University, Lviv, Oblast, Ukraine

Blockchain technology includes numerous elements such as distributed ledgers, decentralization, authenticity, privacy, and immutability. It has progressed past the hype to find actual use cases in industries like healthcare. Blockchain is an emerging area that relies on a consensus algorithm and the idea of a digitally distributed ledger to eliminate any intermediary risks. By enabling them to trace data provenance and any changes made, blockchain technology can enable different healthcare stakeholders to share access to their networks without violating data security and integrity. The healthcare industry faces challenges like fragmented data, security and privacy concerns, and interoperability issues. Blockchain technology offers potential solutions by ensuring secure, tamper-proof storage across multiple network nodes, improving interoperability and patient privacy. Encrypting patient data further enhances security and reduces unauthorized access concerns. Blockchain technology, deployed over the Internet, can potentially use the current healthcare data by using a patient-centric approach and removing the intermediaries. This paper discusses the effective utilization of blockchain technology in the healthcare industry. In contrast to other applications, the exoteric evaluation in this paper shows that the innovative technology called blockchain technology has a major role to play in the existing and future applications of the healthcare industry and has significant benefits.

KEYWORDS

blockchain, distributed public ledger, healthcare, electronic health record, personal health record, medical health insurance

1. Introduction

Blockchain technology has been around for at least 20 years. It wasn't until recently that academics and businesses began to consider it strongly. This paper aims to highlight blockchain technology's existing and future applications in the healthcare industry through a thorough analysis. The authors wanted to highlight projects from the scientific and commercial sectors. This article may alternatively be seen as a concise summary and annotation of the idea behind the potential connections between healthcare and cryptocurrency and blockchain technology.

The blockchain's core value is closely related to trust and decentralization. A distributed database called blockchain supports transactions between unreliable entities, instead of having banks or another brokerage firm it acts as a middleman as is the case with traditional transactions, people and organizations may deal directly with one another. Because of this, it is anticipated that the blockchain will alter how people conduct international payments. It is crucial to investigate the potential benefits that this new technology might bring to the healthcare industry. It is clear that instead of just financial transactions, this effort will focus on healthcare data applicability and operations in general.

10.3389/fpubh.2023.1229386

One of the biggest sectors, healthcare accounts for more than 10% of the GDP of developed nations. The expense of providing effective healthcare services is rising, and patient data is fragmented which calls for the need to protect patient data and deliver effective services to be expanding whereas patient data is scattered, and sharing this private information may occasionally be subject to the practice of authorization management. Data is sometimes unavailable and inaccessible; all such problems in healthcare may be resolved with the help of blockchain. Distributed ledger technology is the foundation of blockchain technology, where transactions occur amongst peers rather than through a central authority. Decentralized transactions will be used throughout none of the entities can change any of the transactions once they have been added since they are all immutable which offers security and privacy for the transactions. A single location may be used to store patient data as a result, diagnosing a patient will be simple. IoT and blockchain can be utilized for real-time patient monitoring. As patient claims are properly monitored by health insurance companies, records are kept in a ledger that cannot be changed after it has been added. Blockchain has a wide variety of features that allow its implementation in a variety of applications. Decentralized storage and authentication are one of the key characteristics of blockchain. There are three main types of blockchain: public, private, and federated. Table 1 lists different types of blockchains. Anyone may participate in and validate public transactions on a public blockchain. The public is in charge of maintaining this kind of blockchain. Public blockchains include, for example, (1) Ethereum, (2) Bitcoin, (3) Bit shares, (4) Waves, and (5) Dash. Government agencies are in charge of maintaining private blockchains. Transactions are internally vetted and not accessible to the public. The consortium maintains federated blockchains, the third form of blockchain. This blockchain may or may not make transactions public. Federated blockchains include B3i, EWF, and Corda R3 as examples.

Blockchain has started to infiltrate a variety of different industries over the past few years, including finance and banking, capital markets, trade finance, business, real estate, media, government organizations, etc. The area where blockchain has enormous potential is healthcare. Data security, data access, data sharing, and interoperability are the top needs in the healthcare industry. Confidentiality and security are fundamental needs of the healthcare sector to protect patient medical information. In this digital age, cloud storage has become the most popular way to exchange and retrieve data, but since it is shared across a network, there is a danger of virus attacks and even a chance that personal information may be compromised. The constraints of healthcare requirements include data sharing, data access, data transmission, authenticity, and interoperability.

An EHR system (electronic health records) has replaced the conventional manual filing method, although it is expensive and labor-intensive. Following the introduction of EHR cloud-based systems to address the problems with EHR systems, however, these cloud-based systems still fell short in terms of encryption, data confidentiality, interoperability, and security standards. Every challenge the healthcare industry has, from interoperability to data security and transmission, might be addressed by blockchain technology.

TABLE 1 Blockchain types.

S.no	Blockchain type	Important property
1	Public blockchain	a) Maintained by the Public b) Anyone can participate and validate the transactions
2	Private blockchain	a) Maintained by Government Organizations b) Transactions are not public
3	Federated blockchain	a) Maintained by consortium b) Transactions may not be public

Blockchain technology has several characteristics that include immutability, transparency, distributed ledgers, data security, authentication, and decentralization. This highlights why and how this technology is becoming increasingly popular across all industries, but especially in those like healthcare where issues with legitimacy, dependability, and security are especially problematic. The blockchain architecture is seen in Figure 1 and consists of three layers: an application layer, a transaction layer, and a network layer. Users will communicate with one another using blockchain applications in the application layer, and immutable transactions will be carried out in the transaction layer and on a global ledger. At the network layer, information exchange will take place on a P2P network.

2. Motivation

It is now well acknowledged and perhaps even evident that blockchain will most likely significantly disrupt the healthcare sector. Nevertheless, the authors tried to define the amount of relevance anticipated in as many ways as possible, to measure it, and to quickly characterize it in general terms.

In order to start with peer-reviewed papers, a search at Scopus was made in January 2022 for publications that had both "blockchain" and "healthcare" in their "Article" or "Abstract" or "Keyword" sections. This search returned a total of 1,776 documents (Figure 2). Figure 3 shows their distribution according to their country of origin, scientific field, keyword, and year of publication. The authors can clearly see the apparent quick growth in the number of relevant publications, as well as an intriguing keyword distribution where some keywords (like "security" and "privacy") do not occur as frequently as expected. It is important to note that interest has extended globally, in addition to China and the USA. On January 2023, the authors ran the same query and received 2,828 documents in response. According to this, there are around three articles about blockchain technology in healthcare every day.

A similar search at Solulab turned up the 18 projects financed by independent companies under their Research and Innovation initiative. The acronyms for these R & D projects are provided

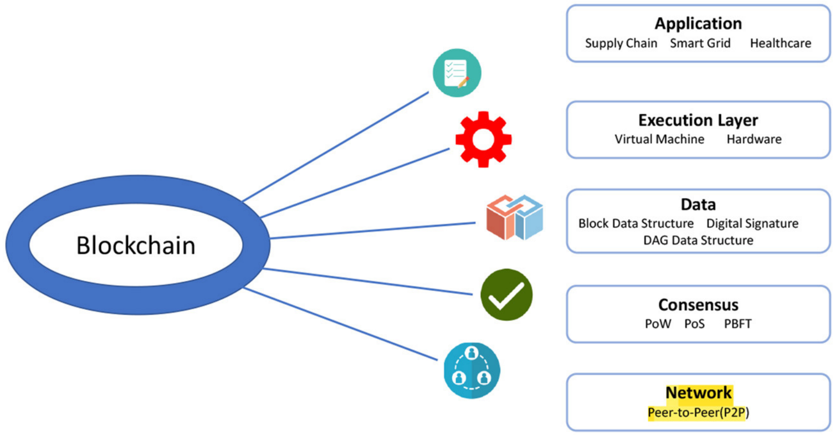


FIGURE 1
Architecture of blockchain.

below, followed by their titles, and they are ranked by Solulab according to how closely they adhere to the keywords. Presents a visual representation of the chronological distribution of these initiatives. These projects' time distribution is graphically shown in Figure 4.

BurstIQ: The organization uses blockchain to enhance the sharing and utilization of medical data.

SoluLab: Solulab uses blockchain to strengthen healthcare cybersecurity and to increase the sharing and use of medical data.

Medicalchain: The blockchain-based infrastructure used by Medicalchain keeps track of the source and safeguards patient identification.

Guardtime: Blockchain technology is used by Guardtime in cybersecurity applications, including healthcare.

Avaneer Health: Avaneer is a startup that seeks to use blockchain technology to improve healthcare efficiency. It is supported by Aetna, Anthem, and Cleveland Clinic. This is accomplished by utilizing a public ledger to provide improved claims processing, safe healthcare data transfers, and upgrade provider directories.

Chronicled: The usage of the Chronicled blockchain network ensures the secure delivery and thorough examination of drug supplies.

ProCredEx: ProCredEx has created a decentralized record system for healthcare credential information, ensuring that the data cannot be altered and remains permanently trackable. This enhances the effectiveness of intricate datasets, allowing for data customization to meet specific organizational requirements and the secure sharing of this data with authorized collaborators.

Robomed: Robomed captures patient data via blockchain and securely distributes it to the patient's healthcare professionals.

Patientory: The blockchain platform used by Patientory makes it possible to store and send sensitive medical data securely.

Doc.ai: The company uses artificial intelligence to decentralize medical data on the blockchain.

Encrypgen: It is now simpler to locate, share, save, and purchase genetic information because of the company's blockchain technology.

Coral Health: Coral Health uses blockchain to streamline administrative procedures, speed up the delivery of treatment, and enhance patient outcomes. The startup establishes faster connections between doctors, scientists, lab workers, and public health authorities by incorporating patients' records into DLT. To ensure that information and treatments are correct, Coral Health also uses smart contracts between patients and medical experts.

Embleema: Embleema is a tool for regulatory analytics and virtual experiments intended to accelerate the drug development process. Users are encouraged to provide their digital consent for the safe, unaltered acquisition of their medical data, which is subsequently recorded on the blockchain of Embleema and examined.

Blockpharma: Blockpharma provides a method to combat medication fraud and counterfeiting. Patients may find out whether they are taking fake medications using the company's app by scanning the supplier base and validating all points of shipping. Using a blockchain-based supply chain management system, Blockpharma claims it can screen out the 15% of medications that are fake worldwide.

Tierion: The blockchain of Tierion verifies all documentation, data, and pharmaceuticals to maintain a complete record of ownership. The company maintains evidence of ownership throughout a medical supply chain using timestamps and credentials.

FarmaTrust: The blockchain solutions from FarmaTrust may be used to manage prescriptions, validate the legitimacy of medical equipment, and protect patient data when they schedule vaccines and diagnostic tests. For example, the firm offers a service that prevents fraudulent pharmaceuticals from entering the supply chain and an app that allows customers to verify that their medications are authentic.

Nebula Genomics: Nebula Genomics is utilizing distributed ledger technology to eliminate extra fees and intermediaries in the genetic research industry. Annually, companies in the biotech and drug industries spend billions of dollars to get genetic data from other sources. While contributing to the development of a

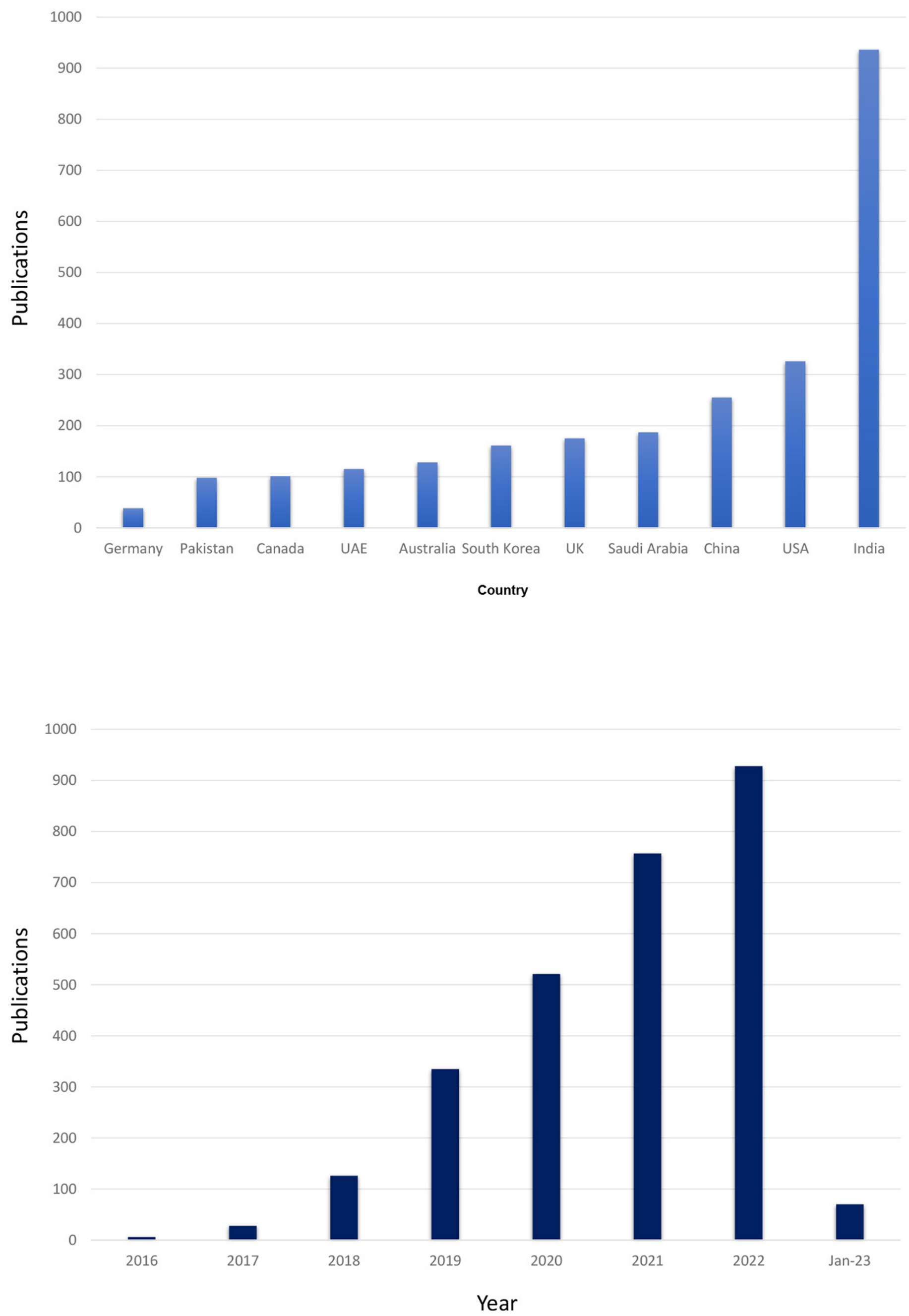


FIGURE 2
Publications per country over years.

substantial genetic library, Nebula Genomics invites users to sell their encrypted genetic data securely.

3. Background and concepts

In 2008 Satoshi Nakamoto implemented the cryptocurrency named Bitcoin based on the web white paper (1). This cryptocurrency works on open source technology and on the decentralized network in simple terms all nodes are connected mutually, and these nodes have the authority to leave and rejoin the network on demand and later receives the authentic record i.e., Proof of Work (PoW) referred to as the blockchain (1). To rejoin the network, they had to perform certain large computations to show evidence of their authentic members. PoW describes and gives proof of what happened when the particular nodes left the network. In cryptocurrency there may be a threat of a Sybil attack and this situation can be solved by claiming PoW from all nodes of the network which verifies transactions. The working of PoW can be understood by understanding the block of the bitcoin structure. The network consists of nodes that are nothing but participants and all of them have an identical ledger copy and these blocks of

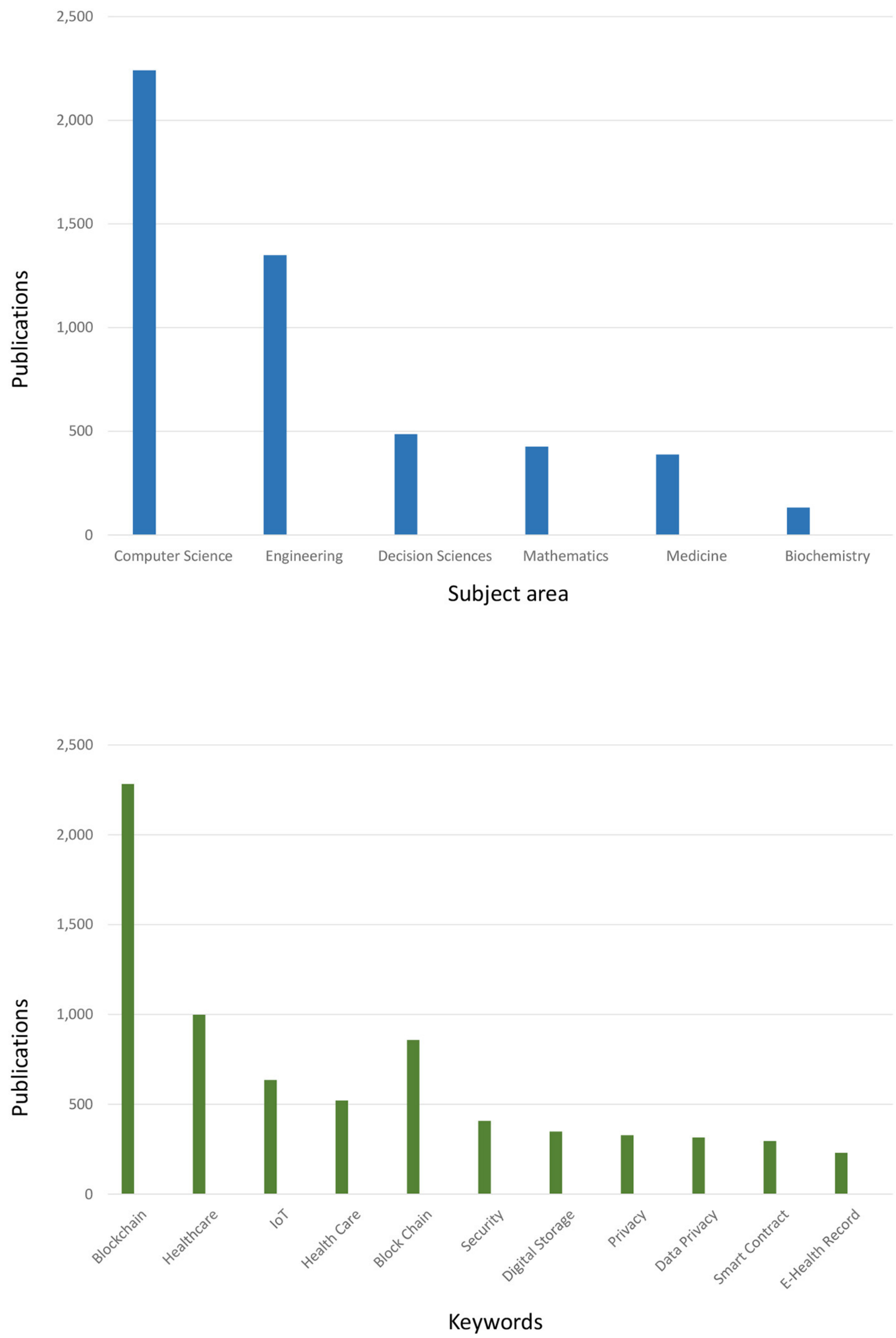


FIGURE 3
Publications in subject area and keywords.

information are attached (2). These blocks consist of transaction data of both sender and receiver, the extent of the transaction, and hash value. Hash values are used to link the blocks, therefore Blockchain is a series of blocks tailored together as illustrated in Figure 5. It reflects, how blocks are linked together in blockchain. The order of blocks linked together will be determined by PoW consensus in Bitcoin. Bitcoins are chained using hashing. Changing the hash value will lead to the invalidation of a block. To validate, the block hash value needs to be recalculated. Bitcoin as being public blockchain technology is susceptible to security and privacy threats; this property is not acceptable for healthcare systems where data privacy is concerned. Bitcoin along with throughput is most desirable for building healthcare applications (3).

Blockchain is best characterized by a decentralized, immutable ledger that records data. It enables entities to communicate with one another without the requirement for a centralized, reliable middleman. The blockchain contains blocks of data, comprising sets of information that grow consistently. Once integrated into the blockchain, these blocks are linked to the preceding and subsequent blocks through cryptographic procedures. All parties can read,

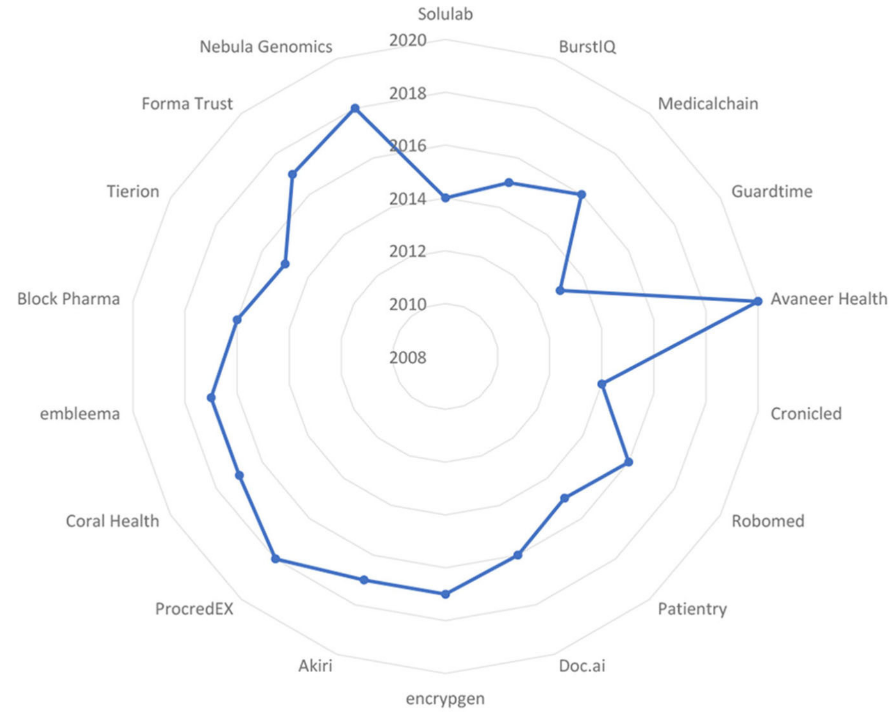


FIGURE 4
Time distribution of projects.

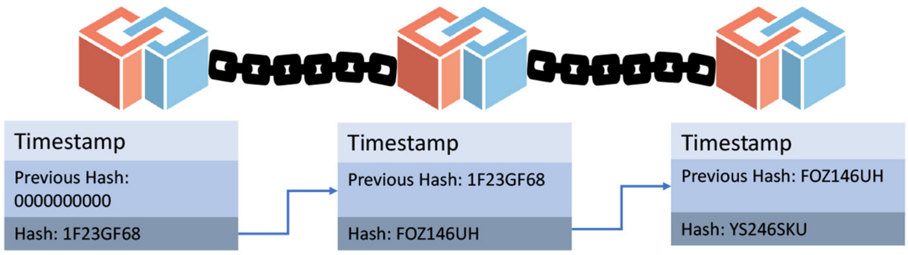


FIGURE 5
Representation of blocks in blockchain. Blockchain blocks include transaction details, including sender and recipient information, transaction size, and hash value. Blockchain is made up of a collection of interconnected blocks that are linked together using hash values.

write, and modify these data records/blocks in the blockchain's original form. Decentralized transactions and data processing are made possible. These characteristics have made blockchain very popular for a variety of uses. Blockchain also supports smart contracts, fully independent contracts that may be executed without a central authority. As of right now, Ethereum is the blockchain that facilitates smart contracts.

3.1. What is blockchain?

3.1.1. Key characteristics

Blockchain is decentralized, so nobody has full control over the data that is uploaded to it. Instead, a P2P network utilizing different consensus mechanisms approves the data that is uploaded to the blockchain. Persistence is another essential component of blockchain technology. Due to the distributed ledger's massive node storage, it is highly challenging to remove anything after it has been published onto the blockchain. Additionally, a lot of blockchains make use of the desirable potential anonymity (or pseudonymity) characteristic. By linking each block in a chain of blocks by containing the hash of the one before it, blockchains offer traceability and transparency. The Merkle tree-based organization of the blocks' transactions enables independent root-to-transaction verification (4, 5).

3.1.2. Type of blockchains

Blockchains come in three basic forms: consortium, public, and private. They have a wide range of characteristics that influence who is able to read from, write to, and access data on the blockchain. All users have access to the data on a public chain, and anybody may contribute and make changes to the consensus and the core software. The two biggest cryptocurrencies, Bitcoin and Ethereum, which are categorized as public permissionless networks, are among the several cryptocurrencies that utilize the public blockchain. Because only a small number of carefully chosen groups of businesses have access to observe and take part in the consensus process, a consortium blockchain may be perceived as being somewhat centralized. A decentralized network that is frequently centralized makes up a private blockchain. A few nodes can connect to the network, and they are often under the supervision of one central authority (4, 6, 7). The definition and classification of the many blockchain types discussed here are still subject to debate. There is currently no widespread agreement on what defining characteristics and consensus procedures are necessary to refer to a piece of technology as "blockchain" (8). For the creation of decentralized applications, there are currently available blockchain frameworks and platforms (Dapps). The most well-known blockchain development platforms to date are Ethereum (decentralized platform) and Hyperledger (framework), both of which let programmers add new blockchain apps to current blockchains and construct new test nets using their protocols (9).

3.1.3. Consensus algorithms

The procedure for approving data records on the decentralized ledger is a crucial component of blockchain technology. A distributed consensus method that checks the data entries accomplishes this. For this, a variety of consensus techniques have been put out and used; the three most popular ones are shown in Table 2 and are given below:

Proof of Work (PoW): PoW is the consensus system that is most closely connected to the blockchain because it is a part of Bitcoin. The process confirms the transaction and creates a new block for the blockchain. Miners compete in this procedure to finish the network transaction first. Competing is involved in mining. Miners receive incentives for successfully confirming a new block. This idea is supported by evidence of the significant electricity needed for Bitcoin mining, which is currently comparable to the requirements of a small nation (10).

Proof of Stake (PoS): With PoS, the node that will serve as an approving node is chosen based on its stake in the blockchain. When it comes to cryptocurrencies, a person's balance in a certain currency represents their investment. The "richest" node may, however, unjustly gain from this. Many hybrid PoS systems have been put out as a solution to this problem, where the approving node is selected using a mix of the stake and some randomization. The second-largest cryptocurrency, Ethereum, plans to transition to Proof of Stake from Proof of Work (4).

Practical Byzantine Fault Tolerance (PBFT): The underlying protocol of PBFT is a Byzantine agreement mechanism. This consensus procedure can't be used in a public blockchain since every node in PBFT must be known to the network, which places limitations on its use. Pre-prepared, prepared, and commit are the three separate stages of the PBFT consensus process. A node must get two-thirds of the votes from the other nodes in order to pass through the three phases. PBFT is now used by Hyperledger Fabric (11, 12).

TABLE 2 Comparison of consensus mechanisms.

Characteristics	PoV	PoS	PBFT
Management of nodes	Accessible	Accessible	Permissioned
Usage of energy	High	Medium	Low
Tolerated an adversary power	<25%	<51%	33.3% faulty
Example	Bitcoin	Ethereum	Hyperledger

3.1.4. Smart contracts

Smart contracts are supported by blockchain infrastructures like Ethereum. These are executed automatically contracts with clauses that are written clearly into the source code. Smart contracts operate independently of any third parties or middlemen since they are automatically implemented based on these set clauses. A blockchain transaction may activate this smart contract feature, and the healthcare industry seems like an appealing use for it (7).

3.2. Significance of blockchain in healthcare industry

Healthcare is a problem-driven, people- and data-intensive industry, and access to, updating, and trust in the information generated by its operations are essential for the sector's overall operations. According to a classification of healthcare operations into accident and emergency, health problem-solving, clinical decision-making, realization, and evaluation of knowledge-based care (Figure 6), it is essential to have a multidisciplinary team of healthcare professionals who treat patients with the best knowledge, technologies, and skills. To help students learn and develop their skill sets, the healthcare industry must work with educational institutions to give them access to patients and a training environment. In exchange, educational institutions give the industry skilled staff.

The healthcare sector faces a number of challenges, including data fragmentation, security and privacy issues, and interoperability problems caused by the usage of numerous standards in healthcare systems. By assuring secure and impenetrable data storage through a distributed ledger system, enabling secure data sharing to improve interoperability, and protecting patient privacy through data encryption, blockchain technology presents a potential solution. In the context of blockchain, various technologies have been implemented to address these issues, such as electronic health records (EHRs), health information exchanges (HIEs), and federated learning. While blockchain has great potential, it is important to understand that it is not a cure-all, and the healthcare sector is still exploring a variety of methods and technologies to meet its changing demands.

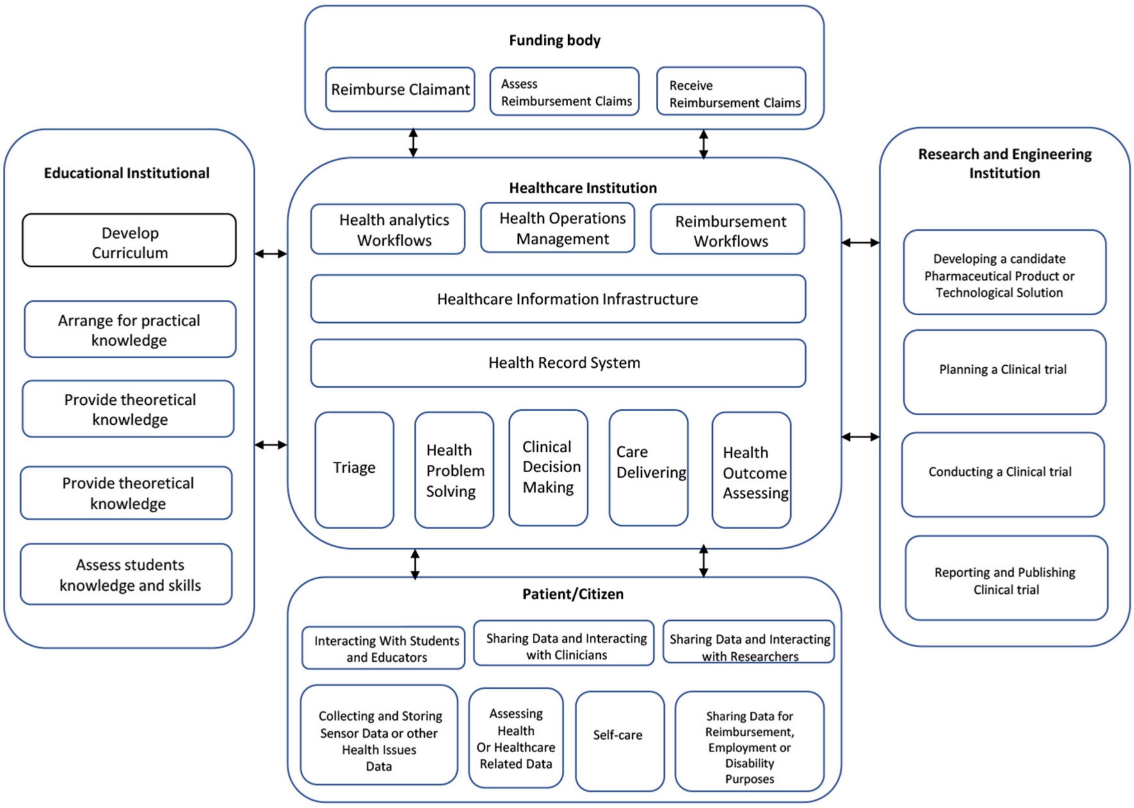


FIGURE 6
Map of healthcare industry.

Health institutions must assist with access to experts, informants, test subjects, and samples when working with organizations and businesses for research and engineering purposes. Health institutions are required to help in the design, planning, execution, and analysis of the studies while taking part in prospective clinical trials. In exchange, the research and engineering organizations offer the healthcare industry the most recent information, practices, and technology. Therefore, the operations of health institutions are closely linked to those of organizations that educate health professionals and conduct biomedical research and engineering (Figure 6). The efficient sharing of patient-related information and evidence, along with reimbursement processes, necessitates the exchange of data between different institutions. It's crucial to safeguard the sensitive data patients entrust to healthcare organizations. Ensuring patients' privacy while sharing data with other entities in the healthcare network requires measures like access control, maintaining data origin, preserving data integrity, and enabling interoperability. The traditional way of implementing access control assumes a level of trust between data owners and the entities holding the data. These entities often manage access restrictions. To enhance the health of individuals and communities through collaborative data access, exchange, and utilization, a variety of information systems, devices, or applications need to seamlessly connect within and across organizational boundaries in a coordinated manner.

Data provenance refers to the origins and historical records of data sources. It can enhance transparency and reliability in electronic health records (EHRs) and foster consumer confidence in EHR software. According to Courtney and Ware, data integrity encompasses data quality and its expected standards. This means that meeting or surpassing these standards directly impacts data reliability. The demand for real-world data from businesses and research units is increasing within healthcare institutions. Simultaneously, the public's trust in healthcare organizations is waning due to instances of unlawful data sharing, widely publicized breaches, and private information theft. Another constraint is the existence of healthcare system mispractices that exploit the same trust level, involving issues like counterfeit medications, fraudulent personnel, and patients. Given this overall scenario, a reassessment and adoption of alternative strategies are imperative (13, 14).

4. Related works

In this study, the keywords such as "Blockchain in healthcare," "healthcare record(s)," "healthcare system(s)," "healthcare and system and record(s)," "healthcare and blockchain," and "healthcare and blockchain" were used to search the Scopus and Google Scholar databases for EHR-related literature. By limiting the inclusion of the research in these influential review publications, this manuscript provides a concise overview. In this study, the word "Healthcare" is used instead of "Healthcare," as "Health care" refers to an industry or system that enables individuals to access the medical care they require (12, 15, 16).

People have expressed a desire to save their medical records as new technologies have emerged. For instance, research revealed that 87% of the health data that US citizens have gathered for themselves and their families are physical copies and that 42% of US citizens have done the same (17). EHR systems include several restrictions (like security) that make it challenging for users to exchange information. Rezaeibagha et al. (16) took into account the implications of security and privacy on EHR systems in their assessment of healthcare systems. They recognized a number of important aspects that affect information security and privacy, including encryption and scaling techniques, laws and regulations applicability, agreement and choice mechanisms, and integration and sharing of information. In the latest review to investigate the effectiveness of EHR systems, Afrizal et al. (18) explored the perspective of both individuals and an organization. Their research revealed organizational limitations, such as a lack of teamwork, inadequate executives, and a shortage of competent personnel. Aside from that, each person had their own limitations, such as a lack of computational resources and apprehension about novel technologies.

Modern technologies reduce the aforementioned restrictions. For instance, there are various ways to use blockchain to reduce hurdles in electronic health record systems (19). Blockchain is a distributed ledger technology that records network member transactions using immutable, reliable, and encrypted data (20). A system is referred to as a completely distributed system when no individual authority manages transaction operations that require the computing labor of several machines (21). UN's sustainable development goal may be greatly enhanced and achieved using blockchain, particularly in the healthcare sector (22). EHRs are one example of a public sector function that can be modernized with blockchain (23).

Zhang et al. (19) investigated the application of blockchain in healthcare systems using health scenarios that highlighted a patient-centric strategy in a framework for safe data sharing. They proposed adopting blockchain in seven different areas, including clinical documentation, patient-supervised cancer information, telemedicine treatment, patient verification, and health insurance disputes. In order to show how blockchain is related to patients' information-sharing behavior, authors focused on their health data. Even though Zhang et al. emphasized the benefits of utilizing it in health record administration, very few current works have offered a foundation for employing blockchain for patient health records. Homans et al. (24) built a blockchain-based management information system for electronic health records to solve security and privacy concerns. The ledger, database, committer, "orderer," endorser, and client were suggested as the six components that make up the framework. Fan et al. did not concentrate on the concepts of privacy labor and digital money and left them for future investigations.

Griggs et al., Fan et al. (25), added to the work done by Fan et al., using Homans (24), by introducing a private blockchain to address privacy concerns in blockchain usage. Public and private blocks come in two varieties. A block is a complete record of every completed and pending transaction. A private blockchain can be a useful option in healthcare administration, according to Griggs et al., given the serious security issues surrounding personal information. Reduced opt-in rates might be the outcome of privacy problems in EHR systems.

The study proposes "PeNLP Parser," a tool developed to extract and visualize exact geographic information about maternal, neonatal, and pediatric healthcare from unstructured data. The application extracts pertinent data and geolocations from the unstructured data using Natural Language Processing (NLP) techniques. By employing PeNLP Parser, healthcare providers and researchers can efficiently access and visualize essential geo-location data, enhancing their ability to make informed decisions and improve maternal and child healthcare services (26).

An integrated ontology is presented in the study by Patience et al. to aid in decision-making in the Maternal, Newborn, and Child Health (MNCH) sector. Context awareness is a feature of the ontology that enables it to take a variety of situational circumstances into account while offering decision help. By utilizing the integrated ontology, which can effectively analyze and understand data relevant to maternity, newborn, and child health to provide insightful analysis and suggestions for healthcare professionals and policymakers, the study seeks to improve decision-making processes in MNCH (27).

Sharma et al. (28) used the technique of the soft system to present qualitative evidence demonstrating that the usage of EHRs assisted with blockchain can increase patient engagement opt-in rates. They worked on the PHC strategy, which consists of a number of separate EHRs that are meant to be available to everyone in order to advance the healthcare system. They demonstrated how their suggested blockchain-based approach may boost patient and doctor trust in the sharing of medical records, while also enhancing the security and privacy of trustless PHC platforms.

The prospective impacts of blockchain on HIE were taken into account by Esmaeilzadeh et al. (29) their findings demonstrated that consumers are particularly planning complete blockchain-based privacy protection tools. Shahnaz et al. (30) provided a framework to reduce the scalability issue in the usage of blockchain in order to enable the adoption of blockchain in EHR. Blockchain-based healthcare systems have both beneficial and detrimental effects on patients and healthcare professionals, which presents new study opportunities (31).

The use of blockchain technology in the healthcare industry administration has lately been the subject of a number of studies, although its exact function in healthcare systems is yet unknown (32). To the best of the authors' knowledge, this is the only study to date that systematically examines the correlation between the intention of patients to share medical information and blockchain technology through mediating effects. The role of external incentives and security/privacy in the information system of a healthcare practitioner is also not well understood conceptually.

5. Applications of blockchain in healthcare

Blockchain technology possesses the potential to enhance the healthcare sector by prioritizing the patient within the system and enhancing the safeguarding, security, and seamless exchange of health information. In essence, the healthcare industry could undergo a substantial transformation through the widespread implementation of blockchain, resulting in comprehensive improvements in safety, security, and openness across all operations. Blockchain has the ability to improve things in this specific industry. It may perform a range of tasks, including controlling epidemics and safely encrypting patient data. Finally, by enabling secure data sharing between multiple healthcare systems with patient authorization, blockchain may enhance digital health (Figure 7).

5.1. Electronic health records

Blockchain technology could be employed to exchange and store patients' EHR. It may offer a supporting system for the exchange of health information that is safer, more transparent, and traceable. With the use of this technology, several data management systems that currently function in isolation might be linked to creating an EHR system that is both interconnected and functional. Therefore can patients and healthcare professionals easily access health information stored on the blockchain. It may be summed up in four easy steps:

1. The patient is examined by the doctor, who also registers the patient's report, lab findings, prescribed medications, and important comments in their current health information system. The patient's government-approved and accepted approved ID-related data fields are then transmitted to the blockchain using APIs. Here, a transaction is established.
2. Each transaction on the blockchain is verified and given a unique public key that would be stored on the blockchain.
3. Using the patient's decryption key, doctors and healthcare organizations may use APIs to build a query that retrieves the encrypted patient data.
4. Patients can give their doctor or the healthcare institution authorization to decrypt their data by giving them the private key, which serves as a password. The information is nonetheless encrypted for those without a secret key.

Asaph et al. introduced a decentralized system for managing medical records, aimed at handling electronic medical records (EMRs). In this system, MedRec provides patients with a comprehensive and reliable log of their medical history. This log is easily accessible and empowers patients with a better understanding of their medical past and any modifications to it, thus restoring their control over their medical information. The authors established a mechanism for patients to initiate sharing of their data across different medical entities using blockchain-based permission management.

MedRec's architecture enables specific permissions to be granted with a focus on maintaining confidentiality at a very detailed level. Additional constraints, such as setting time limits on viewing rights, can be placed within the various metadata segments that constitute a single medical record. These constraints can be independently communicated through smart contract provisions. By utilizing blockchain technology, the ledger ensures a transparent and traceable record of all medical interactions involving patients, doctors, and regulatory bodies (33).

5.2. Genomic data exchange platform

Platforms built on the blockchain aim to solve some of the biggest problems with governance, including the exchange of genetic data. The ultimate objective is to guarantee that organizations and people may exchange data with privacy-preserving algorithms that make it easier to adhere to moral and legal obligations. Even though most new platforms are still in their infancy, they might be regarded as advantageous since they provide fresh solutions to the governance issues associated with the sharing of genetic data. Notably, Blockchain represents more than

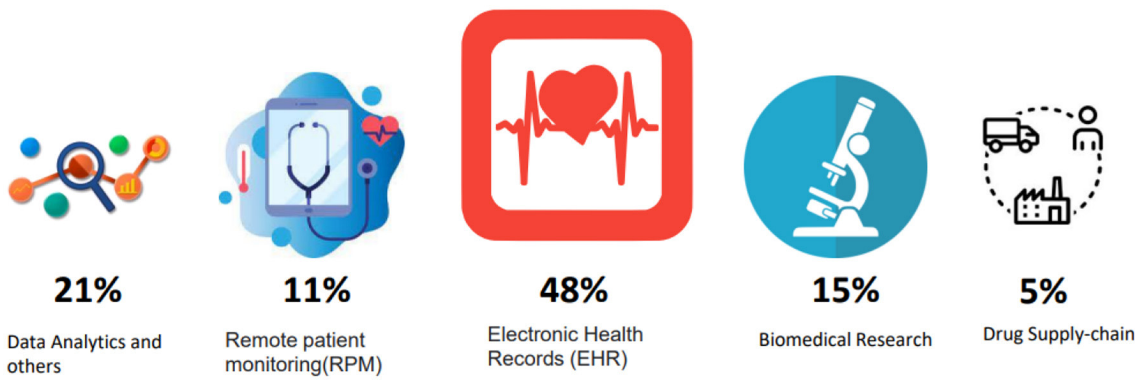


FIGURE 7
Emerging healthcare blockchain uses cases.

TABLE 3 Applications/delivery functions of blockchain in healthcare.

S.no	Application	References
1	Access control	(34–37)
		(38–41)
		(42–46)
		(25, 33, 47–49)
		(50–53)
		(54–59)
		(45, 60–64)
2	Privacy	(48, 65–74)
3	Distributed computing	(75–77)
	HIV	(78)
	LMIC	(79)
	Cancer	(73, 80)
	Diabetes	(81)
	Insurance	(82)
	Dermatology	(83)
	Clinical trial	(46, 77, 84, 85)
	Supply chain	(86–88)
	Genetics data	(89)
	Communication	(51)
	Medical imaging	(78, 90)
	Software design	(81, 91)
	Pharmaceuticals	(92)
	Health education	(47)
	Blood management	(93)
	Radiology oncology	(91, 94, 95)
	Patient-centered care	(85, 96–98)
	National Health system	(62)
4	Service delivery	
	Dyslexia	(99)
	Remote care	(100)
	Dental care	(99)
	DNA compression	(101, 102)
	Hemoglobin test (HbAlc)	(103)
	Arrhythmia classification	(68)

a mere technological foundation; it introduces a novel approach to overseeing open networks that leverages the advantages of decentralized systems, market dynamics, and consumer genetics. As a result, the primary innovation in this context surpasses technological aspects, although it is facilitated by them. Networks built on blockchain hold the potential to amplify data volume while introducing fresh ownership models and fostering active user participation in data-sharing governance. Especially in the realm of blockchain-based solutions, there's the potential to automate data access control processes, thus enhancing transparency and the availability of genetic data. Similarly, the incorporation of smart contracts could significantly bolster the enforcement of access agreements. This effort is noteworthy as it instills confidence among researchers and data custodians that subsequent data uses will adhere to the specified terms and conditions (Table 3).

Furthermore, effective implementation of the Blockchain-driven solution has the potential to reshape cultural norms around data sharing. This could result in a shift away from the dominance of public and commercial genetic test providers in controlling the dissemination of genetic information. This shift would empower patients and individuals to play a more influential role in the data-sharing landscape. Blockchain technology holds the capacity to establish novel shared resources that bridge the gap between market-driven dynamics and public resources. To initiate this transformation, substantial efforts in education, incentive design, ownership structure, and collaborative governance may be required. The aim of blockchain-based platforms is to empower patients and citizens to have agency over their data and participate in data sharing. Nonetheless, it's important to note that legal frameworks are still essential for the success of Blockchain-based solutions.

There are crucial scenarios where self-regulation might fall short, such as when assigning value to specific genetic datasets and determining ownership rights. To ensure that regulations align with the best interests of scientific advancement, society, and innovation, a thorough evaluation of the broader impact of such regulations within the realm of biomedical research is of utmost importance (104).

5.3. Medical imaging

Many scientists have been working on creating a feasible method for storing and distributing medical images in the realm of healthcare in recent years. The use of centralized cloud-based data centers in current practices raises privacy problems when exchanging information across a network, increases maintenance expenses, and necessitates vast storage capacity. The chain of transactions on the blockchain simply contains a list of the key owners who are authorized to view each research; no medical pictures are kept there. The image recipient must issue a signed request to the URL endpoint of the imaging source that generated the research before the actual image transfer can take place. Any person or organization that the owner (patient) who has granted permission to obtain this specific imaging study may be considered the requesting entity. The authors made use of the already existing work by the Integrating the Healthcare Enterprise (IHE) effort, which has established the ITI-43 transaction as a standard form for document retrieval across domains. The image source certifies the validity of the signature, confirms that the repository Unique ID specifies its own public key, confirms that the hashed UID matches to a study it previously released for the patient, and confirms—via the blockchain—that the patient has authorized the requester access to these images. If every requirement is met, the source sends back an ITI-43 response that includes the imaging study. To avoid eavesdropping, the request and the response are both sent via a secure channel at the transport layer (62, 105).

5.4. Pharmaceutical and drug discovery

Pharmaceutical research and development encompass a comprehensive journey, spanning multiple years dedicated to drug discovery, drug development, and regulatory approval within the pharmaceutical supply chain. However, drug counterfeiting occurs when drug producers and regulatory agencies conceal, lack control over, or use obsolete information on the supply of pharmaceuticals. This information results in the production, marketing, and use of fake pharmaceuticals. In situations like these, when data security and privacy protection are top priorities, blockchain is the most appropriate technology. It demonstrates the reliability of medical treatment for the people and the safety of pharmaceuticals sold on the market by utilizing current, genuine digital technologies. When considering the potential uses of blockchain technology in the sector, the pharmaceutical supply chain offers a convincing example: Pharmaceutical drugs are created and produced in specialized facilities before being routinely distributed to wholesalers and eventually patients. A possible way to improve medication research and secure the dependability of the pharmaceutical supply chain is through the incorporation of blockchain technology. The whole drug development process is facilitated and managed by this technology thanks to features like distributed ledgers, smart contracts, asset transfers, and proof of work (Table 4).

Despite its potential benefits, the influx of counterfeit and substandard pharmaceuticals into the legitimate supply chain poses a significant threat to public health. However, blockchain technology holds the promise of mitigating these risks and improving current systems. As the acceptance of blockchain technology becomes widespread, its capacity to revolutionize intercompany interactions is evident. While industries are only beginning to grasp its potential implications, it's essential to recognize that its applications span beyond specific sectors.

The full extent of the impact this transformative technology will have on the global landscape will become apparent over the years. As blockchain technology gains traction, its potential to reshape business interactions becomes increasingly apparent.

5.5. Remote patient monitoring in IoT

The utilization of the Internet of Things (IoT) and Blockchain advancements has found extensive application in various domains, such as remote patient monitoring (RPM). There has been swift development in crafting wearable medical devices within the IoT framework, equipped with diverse functionalities that facilitate the collection and analysis of live sensory data from patients. Data from IoT devices is gathered, analyzed, and stored centrally. However, there may be a number of drawbacks to this centralization, such as single-point failure, data manipulation, privacy concerns, etc. Blockchain's decentralized design can be used to solve these issues. Consequently, using IoT and blockchain to create a smart RPM system is a viable option. RPM is effective for treating a wide range of medical diseases, including diabetes, pediatrics, prenatal care, hypertension, and post-operative treatment. Patients can monitor their health at home with medical tools including blood pressure cuffs for hypertension, pulse oximeters for blood oxygen monitoring, glucometers for blood sugar levels, ECG machines for heart patients, and activity trackers. When one of these devices is attached to a patient, the health readings are automatically taken. The readings are then sent to the healthcare professionals who may monitor for health trends, and changes in conditions, and even be alerted when a patient's condition is likely to deteriorate using danger warnings.

5.6. Parallel Healthcare systems

The authors provided a paradigm for ACP-based Parallel Healthcare systems (PHSs) to improve the precision of diagnosis and efficacy of treatment. PHS uses computerized testing to analyze and evaluate a variety of medical prescriptions, simulating real-world healthcare systems to represent and reflect patient conditions, infections, and prescriptions. Healthcare operations for both humans and machines use real-time advancement as well as data-driven decision support in system arrangement. Additionally, they combined the recently developed blockchain-based PHS, which utilizes a consortium blockchain to interconnect patients, hospitals, healthcare organizations, and societies for critical health data interchange, medical record review, and accessibility to treatment (84).

5.7. Medical body sensor network

In addition to concurrent wireless technologies like wireless personal area networks, WBANs, and WPANs offer a lot of promise in healthcare monitoring systems to assess particular vital data and also to give location-based data (WPANs). A high incidence of both diagnostic and therapeutic studies is being driven by the expanding selection of wearable and subcutaneous medical equipment and their incorporation with wireless sensors. The development of Wireless Body Area Networks has been facilitated by the growing use of wireless networks and the ongoing shrinking of electrical invasive/non-invasive devices (WBANs). A WBAN allows for continuous patient health monitoring without interfering with the patient's regular daily activities. Numerous technologies have demonstrated their effectiveness in enabling WBANs applications by meeting their unique quality of service (QoS) needs, such as remote monitoring, biofeedback, and assisted living. It might be difficult to choose the best technology for a medical application because there are so many technologies that are now accessible.

5.8. Personal Health Records

A Personal Health Records (PHR) is a medical file where a patient manages their own health records as well as other information related to their healthcare. It is an electronic tool that enables people to manage their healthcare information securely. A patient may maintain and share their health information via an electronic PHR, which is secure software. Information entered by the patient from other sources, such as pharmacies and healthcare

TABLE 4 Applications of blockchain in healthcare and related technology used.

S.no	References	Application	Technology
1	Zhang and Lin (34), Thwin and Vasupongayya (106), Yue et al. (107)	Personal Health Information (PHI)	Proxy re-encryption technique, private blockchain, and consortium blockchain
2	Patel (62), Roehrs et al. (98)	Personal Health Records (PHR)	Distributed P2P network system
3	Xia et al. (96), Badr et al. (38), Alexaki et al. (42), Tian et al. (63), Pham et al. (102), Rouhani et al. (108) Li et al. (109), Theodouli et al. (110)	Parallel Healthcare system (PHSs)	Hyperledger framework, Ethereum, Cryptographic functions, and smart contracts
4	Al Omar et al. (111)	Healthcare System (e-health)	Body sensor network
5	Zhao et al. (112)	Remote Patient Monitoring (RPM)	Loosely coupled Blockchain, off-chain storage, and on-chain verification.
6	Jiang et al. (88)	Remote Healthcare system (RHS)	MAM module of the IOTA protocol
7	Alamoodi et al. (113) Brogan et al. (114) Uddin et al. (37)	Remote Patient Monitoring (RPM) in IoT	BSN, Sensor Data Provider, permissioned, consortium-managed blockchain, Private key, public key, and Healthcare Provider Interface
8	Griggs et al. (64), Dwivedi et al. (115)	EHR in a cloud environment	Constrained Goal Model (CGM)
9	Wehbe et al. (116)	Medical insurance storage	Blockchain technology
10	Wang and Song (36)	Multi-site Clinical Trials	Artificial system modeling
11	Zhou et al. (85)	Telecare Medical Information System	Hyperledger Fabric
12	Wang et al. (84)	Healthcare Information Exchange	Blockchain and cloud
13	Choudhury et al. (117)	Medical Data Sharing in Cloud Environment	Blockchain

providers, may also be included. The medical professionals who assisted in the treatment processes might be held accountable for the patient's disease. PHRs can be stored electronically or using a computer program. The sort of information that each individual may access can be managed by users using PHR.

PHR aims to provide an accurate, online-accessible summary of a patient's medical history. Lab reports, patient-reported outcomes, and other information could be included in the PHR. The phrase first appeared in usage between 1956 and 1978. It was first used in paper-based and digital systems.

People think that PHR and EHR are similar. However, this is not true. Doctors maintain an electronic health record (EHR). Hospitals, pharmacies, and doctors' offices can all create personal health records. Its main goal is to give patients the tools they need to manage their information. PHR calls for all the tools necessary for you to manage your health with the guidance of your doctors. It includes information such as doctor's names, medicine allergies, family history, sickness dates, and extra dosages.

5.9. Drug supply chain

At the moment, counterfeit drugs are pharmacology's primary problem. Health Research Funding estimates that 10%–30% of medicines in underdeveloped nations are bogus. The effects that counterfeit pharmaceuticals cause are not just different from those of traditional drugs; they also have distinct consequences on human health. According to the World Health Organization, around 30% of the medications marketed in Africa, Asia, and Latin America are unfortunately fake. In underdeveloped nations, where one in every ten medicines could be fake or don't follow drug standards, this issue is regrettably becoming worse. Monitoring drug safety has become more difficult as a result of the rise of online pharmacies. These medications run via a more convoluted, dispersed supply chain, making it more challenging to detect fakes and providing possibilities for fake pharmaceuticals to penetrate the real supply chain. Concern over the security of the drug supply chain has increased within the public health community, a process that affects everyone. Public health, a process that involves everyone, has grown more concerned about the safety of the drug supply chain. Transparency might make medication supply chain surveillance and inspection considerably more effective and accessible. Laws and regulations pertaining to blockchain are currently being developed because the technology is in its genesis stage. Even blockchain technology itself is evolving (e.g., "smart contracts," "Blockchain 2.0"), therefore more regulatory effect analyses and system simulation stress testing will be required in the future, along with engagement with key stakeholders, to undertake the cost-benefit analysis.

5.10. Fraud detection

The healthcare industry and public entities are very concerned about medical insurance fraud. The cost of healthcare fraud was reported as a loss to health insurance companies in the United States on an annual basis in the tens of billions. The patient's health is in danger from some types of fraud. This happens because the mechanism used to manually process medical insurance claims frequently fails to consider some stakeholders' consent throughout

TABLE 5 Blockchain frameworks in the healthcare domain.

Feature	Bitcoin	Ethereum	Hyperledger
Security	Public	Private	Private
Speed (transactions/s)	7	15	3,000
Scalability	Low	Low	High
Cost of transaction	Low	Low	Moderate
Need of cryptocurrency	Yes	yes	no

the assertion validation procedure. Blockchain is a peer-to-peer decentralized technology that can enable the secure, open, and unchangeable validation of medical claims.

Blockchain will be applied to Electronic Health Information (EHR) to address issues of security and privacy; by integrating blockchain with EHR, patients can effectively manage and save their records. Dissemination of patient data will be handled confidentially by blockchain using the blockchain medical records can be securely audited. With the blockchain, clinical data sharing will be securely and effectively managed with the help of blockchain. Blockchain enables IoT to provide a range of services, including Remote Patient Monitoring. Blockchain-based Remote Patient Monitoring will be maintained confidently. Health insurance companies are adapting blockchains to monitor false insurance claims made by patients. Various applications of blockchain and related technologies are listed in Table 2. Applications and delivery functions are listed in Table 5. Blockchain will be used in the pharmaceutical industry to address challenges like:

- Clinical data sharing.
- Supply chain of drugs.
- To manage clinical trials.
- Prescription management etc.

6. Use cases of blockchain-assisted decentralized applications

Blockchain technology is still in its adolescence stages, and even when prototype apps are created, they may serve just experimental or least functional objectives. However, some publications give implementation information for programs that have been created for specific use cases. Examples of such apps for the EMR use cases are, Hyperledger Fabric-based Healthchain (118), Acile (119), and MedRec (33), both of which were created on the Ethereum platform and Other instances include MeDShare (96), BlockHIE (88), FHIRChain (81), and MedBlock (25).

The administration of the pharmaceutical and drug supply chain, biomedical education and research, handling health insurance claims, and remote patient monitoring are examples of other blockchain use cases that have been discussed. However, there are also potential blockchain use cases that are still conceptual and do not yet have prototypes, such as the usage of blockchain in legal medicine (120).

7. Constraints of blockchain-assisted decentralized apps

The development of blockchain-assisted applications has been slowed by a variety of issues, including interoperability, security and privacy, scalability, speed, and patient involvement. Due to the lack of an open standard, it may be challenging for applications made by different manufacturers or on technology to interact with one another. This creates an interoperability dilemma. Consider taking a look at the other RPM apps: one was programmed on the Ethereum network, and the other was on the Permissioned Blockchain platform. Information exchange between these two systems would be difficult because, despite the encryption mechanisms used, it may still be possible to determine a patient's identity on a public blockchain by correlating enough data that are pertinent to that patient, blockchain-assisted healthcare apps have received criticism for their lack of security and privacy. Additionally, there is a chance that security flaws brought on by hostile deliberate attacks launched against the healthcare blockchain by criminal groups or even governmental entities might compromise patients' privacy. Various cryptocurrencies' blockchain networks have allegedly been the victim of many hacks. The private keys utilized by the blockchain for the cryptographed data are also prone to hacking, which might give unapproved individuals permission to the stored medical data. How well the immutability aspect of blockchain will function with the "right to be forgotten" clause of the EU GDPR, which specifies that users have the right to request the complete deletion of their personal data, raises further doubts. When a patient's clinical history is wiped away, it may be problematic since data once synced on the blockchain may not be altered or changed because of its immutability. Blockchain-based healthcare systems have major scalability challenges, especially in light of the amount of data involved. It is not advisable to put a vast quantity of biological data on the blockchain since doing so will always result in a severe performance hit. There is also the question of speed, since processing with blockchain may result in significant delays. For instance, the current validation technique used by the Ethereum blockchain platform involves participation from each network node. This results in a sizable processing lag, particularly when the input file is substantial. The management of medical records on the blockchain presents another challenge, particularly the inclusion of patients. It's likely that patients won't be able to or desire to get engaged in the processing of their medical data, primarily individuals of younger or elder age.

8. Addressing challenges

There are several constraints and limitations over how blockchain can be implemented in health IT systems, and many methods are being put up to get around these restrictions. For example, it is advised that encrypted health data be stored "off-chain," with just a limited amount of knowledge about the medical record and how to access it, in contemplation of the scalability issue. Thus, the "right to be forgotten" concern raised by the GDPR is similarly addressed. Although the connection to the medical data stored upon that distributed ledger cannot be unpublished, the particular medical data retained off-chain may be removed permanently. This countermeasure has a number of disadvantages, along with a gradual decrease in the built-in redundancy of the blockchain, which increases data availability. To better safeguard the data and maintain patient privacy, healthcare apps use permissioned blockchains, such as the consortium blockchain, instead of the permissionless, public blockchain. Additionally, by utilizing a strong software development strategy and all existing security safeguards during code development, many security vulnerabilities may be addressed. There are procedures in place on blockchains with permits for the healthcare sector that enable the rectification of transactions that go completely bonkers.

9. Future research

The application of distributed ledger technologies like blockchain in the healthcare industry is still in its premature stage, thus researchers must develop more proofs-of-concept and prototypes. This will aid scientists in their efforts to comprehend and advance technology as it pertains to healthcare systems. It is necessary to construct and test a number of the suggested principles, concepts, methods, and architectural designs in order to examine their merits and drawbacks. To check interoperability across diverse blockchain applications, open standards are necessary. Currently, the focus is on evaluating the capabilities of blockchain prototypes to demonstrate principles. Prior to blockchain becoming fully operational in healthcare systems, open protocols for compatibility must be developed. Researchers must immediately begin researching the issues with interoperability and standardization practices. Currently, there is a standards body (ISO/TC 307) where researchers may propose their concepts.

10. Conclusion

Blockchain technology has evolved since it was first used in Bitcoin to become a general-purpose technology with uses in many other sectors, including healthcare. The authors conducted a systematic review, employing the systematic mapping study methodology, to create a comprehensive overview of relevant research. This was done to gain insights into the current status of blockchain technology utilization in the healthcare sector. The study aimed to achieve several specific objectives: identification of healthcare applications utilizing blockchain technology, examination of exemplary apps developed for these applications, exploration of challenges and limitations linked to blockchain-based healthcare apps, analysis of the methodologies employed in creating such apps, and identification of potential avenues for future research. Through a meticulous search and selection process, the team identified 136 papers. These papers were subsequently scrutinized by the authors to address the research inquiries at hand.

The study we conducted revealed various healthcare applications for blockchain technology, including governing electronic medical records, overseeing the pharmaceutical and drug supply chain, advancing biotech research and education, enabling remote patient monitoring, and facilitating healthcare information analytics. To achieve these goals, different blockchain development approaches such as permissioned blockchains, off-chain storage, and smart contracts have been employed.

However, further investigation is necessary to refine, evaluate, and fully grasp the potential of blockchain technology in the healthcare sector. There is also a need for additional research to support ongoing endeavors aimed at resolving challenges related to scalability, latency, interoperability, confidentiality, and security in the implementation of blockchain-based healthcare solutions.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SS and YS focused on Blockchain and its practical applications. YH, MJ, and OV dedicated their efforts to the effective implementation of blockchain in the healthcare sector. All authors have reviewed and approved the final version of the manuscript for publication.

Acknowledgments

We, as the authors, express our gratitude to our esteemed reviewers for their accurate and concise suggestions, which greatly contributed to the optimal presentation of the material. The realization of this work became possible solely through the unwavering determination and bravery of the Ukrainian Army.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The viewpoints presented in this paper are exclusively those of the researchers and do not necessarily reflect the perspectives of their associated institutions, or the viewpoints of the publishing entity, the editorial team, and the reviewers. The publisher does not provide a guarantee or endorsement for any product discussed in this article or for any claims made by its manufacturer.

References

1. Nakamoto S. *Bitcoin: A Peer-to-Peer Electronic Cash System*. Cryptography Mailing List (2009).

2. Tschorsch F, Scheuermann B. Bitcoin and beyond: a technical survey on decentralized digital currencies. *IEEE Commun Surv Tutor*. (2016) 18:2084–123. doi: 10.1109/COMST.2016.2535718

3. Vujičić D, Jagodić D, Ranić S. Blockchain technology, bitcoin, and Ethereum: a brief overview. In: *2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH)*. East Sarajevo: IEEE (2018), p. 1–6. doi: 10.1109/INFOTEH.2018.8345547

4. Zheng Z, Xie S, Dai H, Chen X, Wang H. An overview of blockchain technology: architecture, consensus, and future trends. In: *2017 IEEE International Congress on Big Data (BigData Congress)*. Honolulu, HI: IEEE (2017), p. 557–64. doi: 10.1109/BigDataCongress.2017.85

5. Merkle RC. One way hash functions and DES. In: Brassard G, editor. *Advances in Cryptology—CRYPTO'89 Proceedings*. New York, NY: Springer (2001), p. 428–46. doi: 10.1007/0-387-34805-0_40

6. Nakamoto S. *Bitcoin: A Peer-to-peer Electronic Cash System*. Decentralized Business Review. (2008) p. 21260.

7. Buterin V. A next-generation smart contract and decentralized application platform. *White Paper*. (2014) 3:2–1.

8. Meyer M. *Blockchain Technology: Principles and Applications - IEEE Access*. (2016). Available online at: https://ieeeaccess.ieee.org/closed-special-sections/blockchain-technology-principles-and-applications/ (accessed January 24, 2023).

9. Cachin C. Architecture of the hyperledger blockchain fabric. In: *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, Vol. 310. Chicago, IL (2016), p. 1–4.

10. Malone D, O'Dwyer KJ. *Bitcoin Mining and its Energy Footprint*. IET. (2014). p. 280–85. doi: 10.1049/cp.2014.0699

11. Castro M, Liskov B. Practical byzantine fault tolerance. In: *Proceedings of the Third Symposium on Operating Systems Design and Implementation*, Vol. 99. New Orleans, LA (1999), p. 173–86.

12. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences—a scoping review. *Int J Med Inform*. (2020) 134:104040. doi: 10.1016/j.ijmedinf.2019.104040

13. Courtney R, Ware W. Some informal comments about integrity and the integrity workshop. In: Ruthberg ZG, Polk WT, editors. *Proc.* of the Invitational Workshop on Data Integrity. Gaithersburg, MD: National Institute of Standards and Technology, Special Publication (1989), p. 500–68.

14. Coiera E. *Guide to Health Informatics*. Boca Raton, FL: CRC Press (2015). doi: 10.1201/b13617

15. Hussien HM, Yasin SM, Udzir S, Zaidan AA, Zaidan BB. A systematic review for enabling of develop a blockchain technology in healthcare application: taxonomy, substantially analysis, motivations, challenges, recommendations and future direction. *J Med Syst*. (2019) 43:1–35. doi: 10.1007/s10916-019-1445-8

16. Rezaeibagha F, Win KT, Susilo W. A systematic literature review on security and privacy of electronic health record systems: technical perspectives. *Health Inf Manag J*. (2015) 44:23–38. doi: 10.1177/183335831504400304

17. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc*. (2011) 18:515–22. doi: 10.1136/amiajnl-2011-000105

18. Afrizal SH, Hidayanto AN, Handayani PW, Budiharsana M, Eryando T. Narrative review for exploring barriers to readiness of electronic health record implementation in primary health care. *Healthc Inform Res*. (2019) 25:141–52. doi: 10.4258/hir.2019.25.3.141

19. Zhang P, Schmidt DC, White J, Lenz G. Blockchain technology use cases in healthcare. In: Raj P, Deka GC, editors. *Advances in Computers*, Vol. 111. Amsterdam: Elsevier (2018), p. 1–41. doi: 10.1016/bs.adcom.2018.03.006

20. A Blockchain Research Framework - Business and Information Systems Engineering (2017). Available online at: https://link.springer.com/article/10.1007/s12599-017-0506-0#citeas (accessed January 26, 2023).

21. Treiblmaier H. The impact of the blockchain on the supply chain: a theory-based research framework and a call for action. *Supply Chain Manag*. (2018) 23:545–59. doi: 10.1108/SCM-01-2018-0029

22. Hughes L, Dwivedi YK, Misra SK, Rana NP, Raghavan V, Akella V. Blockchain research, practice and policy: applications, benefits, limitations, emerging research themes and research agenda. *Int J Inf Manage*. (2019) 49:114–29. doi: 10.1016/j.ijinfomgt.2019.02.005

23. Warkentin M, Orgeron C. Using the security triad to assess blockchain technology in public sector applications. *Int J Inf Manage*. (2020) 52:102090. doi: 10.1016/j.ijinfomgt.2020.102090

24. Homans GC. *Social Behavior: Its Elementary Forms*. Unknown. (1974).

25. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: efficient and secure medical data sharing via blockchain. *J Med Syst*. (2018) 42:136. doi: 10.1007/s10916-018-0993-7

26. Usip PU, Ekpenyong ME, Ijebu FF, Usang KJ, Udo IJ. PeNLP parser: an extraction and visualization tool for precise maternal, neonatal and child healthcare geo-locations from unstructured data. In: Agarwal B, Balas VE, Jain LC, editors. *Deep Learning in Biomedical and Health Informatics*. Boca Raton, FL: CRC Press (2021), p. 157–81. doi: 10.1201/9781003161233-8

27. Usip PU, Ekpenyong ME, Ijebu FF, Usang KJ. Integrated context-aware ontology for MNCH decision support. In: Tiwari S, Rodriguez FO, ŽJabbar MA, editors. *Semantic Models in IoT and Ehealth Applications*. Amsterdam: Elsevier (2022), p. 227–43. doi: 10.1016/B978-0-32-391773-5.00017-0

28. Sharma R, Zhang C, Wingreen SC, Kshetri N, Zahid A. Design of blockchain-based precision health-care using soft systems methodology. *Ind Manag Data Syst*. (2020) 120:608–32. doi: 10.1108/IMDS-07-2019-0401

29. Pouyan Esmaeilzadeh TM. *The Potential of Blockchain Technology for Health Information Exchange: Experimental Study from Patients' Perspectives*. (2019). Available online at: https://www.jmir.org/2019/6/e14184/citations (accessed January 26, 2023).

30. Shahnaz A, Qamar U, Khalid A. Using blockchain for electronic health records. *IEEE Access*. (2019) 7:147782–95. doi: 10.1109/ACCESS.2019.2946373

31. El-Gazzar R, Stendal K. Blockchain in health care: hope or hype? *J Med Internet Res*. (2020) 22:e17199. doi: 10.2196/17199

32. Wamba SF, Queiroz MM. Blockchain in the operations and supply chain management: benefits, challenges and future research opportunities. *Int J Inf Manag*. (2020) 52:102064. doi: 10.1016/j.ijinfomgt.2019.102064

33. Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: Using blockchain for medical data access and permission management. In: *2016 2nd International Conference on Open and Big Data (OBD)*. Vienna: IEEE (2016), p. 25–30. doi: 10.1109/OBD.2016.11

34. Zhang A, Lin X. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain. *J Med Syst*. (2018) 42:140. doi: 10.1007/s10916-018-0995-5

35. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. *JMIR Res Protoc*. (2018) 7:e10163. doi: 10.2196/10163

36. Wang H, Song Y. Secure cloud-based EHR system using attribute-based cryptosystem and blockchain. *J Med Syst*. (2018) 42:152. doi: 10.1007/s10916-018-0994-6

37. Uddin MA, Stranieri A, Gondal I, Balasubramanian V. Continuous patient monitoring with a patient centric agent: a block architecture. *IEEE Access*. (2018) 6:32700–26. doi: 10.1109/ACCESS.2018.2846779

38. Badr S, Gomaa I, Abd-Elrahman E. Multi-tier blockchain framework for IoT-EHRs systems. *Procedia Comput Sci*. (2018) 141:159–66. doi: 10.1016/j.procs.2018.10.162

39. Du Y, Liu J, Guan Z, Feng H. A medical information service platform based on distributed cloud and blockchain. In: *2018 IEEE International Conference on Smart Cloud (SmartCloud)*. New York, NY: IEEE (2018), p. 34–9. doi: 10.1109/SmartCloud.2018.00014

40. Srivastava G, Dwivedi AD, Singh R. Automated remote patient monitoring: data sharing and privacy using blockchain. *arXiv*. (2018). [preprint]. doi: 10.48550/arXiv.1811.03417

41. Bhuiyan MZA, Zaman A, Wang T, Wang G, Tao H, Hassan MM. Blockchain and big data to transform the healthcare. In: *Proceedings of the International Conference on Data Processing and Applications*. (2018), p. 62–8. doi: 10.1145/3224207.3224220

42. Alexaki S, Alexandris G, Katos V, Petroulakis NE. Blockchain-based electronic patient records for regulated circular healthcare jurisdictions. In: *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. Barcelona: IEEE (2018), p. 1–6. doi: 10.1109/CAMAD.2018.8514954

43. Cichosz SL, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: an operational concept. *J Diabetes Sci Technol*. (2019) 13:248–53. doi: 10.1177/1932296818790281

44. Shae Z, Tsai JJ. On the design of a blockchain platform for clinical trial and precision medicine. In: *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. Atlanta, GA: IEEE (2017), p. 1972–80. doi: 10.1109/ICDCS.2017.61

45. Yang H, Yang B. A blockchain-based approach to the secure sharing of healthcare data. In: *Proceedings of the Norwegian Information Security Conference*. Oslo (2017), p. 100–11.

46. Gökalp E, Gökalp MO, Çoban S, Eren PE. Analysing opportunities and challenges of integrated blockchain technologies in healthcare. In: Wrycza S, Malankowski J, editors. *Information Systems: Research, Development, Applications, Education: 11th SIGSAND/PLAIS EuroSymposium 2018, Gdansk, Poland, September 20, 2018, Proceedings 11*. Cha: Springer (2018), p. 174–83. doi: 10.1007/978-3-030-00060-8_13

47. Funk E, Riddell J, Ankel F, Cabrera D. Blockchain technology: a data framework to improve validity, trust, and accountability of information exchange in health professions education. *Acad Med*. (2018) 93:1791–4. doi: 10.1097/ACM.0000000000002326

48. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. *J Med Syst*. (2019) 43:1–9. doi: 10.1007/s10916-018-1121-4

49. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. *JMIR mHealth uHealth*. (2017) 5:e7938. doi: 10.2196/mhealth.7938

50. Sun Y, Zhang R, Wang X, Gao K, Liu L. A decentralizing attribute-based signature for healthcare blockchain. In: *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. Hangzhou: IEEE (2018), p. 1–9. doi: 10.1109/ICCCN.2018.8487349

51. Juneja A, Marefat M. Leveraging blockchain for retraining deep learning architecture in patient-specific arrhythmia classification. In: *2018 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. Las Vegas, NV: IEEE (2018), p. 393–7. doi: 10.1109/BHI.2018.8333451

52. Nagasubramanian G, Sakthivel RK, Patan R, Gandomi AH, Sankayya M, Balusamy B. Securing e-health records using keyless signature infrastructure blockchain technology in the cloud. *Neural Comput Appl*. (2020) 32:639–47. doi: 10.1007/s00521-018-3915-1

53. Mikula T, Jacobsen RH. Identity and access management with blockchain in electronic healthcare records. In: *2018 21st Euromicro Conference on Digital System Design (DSD)*. Prague: IEEE (2018), p. 699–706. doi: 10.1109/DSD.2018.00008

54. Tang F, Ma S, Xiang Y, Lin C. An efficient authentication scheme for blockchain-based electronic health records. *IEEE Access*. (2019) 7:41678–89. doi: 10.1109/ACCESS.2019.2904300

55. Cha HJ, Yang HK, Song YJ. A study on access structure management of CP-ABTD based blockchain for medical information monitoring system. *Adv Sci Lett*. (2018) 24:2026–30. doi: 10.1166/asl.2018.11838

56. Zhang X, Poslad S, Ma Z. Block-based access control for blockchain-based electronic medical records (EMRs) query in eHealth. In: *2018 IEEE Global Communications Conference (GLOBECOM)*. Abu Dhabi: IEEE (2018), p. 1–7. doi: 10.1109/GLOCOM.2018.8647433

57. Franceschi M, Morelli D, Plans D, Brown A, Collomosse J, Coutts L, et al. ComeHere: Exploiting ethereum for secure sharing of health-care data. In: *Euro-Par 2018: Parallel Processing Workshops: Euro-Par 2018 International Workshops, Turin, Italy, August 27-28, 2018, Revised Selected Papers 24*. Cham: Springer (2019), p. 585–96. doi: 10.1007/978-3-030-10549-5_46

58. Cao S, Zhang G, Liu P, Zhang X, Neri F. Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain. *Inf Sci*. (2019) 485:427–40. doi: 10.1016/j.ins.2019.02.038

59. Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. Analyzing the performance of a blockchain-based personal health record implementation. *J Biomed Inform*. (2019) 92:103140. doi: 10.1016/j.jbi.2019.103140

60. Mohsin AH, Zaidan A, Zaidan B, Albahri OS, Albahri AS, Alsalem M, et al. Based blockchain-PSO-AES techniques in finger vein biometrics: a novel verification secure framework for patient authentication. *Comput Stand Interfaces*. (2019) 66:103343. doi: 10.1016/j.csi.2019.04.002

61. Chakraborty S, Aich S, Kim HC. A secure healthcare system design framework using blockchain technology. In: *2019 21st International Conference on Advanced Communication Technology (ICACT)*. Pyeongchang: IEEE (2019), p. 260–64. doi: 10.23919/ICACT.2019.8701983

62. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. *Health Informatics J*. (2019) 25:1398–411. doi: 10.1177/1460458218769699

63. Tian H, He J, Ding Y. Medical data management on blockchain with privacy. *J Med Syst*. (2019) 43:1–6. doi: 10.1007/s10916-018-1144-x

64. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. *J Med Syst*. (2018) 42:1–7. doi: 10.1007/s10916-018-0982-x

65. Ito K, Tago K, Jin Q. i-Blockchain: a blockchain-empowered individual-centric framework for privacy-preserved use of personal health data. In: *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*. Hangzhou: IEEE (2018), p. 829–33. doi: 10.1109/ITME.2018.00186

66. Vora J, Nayyar A, Tanwar S, Tyagi S, Kumar N, Obaidat MS, et al. BHEEM: A blockchain-based framework for securing electronic health records. In: *2018 IEEE Globecom Workshops (GC Wkshps)*. Abu Dhabi: IEEE (2018), p. 1–6. doi: 10.1109/GLOCOMW.2018.8644088

67. Ramani V, Kumar T, Bracken A, Liyanage M, Ylianttila M. Secure and efficient data accessibility in blockchain based healthcare systems. In: *2018 IEEE Global Communications Conference (GLOBECOM)*. Abu Dhabi: IEEE (2018), p. 206–12. doi: 10.1109/GLOCOM.2018.8647221

68. Wutthikarn R, Hui YG. Prototype of blockchain in dental care service application based on hyperledger composer in hyperledger fabric framework. In: *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. Chiang Mai: IEEE (2018), p. 1–4. doi: 10.1109/ICSEC.2018.8712639

69. Zaghloul E, Li T, Ren J. Security and privacy of electronic health records: decentralized and hierarchical data sharing using smart contracts. In: *2019 International Conference on Computing, Networking and Communications (ICNC)*. Honolulu, HI: IEEE (2019), p. 375–9. doi: 10.1109/ICCNC.2019.8685552

70. Nortey RN, Yue L, Agdedanu PR, Adjeisah M. Privacy module for distributed electronic health records (EHRs) using the blockchain. In: *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*. Suzhou: IEEE (2019), p. 369–74. doi: 10.1109/ICBDA.2019.8713188

71. Weng Y, Zhou T, Li Y, Qiu X. Nas-unet: Neural architecture search for medical image segmentation. *IEEE Access*. (2019) 7:44247–57. doi: 10.1109/ACCESS.2019.2908991

72. Liu J, Wang Z, Sun D, Wang X. HIV digital vaccine strategy: an application of blockchain technology in preventing the spread of HIV. *Tech Rep*. (2018). doi: 10.2196/preprints.12308

73. Kim THJ, Lampkins J. BRICS: blockchain-based resilient information control system. In: *2018 IEEE International Conference on Big Data (Big Data)*. Seattle, WA: IEEE (2018), p. 5363–5. doi: 10.1109/BigData.2018.8621993

74. Rupasinghe T, Burstein F, Rudolph C, Strange S. Towards a blockchain based fall prediction model for aged care. In: *ACSW 2019: Proceedings of the Australasian Computer Science Week Multi-Conference*. Sydney, NSW (2019). 83:1–10. doi: 10.1145/3290688.3290736

75. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. *F1000Res*. (2016) 5:2541. doi: 10.12688/f1000research.9756.1

76. Angeletti F, Chatzigiannakis I, Vitaletti A. The role of blockchain and IoT in recruiting participants for digital clinical trials. In: *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. Split: IEEE (2017), p. 1–5. doi: 10.23919/SOFTCOM.2017.8115590

77. Ortega DR, Oikonomou CM, Ding HJ, Rees-Lee P, Alexandria, Jensen GJ. ETDB-Caltech: a blockchain-based distributed public database for electron tomography. *PLoS ONE*. (2019) 14:e0215531. doi: 10.1371/journal.pone.0215531

78. Tang H, Tong N, Ouyang J. Medical images sharing system based on blockchain and smart contract of credit scores. In: *2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN)*. Shenzhen: IEEE (2018), p. 240–1. doi: 10.1109/HOTICN.2018.8605956

79. Heston T. *A Case Study in Blockchain Healthcare Innovation*. Unknown. (2017).

80. Tseng JH, Liao YC, Chong B. Liao Sw. Governance on the drug supply chain via gcoin blockchain. *Int J Environ Res Public Health*. (2018) 15:1055. doi: 10.3390/ijerph15061055

81. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: applying blockchain to securely and scalably share clinical data. *Comput Struct Biotechnol J*. (2018) 16:267–78. doi: 10.1016/j.csbj.2018.07.004

82. Azbeg K, Ouchetto O, Andaloussi SJ, Fetjah L, Sekkaki A. Blockchain and IoT for security and privacy: a platform for diabetes self-management. In: *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*. Brussels: IEEE (2018), p. 1–5. doi: 10.1109/CloudTech.2018.8713343

83. Till BM, Peters AW, Afshar S, Meara JG. From blockchain technology to global health equity: can cryptocurrencies finance universal health coverage? *BMJ Global Health*. (2017) 2:e000570. doi: 10.1136/bmjgh-2017-000570

84. Wang S, Wang J, Wang X, Qiu T, Yuan Y, Ouyang L, et al. Blockchain-powered parallel healthcare systems based on the ACP approach. *IEEE Trans Comput Soc Syst*. (2018) 5:942–50. doi: 10.1109/TCSS.2018.2865526

85. Zhou L, Wang L, Sun Y. MIStore: a blockchain-based medical insurance storage system. *J Med Syst*. (2018) 42:149. doi: 10.1007/s10916-018-0996-4

86. Bocek T, Rodrigues BB, Strasser T, Stiller B. Blockchains everywhere-a use-case of blockchains in the pharma supply-chain. In: *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. Lisbon: IEEE (2017), p. 772–7. doi: 10.23919/INM.2017.7987376

87. Castaldo L, Cinque V. Blockchain-based logging for the cross-border exchange of eHealth data in Europe. In: Gelenbe E, Campegiani P, Czachrski T, editors. *Security in Computer and Information Sciences: First International ISCIS Security Workshop 2018, Euro-CYBERSEC 2018, London, UK, February 26-27, 2018, Revised Selected Papers 1*. Cham: Springer (2018), p. 46–56. doi: 10.1007/978-3-319-95189-8_5

88. Jiang S, Cao J, Wu H, Yang Y, Ma M, He J. Blochie: a blockchain-based platform for healthcare information exchange. In: *2018 IEEE International Conference on Smart Computing (Smartcomp)*. Taormina: IEEE (2018), p. 49–56. doi: 10.1109/SMARTCOMP.2018.00073

89. Zhang S, Kim A, Liu D, Nuckchady SC, Huang L, Masurkar A, et al. Genie: a secure, transparent sharing and services platform for genetic and health data. *arXiv*. (2018) [preprint]. doi: 10.48550/arXiv.1811.01431

90. Rahman MA, Hossain MS, Rashid MM, Barnes SJ, Alhamid MF, Guizani M, et al. blockchain-based non-invasive cyber-physical occupational therapy framework: BCI perspective. *IEEE Access*. (2019) 7:34874–84. doi: 10.1109/ACCESS.2019.2903024

91. Ackerman A, Chang A, Diakun-Thibault N, Forni L, Landa F, Mayo J, et al. *Blockchain and Health it: Algorithms, Privacy and Data*. Project PharmOrchard of MIT's Experimental Learning "MIT FinTech: Future Commerce", White Paper August. (2016).

92. Mannaro K, Baralla G, Pinna A, Ibba S. A blockchain approach applied to a teledermatology platform in the Sardinian region (Italy). *Information*. (2018) 9:44. doi: 10.3390/info9020044

93. Kumar R, Tripathi R. Traceability of counterfeit medicine supply chain through Blockchain. In: *2019 11th international conference on communication systems and networks (COMSNETS)*. Bengaluru: IEEE (2019), p. 568–70. doi: 10.1109/COMSNETS.2019.8711418

94. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. How blockchain could empower ehealth: An application for radiation oncology. In: Begoli E, Wang F, Luo G, editors. *Data Management and Analytics for Medicine and Healthcare: Third International Workshop, DMAH 2017, Held at VLDB 2017, Munich, Germany, September 1, 2017, Proceedings 3*. Springer (2017), p. 3–6. doi: 10.1007/978-3-319-67186-4_1

95. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. In: *AMIA Annual Symposium Proceedings*, Vol. 2017. Washington, DC: American Medical Informatics Association (2017), p. 650.

96. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MeDShare: trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access*. (2017) 5:14757–67. doi: 10.1109/ACCESS.2017.2730843

97. Lee SJ, Cho GY, Ikeno F, Lee TR. BAQALC: blockchain applied lossless efficient transmission of DNA sequencing data for next generation medical informatics. *Appl Sci*. (2018) 8:1471. doi: 10.3390/app8091471

98. Roehrs A, Da Costa CA, da Rosa Righi R. OmniPHR: A distributed architecture model to integrate personal health records. *J Biomed Inform*. (2017) 71:70–81. doi: 10.1016/j.jbi.2017.05.012

99. Yli-Huumo J, Ko D, Choi S, Park S. Smolander K. Where is current research on blockchain technology?–a systematic review. *PLoS ONE*. (2016) 11:e0163477. doi: 10.1371/journal.pone.0163477

100. Celiz RC, De La Cruz YE, Sanchez DM. Cloud model for purchase management in health sector of peru based on IoT and blockchain. In: *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. Krabi: IEEE (2018), p. 328–34. doi: 10.1109/IEMCON.2018.8615063

101. Iyengar A, Kundu A, Sharma U, Zhang P. A trusted healthcare data analytics cloud platform. In: *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. Vienna: IEEE (2018), p. 1238–49. doi: 10.1109/ICDCS.2018.00123

102. Pham HL, Tran TH, Nakashima Y. A secure remote healthcare system for hospital using blockchain smart contract. In: *2018 IEEE Globecom Workshops (GC Wkshps)*. Abu Dhabi: IEEE (2018), p. 1–6. doi: 10.1109/GLOCOMW.2018.8644164

103. Rahman MA, Hassanain E, Rashid MM, Barnes SJ, Hossain MS. Spatial blockchain-based secure mass screening framework for children with dyslexia. *IEEE Access*. (2018) 6:61876–85. doi: 10.1109/ACCESS.2018.2875242

104. Shabani M. Blockchain-based platforms for genomic data sharing: a decentralized approach in response to the governance problems? *J Am Med Inform Assoc*. (2019) 26:76–80. doi: 10.1093/jamia/ocy149

105. Bartel M, Boyer J, Fox B, LaMacchia B, Simon E. *XML-Signature Syntax and Processing, 2002*. (2003). Available online at: https://www.w3.org/TR/xmldsig-core/ (accessed May 20, 2023).

106. Thwin TT, Vasupongayya S. Blockchain based secret-data sharing model for personal health record system. In: *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*. Krabi: IEEE (2018), p. 196–201. doi: 10.1109/ICAICTA.2018.8541296

107. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *J Med Syst*. (2016) 40:1–8. doi: 10.1007/s10916-016-0574-6

108. Rouhani S, Butterworth L, Simmons AD, Humphery DG, Deters R. MediChain TM: a secure decentralized medical data asset management system. In: *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. Halifax, NS: IEEE (2018), p. 1533–8. doi: 10.1109/Cybermatics_2018.2018.00258

109. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S. Blockchain-based data preservation system for medical data. *J Med Syst*. (2018) 42:1–13. doi: 10.1007/s10916-018-0997-3

110. Theodouli A, Arakliotis S, Moschou K, Votis K, Tzovaras D. On the design of a blockchain-based system to facilitate healthcare data sharing. In: *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. New York, NY: IEEE (2018), p. 1374–9. doi: 10.1109/TrustCom/BigDataSE.2018.00190

111. Al Omar A, Bhuiyan MZA, Basu A, Kiyomoto S, Rahman MS. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. *Future Gener Comput Syst*. (2019) 95:511–21. doi: 10.1016/j.future.2018.12.044

112. Zhao H, Bai P, Peng Y, Xu R. Efficient key management scheme for health blockchain. *CAAI Trans Intell Technol*. (2018) 3:114–8. doi: 10.1049/trit.2018.0014

113. Alamoodi AH, Zaidan B, Zaidan AA, Samuri SM, Ismail AR, Zughoul O, et al. A review of data analysis for early-childhood period: taxonomy, motivations, challenges, recommendation, and methodological aspects. *IEEE Access*. (2019) 7:51069–103. doi: 10.1109/ACCESS.2018.2889540

114. Brogan J, Baskaran I, Ramachandran N. Authenticating health activity data using distributed ledger technologies. *Comput Struct Biotechnol J*. (2018) 16:257–66. doi: 10.1016/j.csbj.2018.06.004

115. Dwivedi AD, Srivastava G, Dhar S, Singh R. A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors*. (2019) 19:326. doi: 10.3390/s19020326

116. Wehbe Y, Al Zaabi M, Svetinovic D. Blockchain AI framework for healthcare records management: constrained goal model. In: *2018 26th Telecommunications forum (TELFOR)*. Belgrade: IEEE (2018), p. 420–5. doi: 10.1109/TELFOR.2018.8611900

117. Choudhury O, Fairoza N, Sylla I, Das A. A blockchain framework for managing and monitoring data in multi-site clinical trials. *arXiv*. (2019). [preprint]. doi: 10.48550/arXiv.1902.03975

118. Ahram T, Sargolzaei A, Sargolzaei S, Daniels J, Amaba B. Blockchain technology innovations. In: *2017 IEEE Technology and Engineering Management Conference (TEMSCON)*. Santa Clara, CA: IEEE (2017), p. 137–41. doi: 10.1109/TEMSCON.2017.7998367

119. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustain Cities Soc*. (2018) 39:283–97. doi: 10.1016/j.scs.2018.02.014

120. Roman-Belmonte JM. la Corte-Rodriguez HD, Rodriguez-Merchan EC. How blockchain technology can change medicine. *Postgrad Med*. (2018) 130:420–7. doi: 10.1080/00325481.2018.1472996

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 23 October 2023
DOI 10.3389/fpubh.2023.1259158

Check for updates

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Habtamu Alganeh Guadie,
Bahir Dar University, Ethiopia
Simon Robert Stones,
Envision Pharma Group, United Kingdom

*CORRESPONDENCE
Zhenzhen Xiong
✉ xzz62308631@163.com
Zhe Li
✉ jay_li@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 15 July 2023
ACCEPTED 29 September 2023
PUBLISHED 23 October 2023

CITATION
Zhang M, Zhang H, Zhu R, Yang H, Chen M, Wang X, Li Z and Xiong Z (2023) Factors affecting the willingness of patients with type 2 diabetes to use digital disease management applications: a cross-sectional study.
Front. Public Health 11:1259158.
doi: 10.3389/fpubh.2023.1259158

Factors affecting the willingness of patients with type 2 diabetes to use digital disease management applications: a cross-sectional study

Mingjiao Zhang[1,2†], Hao Zhang[2†], Rong Zhu[3†], Huiqi Yang[4], Mengjie Chen[1], Xiaoxia Wang[1], Zhe Li[5,6]* and Zhenzhen Xiong[1]*

[1]School of Nursing, Chengdu Medical College, Chengdu, Sichuan, China, [2]West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China, [3]The 3rd Affiliated Hospital of Chengdu Medical College, Pidu District People's Hospital, Chengdu, Sichuan, China, [4]Nanbu Country People's Hospital, Nanchong, Sichuan, China, [5]Mental Health Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China, [6]Sichuan Clinical Medical Research Center for Mental Disorders, Chengdu, Sichuan, China

Background: The global burden of type 2 diabetes has significantly increased, leading to a considerable impact on healthcare systems worldwide. While the advent of mobile healthcare has provided some relief by addressing the shortage of certain medical resources, its adoption among the Chinese population remains relatively low. To extend the benefits of mHealth to a greater number of Chinese diabetic patients, it is essential to investigate the factors that influence their willingness to utilize it and implement targeted interventions based on these influencing factors. The Technology Acceptance Model (TAM) is widely employed to examine users' ultimate usage behaviors, and previous studies have indicated the potential relevance of the Perceived Risk (PR) theory and the eHealth Literacy Theory to users' usage behaviors.

Objective: Our objective was to investigate the determinants that affect the willingness of Chinese patients diagnosed with type 2 diabetes patients to utilize digital disease management applications (DDMAs).

Methods: We conducted a cross-sectional study of patients with type 2 diabetes in three tertiary general hospitals in Chengdu using questionnaires designed by the investigators. Participants were sampled using a convenience sampling method. The questionnaire comprised three sections: socio-demographic profile and medical history; current awareness and willingness to use digital disease management applications; and the current level of e-health literacy. Structural equation modeling was employed to assess the impact of patient awareness of DDMAs and e-health literacy on the willingness to use such DDMAs.

Results: (1) Patients' attitudes toward using DDMAs were significantly influenced by perceived ease of use ($\beta = 0.380$, $P < 0.001$) and perceived usefulness ($\beta = 0.546$, $P < 0.001$); (2) Electronic health literacy exerted a significant impact on patients' perceived usefulness ($\beta = 0.115$, $P = 0.018$) and perceived ease of use ($\beta = 0.659$, $P < 0.001$); (3) Patients' willingness to use was significantly influenced by perceived usefulness ($\beta = 0.137$, $P < 0.001$) and use attitude ($\beta = 0.825$, $P < 0.001$).

Conclusions: The present research findings hold both theoretical and practical significance, and can serve as a guide for healthcare practitioners and researchers to gain a deeper comprehension of the acceptance of digital disease management applications (DDMAs) among type 2 diabetes patients.

KEYWORDS

type 2 diabetes, mobile medical, TAM, intention to use, eHealth literacy theory

1. Background

The prevalence of diabetes and its complications have had a profound impact on global health. Over the past years, the number of diabetes patients has skyrocketed. In 2015, the number of diabetes patients worldwide reached 415 million, far exceeding the predicted number of 340 million by 2030 (1). The International Diabetes Federation reported that in 2019, the prevalence of diabetes was 9.3%, affecting 463 million people globally, with a projected rise to 10.9%, or 700 million people, by 2045 (2). Type 2 diabetes mellitus (T2DM) is predominant in adults, while children and adolescents also tend to develop it (3). Self-management plays a critical role in diabetes control, as it not only regulates the progression of the disease but also improves patients' health status and quality of life (4, 5). Therefore, people with diabetes should acquire more knowledge and skills related to the disease and enhance their ability to self-manage it (6, 7).

In the face of a large number of patients with various illnesses, traditional medical resources have become strained, and are unable to cater to the medical needs of all patients (8). Digital disease management applications (DDMAs) belong to the category of mobile health (mHealth) and have emerged as a new solution to address the aforementioned dilemma. These apps have a wide range of features including blood glucose recording, insulin management, diet carbohydrate calculation, medication reminder, doctor consultation, diet advice, and health knowledge (9–11). Numerous studies have demonstrated that diabetes patients can manage their illness through these digital applications, which can effectively regulate their health and reduce diabetes-related laboratory indicators (12, 13). Additionally, the use of mobile medicine has the potential to reduce the cost of managing the disease for patients (14). Moreover, especially in the context of public health emergencies such as the COVID-19 pandemic, the use of mobile medicine has reduced the influx of patients to hospitals, thereby reducing the risk of infection (15).

Despite the potential benefits offered by digital disease management applications and the increasing interest from patients, their implementation in practice remains limited (16). To promote the use of digital disease management, scholars in many countries have investigated the factors that affect patients' willingness to use mobile medicine. For example, British scholars Lee et al. have explored the attitude of diabetes patients toward the use of mobile medicine and found that technical considerations, service awareness, and empowerment are the main factors that affect their use (17). Iranian scholar Rangraz Jeddi et al. have investigated the use of smartphone apps to manage diseases in patients with T2DM and found that younger participants were more interested in using such apps (18). However, due to differences in national conditions between developed and developing countries, the influencing factors for the willingness to use digital disease management applications are different. China, being a developing country, has limited evidence on the influencing factors for T2DM patients' willingness to use such applications.

TAM constitutes an information technology framework designed to elucidate user adoption and engagement with emerging technologies. The model postulates that an individual's intention and conduct pertaining to technology adoption hinges upon their perception of the technology's usefulness and ease of use (19). Perceived usefulness and perceived ease of use constitute pivotal determinants within the technology acceptance model, which indirectly shapes users' inclination to adopt through their attitudes toward usage (20). Presently, the Technology Acceptance Model finds application in numerous domains, particularly in the realms of social media (21) and the Internet (22). Evidently, this model serves as a common tool for exploring the factors that impact users' willingness to engage with Internet technologies, making it equally applicable to digital disease management applications as part of the Internet technology landscape. The perceived risk theory posits that every consumer transaction carries a certain degree of risk, implying that the dimension of perceived risk is context-dependent (23) and is now gradually being extended to health-related contexts (24). Furthermore, this theory lends itself to the examination of behavioral intentions regarding usage. Scholars following the Norman perspective define eHealth literacy as the capacity to locate, comprehend, and assess health-related information derived from electronic sources and apply this information to address physical health concerns (25). It has been postulated that eHealth literacy exerts a positive influence on the willingness to embrace mHealth (26). This theory also finds relevance in the present study, emphasizing the significance of eHealth literacy among individuals with type 2 diabetes, and investigating whether it directly impacts the intention to utilize such technologies. Hence, this study endeavors to probe the usage intentions of T2DM patients and the factors that influence digital disease management applications, grounded in the patients' viewpoint and employing the technology acceptance model (TAM), perceived risk theory (PR), and eHealth literacy theory (E-HLT).

2. Methods

2.1. Study design

The current cross-sectional investigation was carried out in 2021 in the southwest region of China, utilizing the convenience sampling technique to select participants. The study included patients with T2DM who visited outpatient and inpatient services in three hospitals located in Chengdu. Inclusion criteria: ① individuals aged 18 years and above; ② Individuals diagnosed with type 2 diabetes in a secondary or tertiary healthcare facility, following the 1999 diagnostic criteria for diabetes established by the World Health Organization (WHO); ③ individuals possessing complete cognitive and behavioral abilities; ④ individuals who exhibited clear awareness, normal thinking, and expression abilities, and were willing to participate after giving informed consent. Patients who declined to participate were excluded from the study. A convenience sampling method was employed to select diabetic patients who fulfilled the eligibility criteria for participation in the survey. Following the formula for determining sample size in multifactor analysis, the sample size should be 15–20 times the number of variables influencing the analysis (27). In this study, there were a total of 28 measured variables, thus necessitating a minimum sample size of 420 cases, calculated as 15 times the number of variables. Considering a 20% allowance for invalid or

10.3389/fpubh.2023.1259158

missing questionnaires, a minimum of 525 questionnaires were distributed for the purpose of this research study.

2.2. Questionnaire

Our questionnaire consisted of three sections. The first section collected socio-demographic information from participants, such as age, gender, education level, city of residence, and monthly income. The second section utilized the Chinese version of the E-health literacy scale (eHEALS) instrument, which includes tests of application ability, judgment ability, and decision-making ability related to network health information and services. The scale has a Cronbach coefficient of 0.913, and the factor analysis load factor is between 0.692 and 0.869. The third section asked participants to provide their cognitive attitudes and willingness to use digital disease management applications, answering questions formulated by the different instrument items represented in the Technology Acceptance Model (TAM) and Perceived Risk Theory. The questionnaire was divided into five dimensions: perceived usefulness (4 items), perceived ease of use (4 items), perceived safety (4 items), attitude toward use (5 items), and intention to use (3 items). The questionnaire had good internal consistency, with Cronbach's a coefficient >0.8 for each dimension of the willingness to use questionnaire, a KMO value of 0.941 in the validity test, and 77.63% of the total variance explained cumulatively by the five dimensional factor rotations. To increase the research validity, five local experts with extensive experience in relevant fields supported the validation of the instrument items and offered wording modifications. Furthermore, 20 individuals with type 2 diabetes (T2DM) were randomly chosen for the pre-survey, and 20 valid questionnaires were collected. The reliability of the questionnaire related to awareness and willingness to use digital disease management applications was assessed using SPSS 22.0, indicating that the internal consistency of the questionnaire was satisfactory. Specifically, the Cronbach's α coefficient for the willingness to use questionnaire was found to be 0.944. Subsequently, the written language and the questions were refined based on the feedback received from the participants. Table 1 presents the questionnaire items employed in this study. Each item was measured using a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5).

2.3. Data collection

The present study received approval from the Ethics Committee of the First Affiliated Hospital of Chengdu Medical College (02020CYFYIRB-BA-129-F01). Wen Juan Xing, a Chinese online survey platform, was used to conduct the study. The researcher got written informed consent from the patients and thoroughly explained the aim of the research and the questionnaire's format to them before to the survey. Patients answered the surveys in the researcher's presence to guarantee clarity and correctness. A total of 570 questionnaires were distributed, with 559 valid questionnaires recovered, resulting in an effective recovery rate of 98.07%.

TABLE 1 Construct items of the instrument.

Instrument items		Questions
Perceived usefulness	PU1	Diabetes management APP can manage diabetes anytime and anywhere, which I think is very convenient.
	PU2	Using diabetes management APP can improve my effectiveness in diabetes management.
	PU3	Diabetes management APP can meet my needs for diabetes management.
	PU4	In conclusion, I think the use of diabetes management APP is very useful for diabetes management.
Perceived ease of use	PEOU1	I think it's easy for me to learn to use diabetes management APP.
	PEOU2	I don't think it will take me too much time to use diabetes management APP for diabetes management.
	PEOU3	I think it's easy to master diabetes management APP.
	PEOU4	In short, I think it is easy to use diabetes management APP to manage diabetes.
Perceived safety	PS1	I'm worried that using diabetes management APP will leak my personal information.
	PS2	I'm worried that my diabetes management APP password will be stolen, resulting in economic losses.
	PS3	I'm worried about the system vulnerability of the third-party platform of diabetes management APP.
	PS4	I believe that the use of diabetes management APP is reliable and safe.
Attitude toward using	ATT1	Compared with my peers, I am more willing to try diabetes management APP to manage diabetes.
	ATT2	Even if there are risks, I intend to use diabetes management APP to manage diabetes.
	ATT3	It is a wise choice to use diabetes management APP to manage diabetes.
	ATT4	I think using diabetes management APP to manage diabetes is in line with my lifestyle.
	ATT5	In general, I support the use of diabetes management APP to manage diabetes.
Intention to use	ITU1	I'm happy to try the management method of diabetes management APP.
	ITU2	I will recommend diabetes management APP to my friends and relatives.
	ITU3	I will actively use diabetes management APP to manage diabetes.

2.4. Statistical analysis

The collected data was subjected to statistical analysis using SPSS version 22.0 and AMOS version 23.0. Firstly, demographic data were evaluated in terms of frequency and percentage,

TABLE 2 Respondents' demographic information ($n = 559$).

		Frequency	Percentage (%)
Gender	Male	351	62.8
	Female	208	37.2
Age	≤45	83	14.8
	46–59	260	46.5
	≥60	216	38.6
Marital status	Married	510	91.2
	Unmarried	49	8.8
Place of residence	Urban	459	82.1
	Rural	100	17.9
Education level	Primary school and below	130	23.3
	Middle school diploma	230	41.1
	High school diploma	133	23.8
	Bachelor's degree and above	66	11.8
Monthly income (RMB)	Below 2,000	162	29.0
	2,001–5,000	227	40.6
	5,001–10,000	121	21.6
	Above 10,000	49	8.8
Occupation	Personnel of government and public institutions	46	8.2
	Enterprise employees	102	18.2
	farmer	79	14.1
	Self-employed businesses	130	23.3
	Retired	202	36.1
Duration of the disease	≤3	201	36.0
	4–10	197	35.2
	>11	161	28.8
Chronic complications of diabetes	Yes	378	67.6
	No	181	32.4
Smartphone proficiency	Very unskilled	44	7.9
	Not skilled	119	21.3
	commonly	192	34.3
	Relatively skilled	144	25.8
	Great facility	60	10.7

while the mean and standard deviation of each dimension of the digital disease management application cognition and willingness to use scale were calculated. AMOS 23.0 was used to create a structural equation model to study the path association between variables. The significance of the differences was evaluated using a bilateral test with a threshold of $P < 0.05$. The Bayesian Positive Free Sampling and Bootstrap methods were used to estimate the effect value and 95% confidence interval. Statistical significance was determined at a $P < 0.05$.

3. Results

3.1. Description of respondents

The sample population consisted of 62.8% males and 37.2% females, with 14.8% of respondents aged between 18 to 45 years, 46.5% between 46 to 59 years, and 38.6% over 60 years old. Among the respondents, the majority (64.4%) had attained a middle school education level or below, followed by those with a high school diploma (23.8%) and those with an

TABLE 3 Means, reliability, and convergent validity.

Construct	Items	Mean (SD)	Standardized factor loading	CR	AVE
Perceived usefulness	PU1	3.76 (1.71)	0.837	0.896	0.683
	PU2	3.69 (1.06)	0.828		
	PU3	3.63 (1.06)	0.801		
	PU4	3.69 (1.05)	0.840		
Perceived ease of use	PEOU1	3.14 (1.27)	0.878	0.900	0.695
	PEOU2	3.36 (1.19)	0.691		
	PEOU3	3.16 (1.27)	0.869		
	PEOU4	3.14 (1.25)	0.881		
Perceived safety	PS1	2.87 (1.34)	0.946	0.967	0.880
	PS2	2.92 (1.34)	0.958		
	PS3	2.82 (1.26)	0.953		
	PS4	3.27 (0.95)	0.894		
Attitude Toward Using	ATU1	3.35 (1.30)	0.793	0.858	0.551
	ATU2	3.04 (1.26)	0.802		
	ATU3	3.47 (1.13)	0.732		
	ATU4	3.21 (1.26)	0.787		
	ATU5	3.78 (1.12)	0.573		
Intention to use	ITU1	3.49 (1.21)	0.781	0.820	0.603
	ITU2	3.36 (1.17)	0.747		
	ITU3	3.46 (1.21)	0.800		

undergraduate degree or higher (11.8%). Furthermore, 40.6% of respondents reported a monthly income between RMB 2,001 and 5,000, with 29.0% earning <RMB 2000, 21.6% earning between RMB 5,001 and 10,000, and 8.8% earning over RMB 10,000. Most of the respondents (82.1%) resided in cities, with 71.2% reporting a T2DM diagnosis for <10 years, and 67.6% indicating the presence of chronic complications related to diabetes. Table 2 provides a detailed summary of the respondents' socio-demographic information.

3.2. Measurement model

The composite reliability (CR) and average variance extracted (AVE) were used to measure the reliability of each component. The CR values for the constructs in this study ranged from 0.820 to 0.967, exceeding the proposed threshold value of 0.70 (28) (Table 3). Similarly, the AVE values ranged from 0.551 to 0.880, exceeding the suggested threshold of 0.50 (29) (Table 3). These findings indicate that the constructs and measurements used in this study are reliable and well-constructed.

3.3. Structural model

3.3.1. Model fit

A structural model was developed using the maximum likelihood (ML) method to identify any explanatory relationships. The initial model included two paths that did not show a statistically significant difference in terms of the difference in intention to use: electronic health literacy and perceived safety. After considering various factors such as correction index, standardized residuals, path coefficient p-values, and removing insignificant paths and variables, a modified model with satisfactory fit was achieved. Whilst evaluating the goodness of fit indices, it is recommended to use more than one indicator to evaluate model fit (30). Apart from Goodness of Fit Index (GFI) which measured slight below 0.90, all fit indices fulfill the accepted values. Both Comparative Fit Index (CFI) and Normed Fit Index (NFI) were estimated at 0.959 and 0.946 and indicates good fit (31). The Root Mean Square Error of Approximation (RMSEA) measured below 0.10 and was considered a good fit (32). The fitness indices of the proposed theoretical model are presented in Table 4.

3.3.2. Analysis of influencing factors

Figure 1 and Table 5 present the results of the structural equation analysis. Perceived ease of use ($\beta = 0.511$, $P < 0.001$) and electronic health literacy ($\beta = 0.115$, $P < 0.05$) were found to predict perceived usefulness. Electronic health literacy ($\beta = 0.659$, $P < 0.001$) was found to predict perceived ease of use. Furthermore, perceived usefulness ($\beta = 0.137$, $P < 0.001$) and use attitude ($\beta = 0.825$, $P < 0.001$) were found to predict intention to use. In conclusion, the research model explained 43.4% of the variance in perceived ease of use, 35.2% of the variance in perceived usefulness, 68.7% of the variance in attitude toward using, and 87.2% of the variance in intention to use.

4. Discussion

This research emanates from China, delving into the inclination to engage with DDMAs among individuals afflicted with type 2 diabetes mellitus. The findings validate that all of the envisaged factors, barring perceived safety, exhibit robust and affirmative correlations with the proclivity to employ DDMAs among patients with type 2 diabetes mellitus. Notably, perceived usefulness and attitude toward utilization directly and positively impact the disposition to adopt, while the influence of perceived ease of use on willingness to use is mediated indirectly through perceived usefulness or attitude toward utilization. Furthermore, the impact of e-health literacy on willingness to use operates indirectly through perceived ease of use or perceived usefulness, thereby indirectly influencing the inclination to embrace these applications.

TABLE 4 Model fit.

Fitting index	CMIN/DF	GFI	AGFI	RMSEA	CFI	NFI
Result	3.951	0.845	0.814	0.073	0.951	0.946
Criteria	<3 excellent; <5 Acceptable	>0.9 excellent; >0.7 Acceptable	>0.9 excellent; >0.7 Acceptable	<0.05 excellent; <0.08 Acceptable	>0.9 excellent; >0.7 Acceptable	>0.9 excellent; >0.7 Acceptable

CMIN/DF, chi-square fit statistics/degree of freedom; GFI, goodness of fit index; AGFI, adjusted goodness of fit index; RMSEA, root mean square error of approximation; CFI, comparative fit index; NFI, normed fit index.

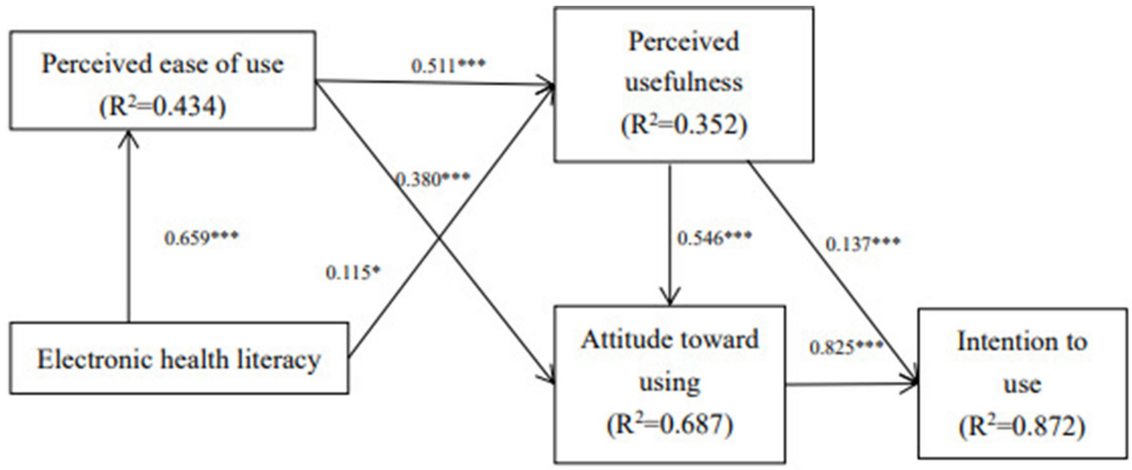


FIGURE 1
The results of the structural model ($^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

4.1. Willingness to use digital disease management applications

According to the study, the inclination of individuals with T2DM toward utilizing DDMAs is moderately to highly favorable, with scores exceeding 3 (out of 5) for each of the dimensions related to willingness to use, analogous to the discoveries of Mehbodniya (12) and Rangraz (18). Moreover, the study comprised 40% of the total population aged 60 years and above, who have a moderate to high intention to use DDMAs, consistent with the research conducted by Jaana M (33). This emphasis on the older population is justified by their significantly higher prevalence of T2DM compared to the general adult population, highlighting their heightened need for support from DDMAs. Additionally, this reflects the widespread presence of the internet and patients' confidence in the capability of DDMAs to facilitate effective disease management. The swift growth of mHealth in China over the years has led to reduced costs of utilization. Furthermore, China's policies have encouraged the advancement of "Internet + medicine" to ensure the availability of mobile healthcare benefits to all residents (34). MHealth has gained popular acceptance, particularly after COVID-19, making it easier for people with chronic diseases to manage their conditions and minimizing the need for hospital visits (35). Various channels have enabled patients with T2DM to become acquainted with or employ mHealth, gradually comprehending the benefits of this novel disease management model. Several avenues have facilitated the introduction and utilization of mobile health (mHealth) among patients with type 2 diabetes (T2DM), gradually imparting an understanding of the advantages associated with this innovative disease management approach. Consequently, there exists a necessity to further diminish the barriers for older individuals in accessing Internet-based healthcare services. Community and healthcare institutions should offer ample training opportunities to assist them in comprehending and mastering DDMAs (36). Simultaneously, DDMAs should be thoughtfully designed, considering the specific requirements and usability of this demographic, with a focus on maximizing ease of operation and addressing the disease management needs of this group.

4.2. Factors affecting willingness to use DDMAs

The Technology Acceptance Model (TAM) is a well-established and reliable research model used for forecasting user acceptance and adoption of health information technologies (21). This study is similar to previous studies applying the TAM in the context of general mobile medicine. The findings indicate that perceived ease of use and perceived usefulness are critical patient-centered factors that positively impact the willingness of patients with type 2 diabetes to use DDMAs, thereby validating the TAM model (37). These results are consistent with the findings of other studies that have explored the adoption of mHealth based on TAM, including Wang et al. (38), Nezamdoust et al. (39), and Harakeh et al. (37). Additionally, the research of Breil et al. (40) and

Palos-Sanchez et al. (41). also supports the notion that perceived usefulness bears a direct impact on the inclination to adopt, while perceived ease of use can affect the willingness to use through its influence on perceived usefulness. Additionally, these two factors can synergistically interplay to shape attitudes toward utilization, consequently influencing the predisposition to adopt. Perceived usefulness refers to the subjective perception of how much patients believe DDMAs can benefit their disease management. T2DM patients are more likely to use DDMAs to manage their disease when they believe that these apps can help manage their disease (42). This underscores the necessity for DDMA developers to meticulously enhance application functionalities, aligning them with patients' requirements for effective disease management. By augmenting the user experience, the goal is to catalyze the wider adoption of DDMAs. Perceived ease of use refers to the degree to which patients find the diabetes management app effortless to use. T2DM patients are more positive about using the diabetes management app when they feel it is less difficult to operate and the setting functions are easier to understand (43). Particularly in the context of China, where individuals with type 2 diabetes are predominantly found among the middle-aged and older adult demographics (44), the formulation of disease management applications necessitates a meticulous consideration of the physiological decline associated with older adult patients. Factors like visual impairment and reduced manual dexterity should be attentively taken into account during the design process (45). These applications ought to be structured for effortless installation and operation, with an interface that emphasizes simplicity. Moreover, provisions for specialized assistance modes can also be incorporated to cater to the needs of older adult users. The path coefficient between perceived usefulness and use attitude is greater than that between perceived ease of use and use attitude. This indicates that patients attribute higher importance to the concrete advantages offered by disease management applications in comparison to the user-friendliness of these applications. Moreover, perceived usefulness directly impacts patients' willingness to use DDMAs, indicating that T2DM patients are highly concerned about the effectiveness of glucose management brought about by using DDMAs (46, 47). But the positive impact of perceived ease of use on perceived usefulness cannot be ignored. Patients may doubt the usefulness of mobile medical technology if they find it inconvenient or challenging to use, which can affect their attitude toward using it (48, 49).

This study further corroborates that electronic health literacy indirectly affects patients' willingness to use DDMAs. The findings demonstrate that electronic health literacy has a favorable impact on patients' perceived usefulness and perceived ease of use, aligning with previous research (50). Song et al. (51) posited e-health literacy as a potential variable influencing the sustained utilization of mHealth services by patients. They proposed that individuals possessing higher e-health literacy are more adept at effectively engaging with DDMAs for the purpose of managing their health conditions. Consequently, this elevated e-health literacy contributes to an augmented perception of the utility and user-friendliness of DDMAs (52). However, the path coefficient between e-health literacy and perceived ease of use is substantially greater than that of e-health literacy and perceived usefulness, signifying a more significant impact of e-health literacy

TABLE 5 Structural equation model analysis results.

Path	Path coefficients	S.E	C.R.	*P* value
PU→ ITU	0.137	0.043	3.915	<0.001
PEOU→ PU	0.511	0.039	10.250	<0.001
PU→ ATU	0.546	0.046	14.755	<0.001
PEOU→ ATU	0.380	0.034	10.986	<0.001
ATU→ ITU	0.825	0.039	20.852	<0.001
E-heals→ PU	0.115	0.035	2.374	0.018
E-heals→ PEOU	0.659	0.033	18.485	<0.001

PU, Perceived usefulness; PEOU, Perceived ease of use; ATU, Attitude toward using; ITU, Intention to use; E-heals, Electronic health literacy.

on perceived ease of use. This parallels the discoveries made by Chisolm et al. (53). Enhancing users' e-health literacy to streamline their online access to requisite health information becomes notably feasible when confronted with novel web-based offerings. With the advent of the internet, patients with elevated electronic health literacy have greater access to pertinent health information and management services, thus facilitating their adoption of relevant mobile health services (54). Consequently, the electronic health literacy of patients emerges as a pivotal concern, demanding the implementation of strategies to elevate patients' proficiency in electronic health literacy. This equivalently signifies the necessity for relevant authorities to place substantial emphasis on providing technical educational support to older adult individuals grappling with diabetes. Embracing novel disease management paradigms and enhancing the efficacy and potential of disease management strategies are also imperative considerations. In contrast to prior research, this study integrates the perceived risk theory into the analysis. Surprisingly, the results show that perceived safety has no impact on patients' attitudes or willingness to use DDMAs (55, 56). Despite the complexities and potential security risks associated with online information, including online fraud and data breaches, patients still exhibit a willingness to use these digital tools (57, 58). One possible explanation for this is that older adult Chinese patients with T2DM, who make up a significant portion of the population, are less concerned with their online information exposure than younger people. During the questionnaire survey, some older adult patients did not believe that personal information leakage would impact their daily life or cause economic losses, which suggests a lack of awareness regarding information security (59). Additionally, diabetes management mHealth primarily involves recording patients' blood glucose, medication, diet, and exercise, with less involvement in their financial information. As a result, patients' perceived security is insufficient to affect their willingness to use such tools. Nonetheless, within the domain of perceived security, the entry score reveals a lower perceived security rating. This underscores that patients continue to harbor concerns regarding potential privacy breaches and the security of their personal assets arising from the utilization of DDMAs. Therefore, developers of DDMAs are compelled to intensify their focus on enhancing application security to effectively safeguard patients' personal privacy.

4.3. Theoretical and practical implications

In this study, a theoretical model was constructed to investigate the factors that influence the willingness of T2DM patients to use DDMAs. The results showed that e-health literacy had an impact on patients' perceived ease of use and perceived usefulness toward DDMAs. This provides new insights into improving the willingness of patients to use DDMAs. In contrast, the perceived safety may not significantly affect patients' intention to use DDMAs, suggesting that it is not the main concern for patients who decide to adopt DDMAs. However, due to the unique characteristics of the study population, future studies should also consider perceived security in their measurements.

Past scholarly inquiries within this realm have predominantly centered on evaluating the effectiveness of patients' engagement with disease management applications, resulting in fewer examinations of their inclination to embrace these tools. This is notably prominent when considering patients dealing with type 2 diabetes in the Chinese context. This study broadens the scope of the field, presenting significant and relevant insights that hold practical value for mHealth designers and educators in the field of diabetes health. By identifying the impact of e-health literacy on perceived usefulness and perceived ease of use, as well as the impact of perceived usefulness and perceived ease of use on use attitude, the findings can inform the development of personalized and user-friendly DDMAs services that cater to individual users' needs and preferences. Designers can reduce the complexity of DDMAs operation by providing guidance and prompts at each step of the process, which can be particularly helpful for older adult diabetes patients. Although the research did not find a significant impact of perceived security on user willingness to use, patients' low perceived security scores indicate that personal privacy protection should still be a consideration in DDMAs design. DDMAs providers should specify how personal information is used and ensure the safety of all personal data collected.

5. Limitations and future work

This study has some limitations that should be acknowledged. Primarily, the exclusivity of the survey's focus on Chinese respondents, coupled with the sole reliance on quantitative research methodologies, might have contributed to an aspect of incompleteness within the findings. Subsequent research endeavors could consider amalgamating qualitative and quantitative methodologies to offer a more comprehensive understanding of patients and yield a richer repository of information. As cultural differences may affect social norms, as well as users' perceptions and attitudes, it is crucial to confirm the research findings in other cultural contexts before generalizing them. Thus, we suggest conducting a cross-cultural comparison of the factors that influence the willingness to use DDMAs. Secondly, as the actual adoption rate of DDMAs in China is currently low, the results of this study may only be valid for predicting patients' behavior at this stage. Therefore, the population should be monitored in the future. Thirdly, the study only examined patients' willingness to use DDMAs, and further research could explore their actual usage behavior. Lastly, the evaluation of patients' smartphone proficiency rested on self-reports, potentially engendering subjectivity in the research outcomes. To enhance the precision of results, future studies could incorporate objective tools for assessing smartphone proficiency.

6. Conclusions

Our study delved into the impact of patient-centered factors on the patients' inclination to adopt DDMAs, and integrated the technology acceptance model (TAM), perceived risk theory, and e-health literacy to substantiate the acceptance model. Our complete model describes the variation in DDMA use willingness of 87.2%. Precisely, patient-centered factors exert diverse influences on their attitudes and behavioral propensities to utilize DDMAs. Specifically, patient-centered elements wield a diverse range of impacts on their attitudes and inclinations toward utilizing DDMAs. Among these, perceived usefulness and attitude toward utilization emerge as pivotal determinants of the propensity to adopt. For both diabetes health educators and developers of DDMAs, careful attention should be accorded to these determinants. This entails catering to patients' exigencies for disease management and perpetually refining and enhancing application functionalities. During the preliminary phases of DDMA design, a profound exploration of the type 2 diabetes patient group becomes imperative. This can be accomplished through surveys, interviews, and other modes to glean genuine patient requirements and precisely delineate the trajectory of functional development. Likewise, consulting medical professionals and pertinent guidelines is essential to elucidate the disease management focal points. Subsequent to this, functionalities can be meticulously set, guided by scientific principles and methodologies to ensure the platform's empirical soundness. Moreover, the indirect effects stemming from perceived ease of use and e-health literacy warrant substantial attention. Perceived ease of use considerably influences perceived usefulness and attitudes toward utilization. Simultaneously, e-health literacy exerts its influence on perceived ease of use. Hence, patient habits concerning DDMA usage demand consideration. Particularly for older adult type 2 diabetes patients, employing more intuitive formats such as imagery, comics, and videos can alleviate reading complexities and amplify engagement. Furthermore, our study underscores the significance of extending technological education support to type 2 diabetes patients. This holds true especially for the substantial portion of older adult or less formally educated patients. By facilitating patients' seamless integration into the trajectory of information technology advancement, they can efficaciously partake in and reap the benefits of technological development.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

MZ: Formal analysis, Investigation, Methodology, Software, Validation, Writing—original draft. HZ: Investigation, Methodology, Writing—original draft. RZ: Investigation, Writing—review and editing. HY: Investigation, Data curation, Writing—original draft. MC: Investigation, Methodology, Data curation, Writing—original draft. XW: Investigation, Methodology, Data curation, Writing—original draft, Writing—review and editing. ZL: Methodology, Supervision, Validation, Data curation, Formal analysis, Writing—review and editing. ZX: Funding acquisition, Methodology, Supervision, Data curation, Formal analysis, Validation, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Clinical Medicine Research Centre for Geriatric Diseases Open Subjects (Chengdu Medical College-NanBu People's Hospital Joint Fund Project) 2022LHNBSYB-05, Chengdu Medical College—Chengdu Pidu District People's Hospital Joint Research Fund Nursing Special Project (2021LHHL-03), and 2022 Chengdu Philosophy and Social Science Planning Project (2022BZ148).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract.* (2014) 103:137–49. doi: 10.1016/j.diabres.2013.11.002

2. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. Idf diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract.* (2018) 138:271–81. doi: 10.1016/j.diabres.2018.02.023

3. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol.* (2018) 14:88–98. doi: 10.1038/nrendo.2017.151

4. Hildebrand JA, Billimek J, Lee JA, Sorkin DH, Olshansky EF, Clancy SL, et al. Effect of diabetes self-management education on glycemic control in latino adults with type 2 diabetes: a systematic review and meta-analysis. *Patient Educ Couns.* (2020) 103:266–75. doi: 10.1016/j.pec.2019.09.009

5. Emery KA, Robins J, Salyer J, Thurby-Hay L, Djira G. Type 2 diabetes self-management variables and predictors. *Clin Nurs Res.* (2022) 31:1250–62. doi: 10.1177/10547738211067322

6. Afroz A, Alramadan MJ, Hossain MN, Romero L, Alam K, Magliano DJ, et al. Cost-of-illness of type 2 diabetes mellitus in low and lower-middle income countries: a systematic review. *BMC Health Serv Res.* (2018) 18:972. doi: 10.1186/s12913-018-3772-8

7. Zurita-Cruz JN, Manuel-Apolinar L, Arellano-Flores ML, Gutierrez-Gonzalez A, Najera-Ahumada AG, Cisneros-González N. Health and quality of life outcomes impairment of quality of life in type 2 diabetes mellitus: a cross-sectional study. *Health Qual Life Outcomes.* (2018) 16:94. doi: 10.1186/s12955-018-0906-y

8. Geng J, Chen X, Bao H, Qian D, Shao Y, Yu H. Patients' preferences for health insurance coverage of new technologies for treating chronic diseases in china: a discrete choice experiment. *BMJ Open.* (2020) 10:e038051. doi: 10.1136/bmjopen-2020-038051

9. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. *Diabetologia.* (2019) 62:877–87. doi: 10.1007/s00125-019-4864-7

10. Alfonsi JE, Choi EEY, Arshad T, Sammott SS, Pais V, Nguyen C, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial. *JMIR mHealth and uHealth.* (2020) 8:e22074. doi: 10.2196/22074

11. Buss VH, Varnfield M, Harris M, Barr M. A mobile app for prevention of cardiovascular disease and type 2 diabetes mellitus: development and usability study. *JMIR Human Fact.* (2022) 9:e35065. doi: 10.2196/35065

12. Mehbodniya A, Suresh Kumar A, Rane KP, Bhatia KK, Singh BK. Smartphone-based mhealth and internet of things for diabetes control and self-management. *J Healthc Eng.* (2021) 2021:2116647. doi: 10.1155/2021/2116647

13. Nkhoma DE, Soko CJ, Bowrin P, Manga YB, Greenfield D, Househ M, et al. Digital Interventions Self-Management Education For Type 1 And 2 Diabetes: A Systematic Review And Meta-Analysis. *Comput Methods Programs Biomed.* (2021) 210:106370. doi: 10.1016/j.cmpb.2021.106370

14. He Q, Zhao X, Wang Y, Xie Q, Cheng L. Effectiveness of smartphone application-based self-management interventions in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. *J Adv Nurs.* (2022) 78:348–62. doi: 10.1111/jan.14993

15. Khoshrounejad F, Hamednia M, Mehrjerd A, Pichaghsaz S, Jamalirad H, Sargolzaei M, et al. Telehealth-based services during the covid-19 pandemic: a systematic review of features and challenges. *Front Public Health.* (2021) 9:711762. doi: 10.3389/fpubh.2021.711762

16. Quaosar G, Hoque MR, Bao Y. Investigating factors affecting elderly's intention to use m-health services: an empirical study. *Telemed J E Health.* (2018) 24:309–14. doi: 10.1089/tmj.2017.0111

17. Lee PA, Greenfield G, Pappas Y. Patients' perception of using telehealth for type 2 diabetes management: a phenomenological study. *BMC Health Serv Res.* (2018) 18:549. doi: 10.1186/s12913-018-3353-x

18. Rangraz Jeddi F, Nabovati E, Hamidi R, Sharif R. Mobile phone usage in patients with type II diabetes and their intention to use it for self-management: a cross-sectional study in Iran. *BMC Med Inform Decis Mak.* (2020) 20:24. doi: 10.1186/s12911-020-1038-y

19. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. *J Med Internet Res.* (2019) 21:e11604. doi: 10.2196/11604

20. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. *J Biomed Inform.* (2010) 43:159–72. doi: 10.1016/j.jbi.2009.07.002

21. Nadal C, Sas C, Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. *J Med Internet Res.* (2020) 22:e17256. doi: 10.2196/17256

22. Liu D, Li Q, Han S. Using extended technology acceptance model to assess the adopt intention of a proposed iot-based health management tool. *Sensors (Basel).* (2022) 22:16. doi: 10.3390/s22166092

23. Wei Y, Wang C, Zhu S, Xue H, Chen F. Online purchase intention of fruits: antecedents in an integrated model based on technology acceptance model and perceived risk theory. *Front Psychol.* (2018) 9:1521. doi: 10.3389/fpsyg.2018.01521

24. Xie C, Jia S, He C. An empirical study on the factors affecting elderly users' continuance intention of shared nurses. *Risk Manag Healthc Policy.* (2020) 13:1849–60. doi: 10.2147/RMHP.S261827

25. Norman CD, Skinner HA. Ehealth literacy: essential skills for consumer health in a networked world. *J Med Internet Res.* (2006) 8:e9. doi: 10.2196/jmir.8.2.e9

26. Sun H, Qian L, Xue M, Zhou T, Qu J, Zhou J, et al. The relationship between ehealth literacy, social media self-efficacy and health communication intention among Chinese nursing undergraduates: a cross-sectional study. *Front Public Health.* (2022) 10:1030887. doi: 10.3389/fpubh.2022.1030887

27. Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. *Chest.* (2020) 158:S12–s20. doi: 10.1016/j.chest.2020.03.010

28. Deng Q, Zeng Z, Zheng Y, Lu J, Liu W. Predictors of physicians' intentions to use clinical practice guidelines on antimicrobial in tertiary general hospitals of china: a structural equation modeling approach. *Antimicrob Resist Infect Control.* (2021) 10:97. doi: 10.1186/s13756-021-00966-z

29. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educ Psychol Meas.* (2013) 76:913–34. doi: 10.1177/0013164413495237

30. Soon JM. Structural modelling of food allergen knowledge, attitude and practices among consumers in malaysia. *Food Res Int.* (2018) 111:674–81. doi: 10.1016/j.foodres.2018.06.001

31. MacCallum RC, Hong S. Power analysis in covariance structure modeling using Gfi and Agfi. *Multivariate Behav Res.* (1997) 32:193–210. doi: 10.1207/s15327906mbr3202_5

32. Soon JM, Wahab IRA, Hamdan RH, Jamaludin MH. Structural equation modelling of food safety knowledge, attitude and practices among consumers in Malaysia. *PLoS ONE.* (2020) 15:e0235870. doi: 10.1371/journal.pone.0235870

33. Jaana M, Paré G. Comparison of mobile health technology use for self-tracking between older adults and the general adult population in canada: cross-sectional survey. *JMIR mHealth and uHealth.* (2020) 8:e24718. doi: 10.2196/24718

34. Gong L, Jiang H, Wu X, Kong Y, Gao Y, Liu H, et al. Exploring users' health behavior changes in online health communities: heuristic-systematic perspective study. *Int J Environ Res Public Health.* (2022) 19:18. doi: 10.3390/ijerph191811783

35. Wu J, Xie X, Yang L, Xu X, Cai Y, Wang T, et al. Mobile health technology combats Covid-19 in China. *J Infect.* (2021) 82:159–98. doi: 10.1016/j.jinf.2020.07.024

36. Li W, Shen S, Yang J, Guo J, Tang Q. Determinants of satisfaction with hospital online appointment service among older adults during the covid-19 pandemic: a cross-sectional study. *Front Public Health.* (2022) 10:853489. doi: 10.3389/fpubh.2022.853489

37. Harakeh Z, Van Keulen H, Hogenelst K, Otten W, De Hoogh IM, Van Empelen P. Predictors of the acceptance of an electronic coach targeting self-management of patients with type 2 diabetes: web-based survey. *JMIR Format Res.* (2022) 6:e34737. doi: 10.2196/34737

38. Wang H, Zhang J, Luximon Y, Qin M, Geng P, Tao D. The determinants of user acceptance of mobile medical platforms: an investigation integrating the Tpb, Tam, and patient-centered factors. *Int J Environ Res Public Health.* (2022) 19:17. doi: 10.3390/ijerph191710758

39. Nezamdoust S, Abdekhoda M, Rahmani A. Determinant factors in adopting mobile health application in healthcare by nurses. *BMC Med Inform Decis Mak.* (2022) 22:47. doi: 10.1186/s12911-022-01784-y

40. Breil B, Salewski C, Apolinário-Hagen J. Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. *JMIR Cardio.* (2022) 6:e31617. doi: 10.2196/31617

41. Palos-Sanchez PR, Saura JR, Rios Martin M, Aguayo-Camacho M. Toward a better understanding of the intention to use mhealth apps: exploratory study. *JMIR mHealth and uHealth.* (2021) 9:e27021. doi: 10.2196/27021

42. Akther T, Nur T, A. Model of factors influencing Covid-19 vaccine acceptance: a synthesis of the theory of reasoned action, conspiracy theory belief, awareness, perceived usefulness, and perceived ease of use. *PLoS ONE.* (2022) 17:e0261869. doi: 10.1371/journal.pone.0261869

43. Kamide H. The effect of social cohesion on interest, usefulness, and ease of use of a driving assistance system in older adults. *Int J Environ Res Public Health.* (2021) 18:21. doi: 10.3390/ijerph182111412

44. Mo D, Liu L, Chen Y, Yang J, Li J, Wood R, et al. Updating characteristics of type 2 diabetes mellitus patients in china: surveys in year 2008 and 2011-12. *Value Health.* (2014) 17:A756. doi: 10.1016/j.jval.2014.08.231

45. Roman MA, Rossiter HB, Casaburi R. Exercise, ageing and the lung. *Eur Respir J.* (2016) 48:1471–86. doi: 10.1183/13993003.00347-2016

46. Padilha JM, Machado PP, Ribeiro AL, Ribeiro R, Vieira F, Costa P. Easiness, usefulness and intention to use a mooc in nursing. *Nurse Educ Today.* (2021) 97:104705. doi: 10.1016/j.nedt.2020.104705

47. Althumairi A, Hariri B, Aljabri D, Aljaffary A. Patient acceptance and intention to use e-consultations during the Covid-19 pandemic in the Eastern Province of Saudi Arabia. *Front Public Health.* (2022) 10:896546. doi: 10.3389/fpubh.2022.896546

48. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. *Healthcare (Basel, Switzerland).* (2021) 9:3. doi: 10.3390/healthcare9030357

49. Zhang Z, Xia E, Huang J. Impact of the moderating effect of national culture on adoption intention in wearable health care devices: meta-analysis. *JMIR mHealth and uHealth.* (2022) 10:e30960. doi: 10.2196/30960

50. Schrauben SJ, Appel L, Rivera E, Lora CM, Lash JP, Chen J, et al. Mobile health (Mhealth) technology: assessment of availability, acceptability, and use in Ckd. *Am J Kidney Dis.* (2021) 77:941–50.e1. doi: 10.1053/j.ajkd.2020.10.013

51. Song T, Deng N, Cui T, Qian S, Liu F, Guan Y, et al. Measuring success of patients' continuous use of mobile health services for self-management of chronic conditions: model development and validation. *J Med Internet Res.* (2021) 23:e26670. doi: 10.2196/26670

52. Nie L, Oldenburg B, Cao Y, Ren W. Continuous usage intention of mobile health services: model construction and validation. *BMC Health Serv Res.* (2023) 23:442. doi: 10.1186/s12913-023-09393-9

53. Chisolm DJ, Hardin DS, McCoy KS, Johnson LD, McAlearney AS, Gardner W. Health literacy and willingness to use online health information by teens with asthma and diabetes. *Telemed J E Health.* (2011) 17:676–82. doi: 10.1089/tmj.2011.0037

54. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. *Int J Cardiol.* (2019) 292:280–2. doi: 10.1016/j.ijcard.2019.05.066

55. Klaver NS, van de Klundert J, van den Broek R, Askari M. Relationship between perceived risks of using mhealth applications and the intention to use them among older adults in the netherlands: cross-sectional study. *JMIR mHealth and uHealth.* (2021) 9:e26845. doi: 10.2196/26845

56. AlHadid I, Abu-Taieh E, Alkhawaldeh RS, Khwaldeh S, Masa'deh R, Kaabneh K, et al. Predictors for E-government adoption of sanad app services integrating Utaut, Tpb, Tam, trust, and perceived risk. *Int J Environ Res Public Health.* (2022) 19:14. doi: 10.3390/ijerph19148281

57. Tang Y, Gu L, Wang L. Deep stacking network for intrusion detection. *Sensors (Basel).* (2021) 22:1. doi: 10.3390/s22010025

58. Mehrtak M, SeyedAlinaghi S, MohsseniPour M, Noori T, Karimi A, Shamsabadi A, et al. Security challenges and solutions using healthcare cloud computing. *J Med Life.* (2021) 14:448–61. doi: 10.25122/jml-2021-0100

59. Jiang Y, Yang F. Association between internet use and successful aging of older chinese women: a cross-sectional study. *BMC Geriatr.* (2022) 22:536. doi: 10.1186/s12877-022-03199-w

frontiers | Frontiers in Big Data

TYPE Original Research
PUBLISHED 23 October 2023
DOI 10.3389/fdata.2023.1239017

OPEN ACCESS

EDITED BY
Plinio Pelegrini Morita,
University of Waterloo, Canada

REVIEWED BY
Jinran Wu,
Australian Catholic University, Australia
Shailesh Tripathi,
Tampere University of Technology, Finland

*CORRESPONDENCE
Nataliya Shakhovska
✉ nataliya.b.shakhovska@lpnu.ua

RECEIVED 16 June 2023
ACCEPTED 02 October 2023
PUBLISHED 23 October 2023

CITATION
Shakhovska N, Kaminskyi R and Khudoba B (2023) Experimental study and clustering of operating staff of search systems in the sense of stress resistance. *Front. Big Data* 6:1239017. doi: 10.3389/fdata.2023.1239017

Experimental study and clustering of operating staff of search systems in the sense of stress resistance

Nataliya Shakhovska*, Roman Kaminskyi and Bohdan Khudoba

Department of Artificial Intelligence, Lviv Polytechnic National University, Lviv, Ukraine

Introduction: The main goal of this study is to develop a methodology for the organization of experimental selection of operator personnel based on the analysis of their behavior under the influence of micro-stresses.

Methods: A human-machine interface model has been developed, which considers the change in the functional state of the human operator. The presented concept of the difficulty of detecting the object of attention contributed to developing a particular sequence of ordinary test images with stressor images included in it and presented models of the flow of presenting test images to the recipient.

Results: With the help of descriptive statistics, the parameters of individual box-plot diagrams were determined, and the recipient group was clustered.

Discussion: Overall, the proposed approach based on the example of the conducted grouping makes it possible to ensure the objectivity and efficiency of the professional selection of applicants for operator specialties.

KEYWORDS

human-machine, operator personnel, decision-making, clustering algorithm, stress management

1. Introduction

In modern information and search automated systems, the main role is played by the human-machine interface for decision-making based on the detection of objects of a given class in the information field. The human operator, as an element of such an interface, carefully reviews the provided information on the information field—the monitor screen, analyzing images of scenes, for example, pictures of territories, economic tables, abstract data, etc. Based on the obtained results of the analysis, the operator forms or chooses from a set of alternatives the appropriate decision, for which he mostly bears some responsibility. Such, let's call them primary solutions, can be both the final and the basis for further functions of this search engine. However, among the factors that in one way or another can negatively affect the quality of the work of this interface, one of the first places is stress in the form of a collection of micro stresses.

Due to the certain complexity of the provided images and their large number and variety, the functional state of a person becomes nervously tense, close to stressing, and in most cases this state is micro-stress. Individually, micro-stresses may be imperceptible to a person, but in the aggregate, their effect can be quite noticeable.

One such source of micro-stress in the field of information search can be a sudden change in the information flow, due to the lack of immediately required additional in-formation or significant masking of the searched object.

The purpose of this experimental study is to develop a methodology for organizing experiments on the selection of applicants for operator positions in information search and similar systems for processing images of scenes of real work situations.

The paper contribution is given as follows:

- The model of the human-machine interface as implementation of reaction to the detected object is proposed. This model is presented as dynamic systems, that al-lows the simulation of the next state of the operator personnel.
- A novel method for the analysis the influence of the flow of micro stressors on operator activity is developed. The sequence of images at discrete moments of time is generated. The operator is exposed on the monitor screen to a sequence of test images with objects of attention of a given class and which operator must implement the corresponding solution. The moments of their exposure and the decisions made by the operator are recorded, and their values are included in the research protocol. Next, hierarchical clustering is used for stress resistance clustering.

The effect on the reaction of a human operator to a deficit or an excess of provided information when making responsible decisions is studied in McEwen and Akil (2020). The model of the operator's stay and exit from stress is considered in Sahin et al. (2019), and the occurrence of neuropsychological overstrain is also explained here. The theoretical foundations of the psycho-diagnostic of stress and methods related to the psychological diagnostic procedure, ethics and stages of the psychological and diagnostic examination are given in Barak and Tsodyks (2023). Various aspects of the influence on the formation of the reliability of operator activity, including the working environment, the functional state of the operator and the intensity of work, are given in Brown and Anderson (2019).

The article Paul and Dykstra (2017) discusses the features of computer training of operators of continuous technological processes in comparison with other subject areas. The possibility of automating the measurement of the properties of the operator's attention using Schulte's tables is confirmed in Kryvenchuk et al. (2019). In Lee (2016) statistical methods of testing results analysis are considered, and the simplest and necessary procedures for statistical processing of knowledge testing results and test quality assessment methods are also presented. The research conducted in Bookstaber et al. (2014) summarized the structure of the stress test and analyzed the methods of generating shock scenarios.

The purpose of the article Battiston and Martinez-Jaramillo (2018) is to study the organizational aspects of stress testing and to determine the place and functions of the supervisory body in the process of stress testing. The methods of quantitative and integral assessment of personnel stress in the process of knowledge verification during attestation are analyzed in Iannello et al. (2017) and DeMenno (2022).

In Allen and Kessel (2002), three strategies for the development of detection systems are proposed, which are based on the principle of balance for purposeful improvement and design of detection systems. As shown in Chapelle (2009), the automation of the process of knowledge control, and the development of computer testing systems is an urgent task.

In work Chou (2000), an analysis of new information and computer training technologies for the development and use of existing interactive educational and training complexes of small arms was carried out. The article McGlohen and Chang (2008) discusses the problems of implementing testing using information and communication technologies. A systematic analysis of organizational problems of the development and application of computer testing technologies in higher education to control students' knowledge was carried out.

The article Zhu et al. (2019) presents an example of realistic image recognition using both manual and automated testing with a decision table. The article Ramanathan et al. (2016) discusses the principles of personnel selection and the most common traditional and non-traditional methods. The material for familiarization and assimilation of the personnel evaluation system to confirm his competence and responsibility is given in Raskin and Kircher (2014). The main theoretical principles and practical aspects of the application of psychophysiological testing to determine the level of professional suitability of polygraph operators are substantiated.

The conceptual contribution of the most valuable reviewed articles is given in Table 1.

The literature review analysis shows the lack of experimental studies related to person's stressful conditions during working time. In general, the research was carried out after the recipient had completed the work. By comparing the work process and psychophysiological indicators, researchers conclude the existence of a person's stressful state and analyze the results of the impact of existing irritants. For gaps filling, the model of human-machine interface will be developed in the paper.

2. Materials and methods

2.1. The model of the human-machine interface

One of the identified large numbers of various tasks of operator activity in computerized workplaces is the task of searching for objects of a given class on the images provided on the monitor and making appropriate decisions when receiving them. With this person and the computer function as a single system in the existing working environment. From the point of view of the general mathematical theory of the system, its descriptive model can be presented as follows. Let $\boldsymbol{P}(\boldsymbol{t})$ is some technological process that is managed by a human operator over time $[0, T_0]$ where $\boldsymbol{T_0} = \{\boldsymbol{t_i}:\boldsymbol{t_1} _< \boldsymbol{t_2}...\boldsymbol{t_N}\ \boldsymbol{i} \in \boldsymbol{T}\}$, and T - real time. The essence of its management is reduced to the functions of solving specific problems by the operator, for example, by identifying some object on the presented image of the scene and making the appropriate decision. An example of such processes is the search for objects of a given class on the image of the controlled territory, in the sequence of images of scenes, elimination of deviations in the parameters of the controlled process, search for the necessary data in document databases, editing of texts, etc., presented by their appearance on the monitor.

Change information about the process $\boldsymbol{P}(\boldsymbol{t})$ happens on the monitor screen in moments $\boldsymbol{t_i} \in \boldsymbol{T_0}$. The operator by focusing attention and visual analysis of the given set of

TABLE 1 References review.

References	The main contribution	Value
Paul and Dykstra (2017)	Cyber operations stress approaches comparison	The work describes the impact of stress on a cyber security worker. Studying the impact of frustration, mistakes on the work of an employee who uses a computer simulator or just a computer helps to understand various states, causes and consequences.
Zhu et al. (2019)	A method for testing intelligent Applications.	Testing smart applications provides insight into how to manually and automatically recognize certain elements. Which approaches are used by machine or human and how can automatic and manual approaches be combined.
Barak and Tsodyks (2023)	Mathematical models of learning in neurobiology	The article describes mathematical models and how they are used in neurobiology. Also, methods of applying data analysis to study the behavior of simulator operators.
Sahin et al. (2019)	Mathematical model of human behavior for simulating evacuation of buildings during emergencies	The author talks about modeling the behavior of people in evacuation situations. Evacuation gives stress to a person. Therefore, the simulation of such situations is related to the simulation of the microstress situation during the operator's work.

images $X=\{x_i{:}x_{1<}x_2...x_N\ i\in N\}$, using his skills, experience and knowledge tries to identify the desired object in these images. As a result of the search and detection of such an object, the operator then chooses, considering the possible consequences of the decision made, makes an appropriate decision from a set of decisions $Y=\{y_j{:}y_{1<}y_2...y_N\ j\in M\}$. Images of situations on the monitor contain all information about the state of what is observed with the help of the monitor and obviously the corresponding technical means of the controlled process $P(t)$.

In the process of its activity, the operator can be in different functional states $C=C(c_i)$, in psychophysiological states from normal to nervous overstrain. These states can be caused both by the influence of the external environment and by the provided information. In other words, in the absence of influence from the work environment, the functional state of the operator is a function of the provided image of the situation that occurred in the process $P(t)$, namely the function $C=C(c_i)$. It is this function that reflects the personality of the operator in the model and was used to study the activity of the operator in a state of micro stress.

The basis of this process is the concept of a system, that is a chain: "process → monitor → operator → computer → decision", which makes up the system together. In set-theoretic terms at this level, the system can be simple and quite naturally defined as a product $S\subset X\times Y$.

In the sense of model development plan the sets X and Y differ from each other in their position on the time axis, they are subject to the moments of time of their implementation. Elements $x\in X$ depends on time $X=\left\{x{:}x_i\equiv x(t_i),\ t_i\in T,\ i=\overline{1,N}\right\}$, but decision elements $Y=\left\{y{:}y_j\equiv y\left(t_j\right),\ t_j\in T,\ j=\overline{1,M}\right\}$ always delay by t_i, which corresponds to the total duration of search, detection, analysis, selection and decision-making until the moment of its implementation by the relevant team, i.e., $t_j=t_i+\Delta t_i$.

The average time of this delay can be used as an estimate of the efficiency of a particular operator during his shift. Over time, the efficiency and accuracy of the operator deteriorates. Each operator has its own optimal change and maximum possible change.

Since the human-machine interface is a dynamic system (Sahin et al., 2019), it makes sense to consider the behavior of the operator depending on the input information. In this general plan, we are talking about his reaction to the detected object, more precisely, the efficiency of detecting the object and making a decision. To establish the relationship between the input and output of the system, which refers to different moments of time, the general mathematical theory of systems introduces the concept of the system's response to an input stimulus (Paul and Dykstra, 2017). In other words, for each operating situation, for each image on the screen for an operator in a normal functional state, there are two such displays:

- making a decision

$$\overline{\rho}=\{\rho_t{:}C_t\times X_t\rightarrow Y_t\ \ \&\, t\in T\}$$

- transition to another state

$$\overline{\varphi}=\left\{\varphi_{tt'}{:}C_t\times X_{tt'}\rightarrow C_{t'}\ \&\, t,t'\in T\ \&\, t'>t\right\}.$$

These reflections can be explained as follows. Being at the moment of time $t\in T$ in the state of C_t, the operator perceives the image of the $x_t\in X$ and detects the desired object on it, analyzes the situation, and makes a decision of $y_t\in Y$, spending the necessary time normally, without changing his functional state (there is no neuropsychic tension).

If the provided image turns out to be such that it is not possible to immediately find the desired object, and the time limit for the search is limited, the operator can switch from a normal state to a nervous state, continuing the search for the object. If we assume that at the moment $t\in T$ the operator has not detected the object and feels that the exposure time is running out, he is aware of the complexity of the situation, and at this moment he can get a shock. This shock can be interpreted as stress or micro stress. The double time index means the time interval $t'>t$ during which the operator passed from a normal to a neuropsychical state.

In general, the model of the human-machine interface as such an information and search system can be represented by a tuple $S=\langle X,Y,C,\rho,\varphi,T\rangle$.

The use of the mathematical apparatus of set theory provides an opportunity to optimize the organization of research, since sets are presented quantitatively, the mappings of which are considered functionally and all of them are considered in time. In addition, based on such a model, it is possible to create a sufficient number of images that simulate the scenes of almost any real operator activity.

2.2. Method for the analysis the influence of the flow of micro stressors on operator activity

The localization of the object of attention on the test images is random, and the ratio of its size to the size of the test image practically amounts to several orders of magnitude (by area). In turn, individual fragments of the image, namely the locations of objects, can be both homogeneous, one-colored, smooth, and contain various other objects. These other objects are related to the background and can be larger or smaller than the object of attention and have an irregular shape. They significantly impair the visibility of the object of attention and are actually obstacles to the search. In the presence of such obstacles, the search process becomes more complicated, requires increased concentration of attention, creates for the operator a certain visual discomfort at first, which turns into additional neuropsychological stress or stress.

The sequence of such images at discrete moments of time reflects the situation observed by the operator, who should make the appropriate decision in case of detection an attention object. On a homogeneous background, the attention of object detection is reduced to a simple reaction of object recognition, however, in the presence of obstacles, noise, the time of such a reaction–searching for an object, will increase depending on their intensity. This, in turn, gives reason to say about the images that some of them have greater and others less difficulty in detecting the object.

On the other hand, individual characteristics, such as groups of operators, will show that the detection time of the object of attention on the same image will be different. Obviously, the more complex the image, the longer the average time to detect an object in this image. So, in this case, the category "complexity" is a subjective characteristic, but in general it has a certain dimension, namely the duration of the search time from the moment of providing the image to the moment of its discovery. The value of the indicator of the complexity of a particular image is the average value of the duration of the search, determined by the data obtained by different operators and in different experiments. The only requirement for this determination of the value of this indicator is the homogeneity of the group of "expert recipients" in terms of their training levels, skills, etc. In other words, the complexity of such images is determined by the time of searching and detecting the object of attention of a given class. The objectivity of such an assessment will be higher the more homogeneous (in terms of training) the group of "expert recipients" is.

2.3. Selection of test images

Let a group of operators be the subject of research, and the purpose of the research is to determine individual stress resistance. The preparation and conducting of the experiment were carried out as follows.

2.3.1. Creating a basic sequence of test images

At this stage, they create test images and form them into a sequence. The images in this sequence simulate the background of the working scene, on which the researcher manually or with the help of random numbers determines the coordinates of the places where the objects will be located. The placement of the object on some images should be quite easy to detect, and on others it should be difficult, i.e., hidden, but so that it can be found. An important point here is that all the details of the object were not closed, the object should look whole. After placing the objects, the received test images now have this status, they are formed into a sequence of scenes, mostly randomly. This sequence is shown to each of the group of recipients, fixing the time of processing each of the images by each recipient. At this stage, the duration of the experiment is determined, the number of test images in a sequence, from which the exposure time of the images is determined.

The duration of such an experiment can be determined depending on the contingent of recipients, for example, for professional operators, it is the duration of a shift or its part (the second half), and for young people, it is the duration of one or two academic hours. Depending on the complexity of the images of the scenes, exposure of the test image is up to 1–2 min. In the experiments conducted by the authors, the sequence included 180 test images with exposure of each of them for 30 s (Figure 1).

2.3.2. Determination of search complexity of test images

The sequence formed in this way is included in the process of professional selection or training of a group of recipients. The task is set and explained to them in advance, and if possible, an imitation of a real workplace and environment is carried out. After the "start" command, test images appear on the monitor screen, the operator, having detected the specified object, presses the corresponding key. When the test image appears, the stopwatch starts, and when the key is pressed, the operator stops it. As a result, the computer records the processing time of this image by the operator and resets the stopwatch.

All applicants work with the same sequence. As a result, for each test image, a sample of time values spent on them by each operator (recipient) will be obtained. The average value of the time spent by them on processing a particular test is actually a time indicator of their search complexity. Since these time indicators are quantitative values, for them it is possible to enter the corresponding scale of search complexity for test images of only this sequence and this group of recipients. Thus, each image will be characterized by its indicator. Figure 1 shows a diagram of times for searching for an object of attention on some subset of test images.

2.3.3. The sequence of stressful events

If we assume that the operator has been processing images of rather low complexity for some time, object detection is easy, and he has adapted (used to) this mode of presenting test images. If an image of high complexity suddenly appears and it was not possible to immediately detect the object, he will be forced to activate and concentrate his attention on the search, which in the mode of time shortage, for example, limiting the exposure to 20 s, will cause him

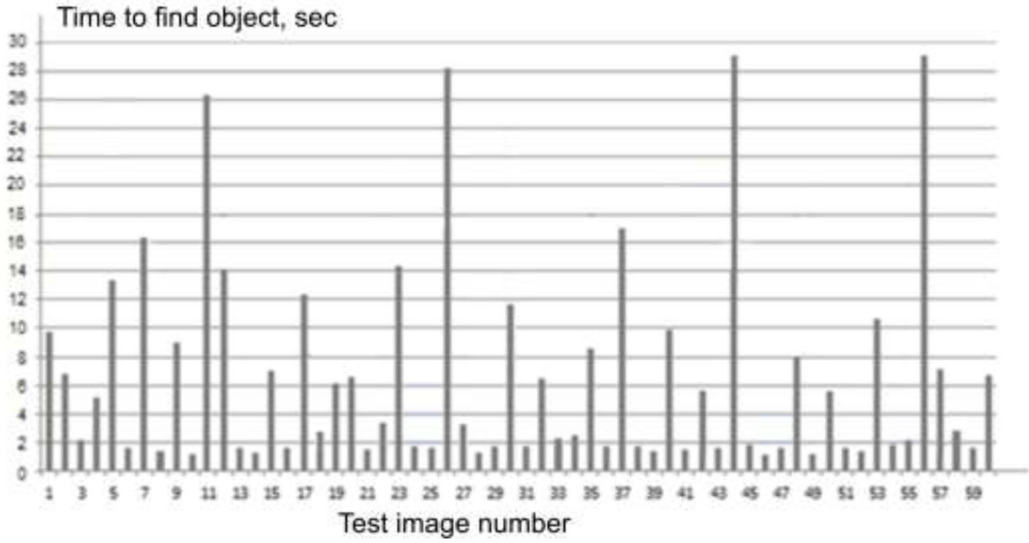


FIGURE 1
Search time for objects in the basic sequence.

(although maybe not) a micro-stress situation. In this regard, it is of interest to form such a sequence, which will consist of test images of low complexity and several images of high complexity are included in it (Figure 2).

Therefore, after the search complexity of the base sequence images is determined, the images with the lowest search complexity are selected, for example, the test images with the smallest search time of <2 s and the longest search time of >18 s. Next, a new basic sequence is made from the images that have minimal complexity (they can be repeated, but not one after the other). This new sequence includes, following the principle of rare events, test images whose search complexity exceeds 18 s. Obviously, all the images in this sequence must be randomly arranged. The view of a fragment of such a sequence is shown in Figure 2. The value of the duration of such micro stress can be represented by the appropriate mathematical model of general stress.

2.4. Options for presenting test images

In general, the organization of experimental research can be presented as follows. Let the human operator be exposed on the monitor screen to a sequence of test images with objects of attention of a given class and which the operator must identify and implement the corresponding solution. The moments of time of their exposure and the decisions made by the operator are fixed, and their values are included in the research protocol.

The main point here is the method of exposure of the test images. So, for the organization of experimental studies, the following three options can be specified for the method of providing test images to the operator on the monitor. Each of them includes two streams of rectangular pulses, synchronized in time along the pulse front. In the first version in Figure 3, the upper sequence corresponds to a regular stream of test images, marked by light rectangular pulses of the same duration and amplitude. Here the exposure of the tests is carried out at the same time intervals. The images of the situations are presented in a regular sequence (Figure 3).

The flow of test images in Figure 3 is a part of the experimental study. The lower sequence of dark impulses corresponds to the duration of the search for the object of attention, its detection as the object being searched for, and decision making. In other words, this sequence reflects the results of the human operator. Each slice of the pulse of this flow corresponds to a reaction-result as the moment of decision-making by the operator. If the operator did not find the desired object in the image, then the dark pulse will be absent in the bottom sequence. In such a graphic presentation, the upper and lower pulses are synchronized along the edge, at the moment of the appearance of the test image, the stopwatch is turned on, which is automatically turned off at the moment of the operator's decision.

In the second version, shown in Figure 4, the exposure of the test images is irregular and has different durations, but the pulse fronts of the upper and lower sequences are also synchronized. In the general case, when using an irregular flow, the duration of the pause between the upper pulses can be different. In addition, a variant is possible when the duration of the pulses of the upper and lower sequences are the same, and the duration of the pauses can be different. This means that after the operator makes a decision, the test image disappears, and the next one appears at the next moment.

In the third option, shown in Figure 5, the exposure of the image continues until the decision is made, after which the next image is displayed on the screen. The last option greatly reduces the duration of research with a limited image base since the duration of the exposure is equal to the duration of the search. In a certain sense, this option has a positive value, especially for laboratory research since it allows you to process more test images during the same experimental period.

The considered options for providing test images and fixing decisions are not equivalent from a psychological point of view. The fact is that with regular exposure, when the operator has very

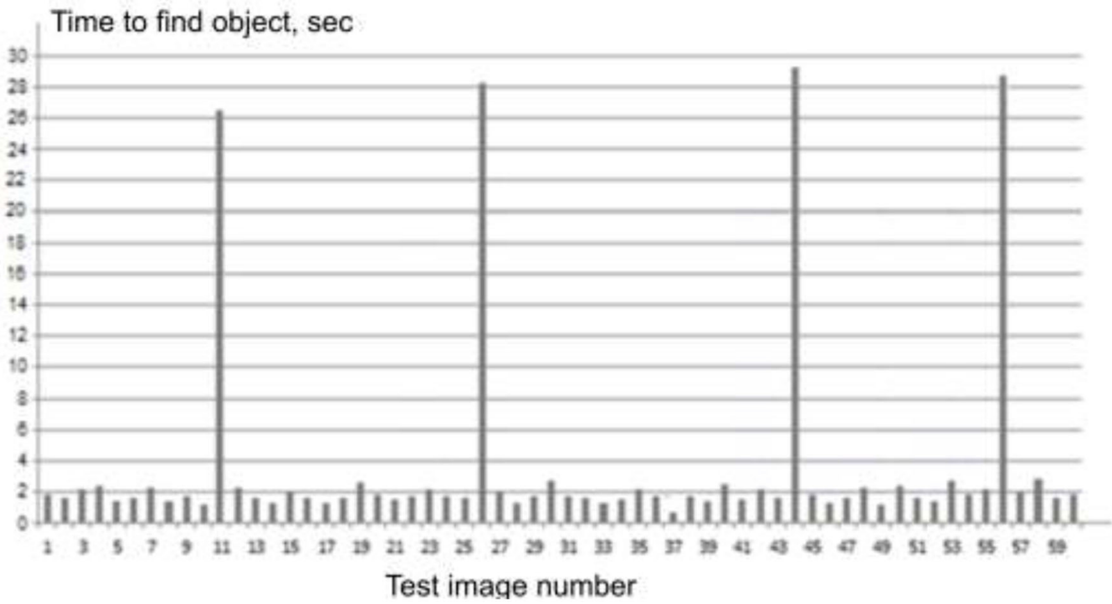


FIGURE 2
View of the experimental sequence.

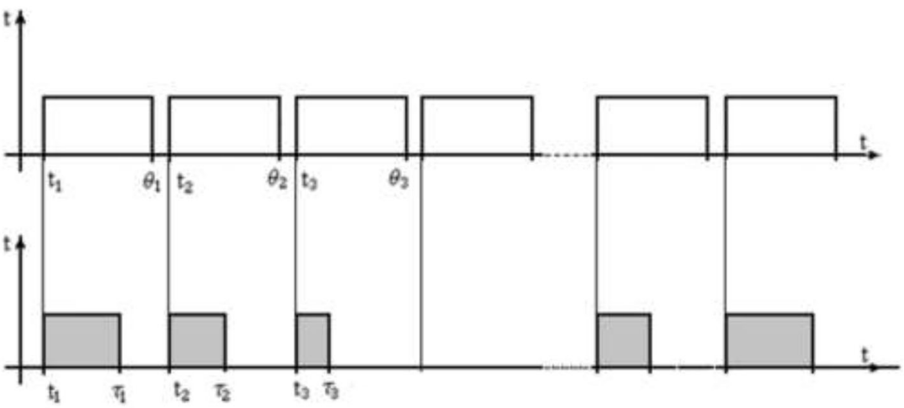


FIGURE 3
Regular stream of test images. Where $t_1, t_2, \ldots$ is starting point, and $\theta_1, \theta_2, \ldots$ are moments of exposure termination; $\tau_1, \tau_2, \ldots$ are moments of decision-making by the operator.

quickly identified the object and made a decision, he has some time left before the exposure of this image is completed. This causes a certain relaxation, and at the moment of the appearance of a new image, the operator sharply mobilizes attention and goes into a tense state. Such an irregular change in concentration and relaxation is a negative element in operator activity. On the other hand, when the next image is presented immediately after making a decision, constant nervous tension is created for the operator due to the need to maintain concentration all the time, which is also negative. Therefore, the choice of one or another mode of providing test images on the monitor screen is not trivial, especially for long-term experiments. Display of experimental research with the help of impulse flows provides a mathematical formulation of experiments and presentation of results by mathematical models of time series, in particular exposure of test images and object detection and decision-making.

3. Results

Nine recipients which are student operators took part in the experimental research. The experiment was conducted in two stages: in the first stage, the search complexity of the given set of test images was determined, and in the second stage, a new sequence was built. The initial volume of the sequence was 180 images of the tests, which were exposed on the monitor screen with a maximum duration of 30 s according to this rule. If the operator detected a given object on the test image in <30 s, the given test image was

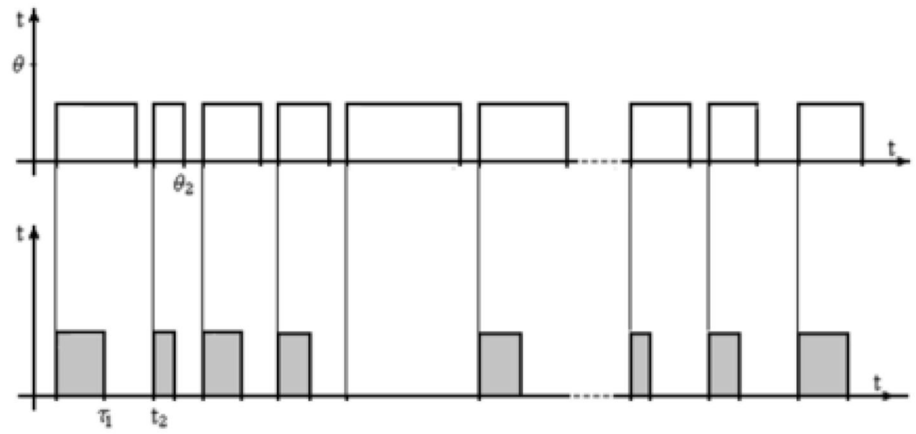


FIGURE 4
Irregular flow of test images.

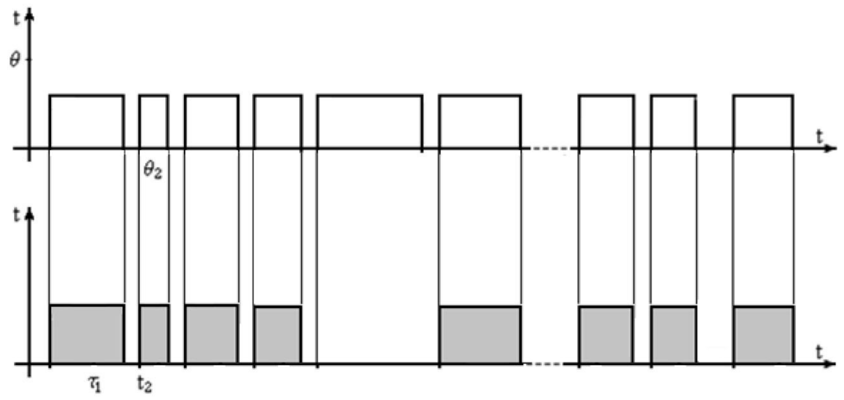


FIGURE 5
Irregular flow of decision-driven test images.

immediately replaced by the next one. In the event that the object was not detected on a given test image within 30 s, then it was replaced by the next one from this sequence.

In the sense of stress resistance, an assumption is made–if the detection time exceeds 2/3 of the exposure time or if the operator did not detect the object during the exposure time of the test image, then it can be assumed that it was in a state of micro stress. Although such categorization may be wrong, in this case, the authors consider it permissible.

The value of the time from the moment of the appearance of the test image to the moment of decision-making was used as an assessment of the search activity of the operator. In addition, to understand the stress state, the exposure time of test images is divided into two parts: the first 1–20 s and the second 20–30 s. The search time was recorded in milliseconds. The procedure for providing images corresponded to the third variant of the impulse flow model, that is, as soon as the operator made a decision about the detected object, the next one from the given sequence was immediately exposed. The volume of the provided sequence was 180 test images. The individual results of the experiments are given in the Table 2. This table presents only correct response.

The following parameters are used: *a*–the number of detected objects within the stress time; *b*–the number of missed objects; *x*–average stress time. Parameters: min, *q*1, *mode*, *q*3, max are parameters of the stress time interval boxplot diagram, namely extremes, quartiles, mode.

Using the parameters of the descriptive statistics of the results of the experiments for each of the recipients in Figure 6 shows constructed boxplot diagrams for the studied group of recipients, as a result of which the general structure of the distribution of individual data is determined.

As the resulting charts show, with the exception of recipient 3 and almost 9, within the interquartile range are symmetric, with recipients 1, 4, 5, and 8, and recipients 2, 6, and 7 having left- and right-sided asymmetries, respectively. This means that more of the rectangle has more variance. It can be assumed that intuitively feeling the completion of the exposure of the test image increases excitement and a certain, mostly significant, nervous tension. The authors assume the following regarding the asymmetry of distributions. For a left-sided distribution, the upper part of the rectangle is larger and corresponds to values greater than the mode value. Their spread between the mode and the third quartile is

TABLE 2 The individual results of the experiments.

Parameter	Operator s								
	1	2	3	4	5	6	7	8	9
a	10	16	11	18	10	15	11	8	15
b	17	10	14	13	11	9	7	13	77
x	24.8	24.1	23.2	23.9	24.2	24.5	24.0	24.4	24.8
Min	20.9	20.3	20.2	20.5	20.2	20.2	20.6	20.9	20.1
*q*1	23.2	21.6	20.6	22.4	23.2	21.9	21.6	21.8	23.8
Mode	24.5	24.3	22.6	23.0	24.2	25.2	24.7	23.5	25.0
*q*3	26.2	26.1	24.7	25.6	25.9	26.4	26.0	27.8	26.5
Max	29.1	29.3	29.6	29.8	27.7	29.4	26.9	28.4	28.3

TABLE 3 Results of the cluster analysis of the group of recipient-operators.

Combining objects	Association number	Distance between objects
Dendrogram parameters		
2 + 6	10	0.494
5 + 7	11	0.746
1 + 8	12	0.883
10 + 4	13	1.000
12 + 11	14	1.158
13 + 3	15	1.384
14 + 9	16	1.822
5 + 16	17	1.974

larger, and therefore it can be assumed that these recipients are characterized by a delayed reaction. For right-sided asymmetry, on the contrary, the lower part of the rectangle corresponds to the values of the search time, which are smaller than the values of the mode, they are characterized by haste of the reaction.

The obtained boxplot diagrams are individual statistical displays of the characteristics of the operators within the stress time interval. These individual characteristics, according to the given table, show the quality of search activity of operators, namely the number of missed and detected objects in the stressful time zone.

When constructing the boxplot diagram, the calculation of emissions was not carried out, since the interval of values here is limited by the condition of 20–30 s. From Figure 6, it can be pointed out that if we place individual boxplots relative to fashion in ascending order, we will get their rating (Figure 6).

During the laboratory study of the formed group of operators, the task of dividing them into subgroups based on close individual indicators arises. Given the small number of classification features and the number of recipients, hierarchical agglomerative cluster analysis was used to determine subgroups. The following individual values of the recipients were taken as signs: the number of detected objects within the stressful time; the number of missed objects; mean stress time, as well as the minimum and maximum time value in the stress time interval, the first and third quartile values, and the mode value. Therefore, the parameters of the boxplot diagram of the stress time interval are taken as the values of the classification features.

4. Discussion

For the clustering procedure, the values of the classification features were reduced to the interval [0, 1] and a flexible strategy was chosen, and the Euclidean metric was used to construct the proximity matrix. The features are normalized for each parameter. The result of the cluster analysis, in the form of dendrogram parameters, is shown in the Table 3 in Figure 7 on the left, which shows the results of combining objects and the distances by which the combination was performed, and the cluster analysis dendrogram itself is shown in this figure on the right.

Visual analysis of the dendrogram gives the following result.

1. At level 1.5, you can visually distinguish three clusters that include such recipients:

- cluster 1: it includes recipients 2, 6, 4, 3;
- cluster 2: it includes recipients 5, 7, 1, 8;
- cluster 3: it only includes recipient 9.

Moreover, the first and second clusters are similar to each other in terms of the homogeneity of recipients, as evidenced by the distance of associations between objects in them, which lies within 0.6–1.4.

2. Recipients vary greatly in their individual characteristics.

For example, the smallest distance between operators 2 and 6, which is equal to 0.5 on the scale of the dendrogram and is 25% percent of the total scale. Pairs 5 and 7 and 1 and 8 differ even more from each other and from pair 2 and 6. Recipients 4 and 3 are very different from pair 2 and 6 even though they all belong to the same cluster.

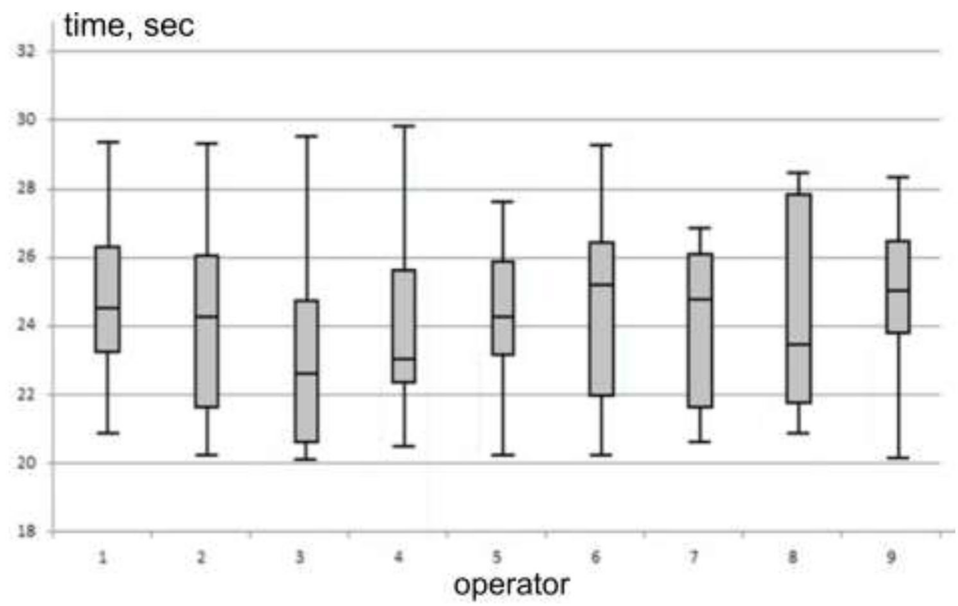


FIGURE 6
Image of individual boxplot diagrams.

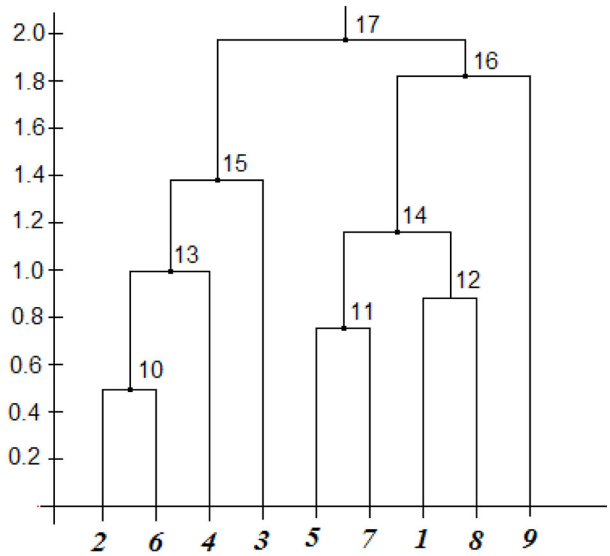


FIGURE 7
Results of the cluster analysis of the group of recipient-operators.

3. Cluster 3 is located separately and as far as possible from the first two, which is quite clear from the analysis of its performance indicators.
4. The distance between these clusters is almost the same and is in the range of 1.822 - 1.974, it can be argued that the subgroups are significantly different from each other.
5. If the division into clusters is carried out at the level of 0.9 - 0.95, it is possible to distinguish three clusters of two operators and three clusters of one operator, although even in these pairs the operators differ quite noticeably.

In this way, the division of the group of operators into professionally homogeneous groups using a computer simulator, the development of image-tests, scenarios provide a sufficiently objective assessment of the results of professional selection, conducting relevant training. In addition, personnel attestation can be carried out in the same way using control materials.

The mentioned before method allows us to find the individual parameters of the recipients such as:

- the recipients in stressful situation;
- mean stress time;
- the minimum and maximum time value in the stress time interval.

Therefore, these parameters are further used to classify recipients according to the level of stress resistance. The developed method can be used for detecting the attention level.

5. Conclusions

Conducting experimental research, especially for the purpose of studying human properties, is a very complex procedure that requires significant preparation and organization, including task development and problem setting. This requires in-depth knowledge of this real operator activity, minimizing the difference between the real operator activity and its reproduction on the simulator. In addition, an important point is the completeness of the received information (data) and their detailed interpretation.

The obtained experimental data gave grounds to objectively assess the qualifications of the recipients, both in the individual plan–the parameters of the boxplot diagram, and in the professional–the parameters of the dendrogram of the hierarchical agglomerative analysis. The latest results of the division of the group of recipients showed the closeness of their characteristics and the possibility of their professional selection for operator specialties.

The given mathematical model of operator activity in information search systems formally describes the operator's work, in particular by the function of transition to another functional state, since the operator is with the same image that led to his stress. The conducted experimental study with nine recipient-operators, according to the scenario in the form of a sequence of test images provided on the monitor screen, provided the result in the form of reactions to micro-stresses. With the help of mathematical processing methods: descriptive statistics, determination of quartiles, construction of boxplot diagrams and hierarchical agglomerative cluster analysis, an objective grouping of this group of recipients was obtained. In this study, each operator is represented by an individual boxplot diagram, and the group of operators is divided into subgroups according to individual test indicators. This made it possible to carry out an objective professional selection for the formation of camera personnel.

The limitation of current work is the ability to deal with micro-stresses but not with panic attacks. This limitation is related to the used dataset and should be investigated separately. In addition, after-effect of micro-stressor should be investigated separately.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, resources, writing—original draft preparation, and supervision: RK. Methodology, formal analysis, writing—review and editing, project administration, and funding acquisition: NS. Software, validation, investigation, data curation, and visualization: BK. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Ukrainian National Research Foundation, grant number 2021.01/0103.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Allen, N., and Kessel, R. (2002). "The roles of human operator and machine in decision aid strategies for target detection," in *RTO HFM Symposium on the Role of Humans in Intelligent and Automated Systems*. Available online at: https://www.researchgate.net/publication/235023362_The_Roles_of_Human_Operator_and_Machine_in_Decision_Aid_Strategies_for_Target_Detection (accessed May 20, 2023).

Barak, O., and Tsodyks, M. (2023). Mathematical models of learning and what can be learned from them. *Curr. Opin. Neurobiol.* 80, 102721. doi: 10.1016/j.conb.2023.102721

Battiston, S., and Martinez-Jaramillo, S. (2018). Financial networks and stress testing: Challenges and new research avenues for systemic risk analysis and financial stability implications. *J. Finan. Stabil.* 35, 6–16. doi: 10.1016/j.jfs.2018.03.010

Bookstaber, R., Cetina, J., Feldberg, G., Flood, M., and Glasserman, P. (2014). Stress tests to promote financial stability: as-sessing progress and looking to the future. *J. Risk Manage. Finan. Institut.* 7, 16–25.

Brown, N. F., and Anderson, T. P. (2019). "Learning rate sensitivity model," in *2019 IEEE Aerospace Conference*. Big Sky, MT: IEEE, 1–6.

Chapelle, C. A. (2009). "Computer-assisted teaching and testing," in *The Handbook of Language Teaching*, 628–644.

Chou, C. (2000). Constructing a computer-assisted testing and evaluation system on the World Wide Web-the CATES experi-ence. *IEEE Trans. Educ.* 43, 266–272. doi: 10.1109/13.865199

DeMenno, M. B. (2022). Environmental sustainability and financial stability: can macroprudential stress testing measure and mitigate climate-related systemic financial risk? *J. Bank. Regulat.* 2, 1–29. doi: 10.1057/s41261-022-00207-2

Iannello, P., Mottini, A., Tirelli, S., Riva, S., and Antonietti, A. (2017). Ambiguity and uncertainty tolerance, need for cognition, and their association with stress. A study among Italian practicing physicians. *Med. Educ.* 22, 1270009. doi: 10.1080/10872981.2016.1270009

Kryvenchuk, Y., Helzynskyy, I., Helzhynska, T., Boyko, N., and Danel, R. (2019). Synthesis control system physiological state of a soldier on the battlefield. *CEUR-Worg S.* 2488, 297–306.

Lee, L. T. (2016). "Workplace stress: a collective challenge," in *National Institute Of Occupational Safety And Health*. Available online at: https://www.ilo.org/safework/info/publications/WCMS_466547/lang--en/index.html (accessed May 20, 2023).

McEwen, B. S., and Akil, H. (2020). Revisiting the stress concept: implications for affective disorders. *J. Neurosci.* 40, 12–21. doi: 10.1523/JNEUROSCI.0733-19.2019

McGlohen, M., and Chang, H. H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behav. Res. Meth.* 40, 808–821. doi: 10.3758/BRM.40.3.808

Paul, C. L., and Dykstra, J. (2017). Understanding operator fatigue, frustration, and cognitive workload in tactical cybersecurity operations. *J. Inform. Warfare* 16, 1–11.

Ramanathan, A., Pullum, L. L., Hussain, F., Chakrabarty, D., and Jha, S. K. (2016). "Integrating symbolic and statistical methods for testing intelligent systems: Applications to machine learning and computer vision," in *2016 Design, Automation & Test in Eu-rope Conference & Exhibition (DATE), Dresden*, 786–791.

Raskin, D. C., and Kircher, J. C. (2014). "Validity of polygraph techniques and decision methods," In *Credibility Assessment*. Cambridge, MA: Academic Press, 63–129.

Sahin, C., Rokne., J, and Alhajj, R. (2019). Human behavior modeling for simulating evacuation of buildings during emergencies. *Physica A*. 528, 121432. doi: 10.1016/j.physa.2019.121432

Zhu, H., Liu, D., Bayley, I., Harrison, R., and Cuzzolin, F. (2019). "Datamorphic testing: a method for testing intelligent applications," in *IEEE International Conference On Artificial Intelligence Testing (AITest)*. Newark, CA: IEEE, 149–156.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 26 October 2023
DOI 10.3389/fpubh.2023.1285390

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Roman Tandlich,
Rhodes University, South Africa
Sanjeeb Sudarshan Bhandari,
UPMC Western Maryland Medical Center, United States

*CORRESPONDENCE
Yavuz Yiğit
✉ yyigit@hamad.qa

RECEIVED 29 August 2023
ACCEPTED 05 October 2023
PUBLISHED 26 October 2023

CITATION
Ahun E, Demir A, Yiğit Y, Tulgar YK, Doğan M, Thomas DT and Tulgar S (2023) Perceptions and concerns of emergency medicine practitioners about artificial intelligence in emergency triage management during the pandemic: a national survey-based study. *Front. Public Health* 11:1285390.
doi: 10.3389/fpubh.2023.1285390

Perceptions and concerns of emergency medicine practitioners about artificial intelligence in emergency triage management during the pandemic: a national survey-based study

Erhan Ahun[1], Ahmet Demir[2], Yavuz Yiğit[3,4]*, Yasemin Koçer Tulgar[5,6], Meltem Doğan[6] David Terence Thomas[7,8] and Serkan Tulgar[9]

[1]Department of Emergency Medicine, Sabuncuoglu Serefeddin Training and Research Hospital, Amasya, Türkiye, [2]Department of Emergency Medicine, Faculty of Medicine, Mugla Sitki Kocman University, Mugla, Türkiye, [3]Department of Emergency Medicine, Hamad Medical Corporation, Doha, Qatar, [4]Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom, [5]Department of Medical History and Ethics, Samsun University Faculty of Medicine, Samsun, Türkiye, [6]Department of Medical History and Ethics, Kocaeli University Faculty of Medicine, Kocaeli, Türkiye, [7]Department of Medical Education, Maltepe University Faculty of Medicine, Istanbul, Türkiye, [8]Department of Pediatric Surgery, Maltepe University Faculty of Medicine, Istanbul, Türkiye, [9]Department of Anesthesiology, Samsun University Faculty of Medicine, Samsun Training and Research Hospital, Samsun, Türkiye

Objective: There have been continuous discussions over the ethics of using AI in healthcare. We sought to identify the ethical issues and viewpoints of Turkish emergency care doctors about the use of AI during epidemic triage.

Materials and methods: Ten emergency specialists were initially enlisted for this project, and their responses to open-ended questions about the ethical issues surrounding AI in the emergency room provided valuable information. A 15-question survey was created based on their input and was refined through a pilot test with 15 emergency specialty doctors. Following that, the updated survey was sent to emergency specialists via email, social media, and private email distribution.

Results: 167 emergency medicine specialists participated in the study, with an average age of 38.22 years and 6.79 years of professional experience. The majority agreed that AI could benefit patients (54.50%) and healthcare professionals (70.06%) in emergency department triage during pandemics. Regarding responsibility, 63.47% believed in shared responsibility between emergency medicine specialists and AI manufacturers/programmers for complications. Additionally, 79.04% of participants agreed that the responsibility for complications in AI applications varies depending on the nature of the complication. Concerns about privacy were expressed by 20.36% regarding deep learning-based applications, while 61.68% believed that anonymity protected privacy. Additionally, 70.66% of participants believed that AI systems would be as sensitive as humans in terms of non-discrimination.

Conclusion: The potential advantages of deploying AI programs in emergency department triage during pandemics for patients and healthcare providers were acknowledged by emergency medicine doctors in Turkey. Nevertheless, they expressed notable ethical concerns related to the responsibility and accountability aspects of utilizing AI systems in this context.

KEYWORDS

specialist, AI, triage, emergency, pandemic, ethics, survey, national

Introduction

Recent attention has been drawn to artificial intelligence (AI) due to its potential to enable the creation of computer systems that can replicate human intelligence and decision-making processes (1). AI has already permeated every aspect of our lives, even if we are not consciously aware of it (2). Recently AI techniques have sent vast waves across healthcare, even fueling an active discussion of whether AI doctors will eventually replace human physicians in the future (3).

Utilizing sophisticated algorithms, AI can 'comprehend' intricate patterns within extensive healthcare data and employ these acquired insights to enhance clinical practices. Furthermore, it can be endowed with the capability to learn and self-correct, thus refining its precision through feedback loops. An AI system aids healthcare practitioners by furnishing them with the most current medical insights from scholarly journals, textbooks, and clinical experiences, thereby ensuring optimal patient care. Additionally, AI systems are pivotal in mitigating the diagnostic and therapeutic errors intrinsic to human clinical practice (3–5). Furthermore, these AI systems extract invaluable information from extensive patient populations, facilitating real-time inferences for health risk alerts and predictions regarding health outcomes (6).

Emergency medicine, like many other medical specialties, has identified a variety of potential AI applications. Diagnosis is one of the most essential applications of AI in emergency care. In order to identify potential diagnoses, AI algorithms can examine patient data, such as symptoms, medical history, and test results, efficiently and swiftly.

AI's role in healthcare extends significantly to include advanced patient triage capabilities. By leveraging AI algorithms to analyze patient data comprehensively, healthcare systems can effectively prioritize individuals based on the severity of their condition, ensuring that those in critical need receive immediate attention and the appropriate care interventions. This not only optimizes resource allocation but also enhances patient outcomes by minimizing delays in treatment.

In emergency medicine, AI's influence in triage is particularly transformative. Beyond diagnosis, AI contributes to the triage process by rapidly assessing the acuity of each case. Through the analysis of various clinical indicators, such as vital signs, medical history, and presenting symptoms, AI systems can swiftly categorize patients, enabling healthcare providers to allocate resources efficiently.

By integrating AI-driven triage systems into emergency departments, healthcare facilities can improve the speed and accuracy of decision-making. AI's ability to analyze vast datasets and adapt to real-time information empowers clinicians to make well-informed decisions, ultimately leading to more precise and timely care delivery. As a result, patients with critical conditions receive immediate attention, while those with less urgent needs are appropriately managed, resulting in enhanced overall healthcare efficiency and patient satisfaction.

The world recently endured a severe COVID-19 pandemic, during which applications of AI were also observed in the health sector. According to reports, methods exist for the analysis of radiological and laboratory results, diagnosis, patient triage in the emergency room, and the development of patient-specific treatments during a pandemic.

As with any pervasive invention, the application of AI in health has sparked ethical concerns, and these debates are ongoing in many fields (7). A few of these concerns include data privacy and security concerns, algorithmic bias, a lack of transparency, autonomy, and accountability, the dehumanization of healthcare, and economic repercussions. For AI to have the potential to improve healthcare outcomes, it must be used ethically and responsibly.

In this survey study, using the opinions of emergency medicine specialists practicing in Turkey, we sought to determine the ethical concerns and perspective of implementing AI in emergency department triage management during an epidemic.

Materials and methods

Study design

This survey-based study investigated the ethical perspectives of emergency specialist physicians regarding the use of artificial intelligence (AI) in the emergency department. This research was approved by the Samsun University Clinical Research Ethics Committee (SÜKAEK) (Approval No. 2022-12-12, 23/11/2022) and conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. All participants gave their informed consent. Participants' confidentiality and anonymity were maintained throughout the duration of the study. Special attention was paid to ensuring that participants did not feel compelled to participate or provide specific responses.

Participant recruitment

Initially, ten emergency specialists were recruited to participate in the study. The research team devised and asked participants open-ended questions to gain insight into their experiences and perspectives regarding the ethical considerations of AI use in the ED.

Survey development

Based on the responses to the open-ended questions, the research team developed a survey for gathering more specific information from the participants. The survey consisted of 15 questions and was designed to assess the emergency specialists' ethical perspectives on the use of AI in the ED.

Survey pilot testing

Pilot testing was conducted with 15 emergency specialist doctors. These participants were asked to provide feedback on any difficulties they had understanding the questions, issues with the appropriateness of certain questions, and grammar and spelling errors. Some changes were made to the survey to improve its clarity, readability, and comprehensiveness in response to the received feedback.

Data collection

The revised survey was distributed to the emergency specialist doctors via email and social media tools between 05/12/2022–15/04/2023. All participants gave their informed consent. Responses were collected anonymously.

Survey content

In the first section of the survey, participants were asked to provide descriptive information such as age, gender, and emergency medicine experience duration. On a 3-point Likert scale, participants were subsequently asked their opinions on a total of 13 questions pertaining to four major ethical topics. Before requesting opinions on each ethical topic, a thorough explanation of the ethical rule was provided. The most important ethical issues were as follows:

Beneficence (A): In this section, participants were asked about the beneficence of AI for triage purposes during the pandemic. The question encompassed two aspects: the usefulness of AI for patients and the usefulness of AI for physicians.

Responsibility and accountability (B): In this section, four questions were posed to gauge the participants' perspectives on responsibility and accountability in the case of complications or adverse events resulting from the use of AI for triage purposes during the pandemic.

Rights to privacy and confidentiality (C): In this section, 5 questions on personal data protection and the right to privacy were posed to participants separately for artificial intelligence and deep learning.

Non-Discrimination (D): In this section, two questions were posed to ascertain the participants' perspectives on nondiscrimination.

On a 3-point Likert scale, participants assessed a total of 20 evaluations. Through a final open-ended question, participants were also given the opportunity to express any ethical concerns not addressed in the questionnaire. The example English translation of the questionnaire is provided in Table 1.

To collect data for the study, a survey was developed using Google Forms and distributed via multiple social media platforms, including WhatsApp, Twitter, and Facebook. In addition, the survey was sent individually via email to groups of emergency medicine specialists. To ensure a reliable sample size, a minimum of 154 participants were required with an 80% level of confidence and a 5% margin of error, given the total population of about 2,500 specialists. The objective was

TABLE 1 Artificial intelligence (AI) triage survey statements.

	Agree	Neutral	Disagree
A. Use of artificial intelligence for triage purposes in emergency services during COVID-19 and similar pandemics (accurate diagnosis timing, fewer complications, etc.)			
A1. It will be beneficial for the patient			
A2. It will be beneficial for the emergency medicine physician.			
B. In the event of misdiagnosis, incorrect treatment or lack of treatment, death, disability, or similar outcomes resulting from the use of artificial intelligence for triage purposes in emergency services during COVID-19 and similar pandemics.			
B1. The responsibility should only be on the practitioner			
B2. The responsibility should not only be on the practitioner, artificial intelligence and manufacturers can also be held accountable			
B3. The responsibility should only be on the artificial intelligence manufacturers or programmers.			
B4. The responsibility for complications occurring in applications varies depending on their nature			
C. The artificial intelligence program processing and retaining patients' data in its memory for the purpose of providing 'better guidance'.			
C1. It is a privacy violation in all circumstances			
C2. Retaining patient data in 'Deep Learning' based applications poses an ethical concern.			
C3. Retaining patient data in 'Artificial Intelligence' based systems raises ethical concerns			
C4. As long as patient data is kept under anonymous records, it does not pose a problem			
C5. The ethical dimension of storing patient data in artificial intelligence, deep learning, and similar systems during extraordinary periods like the COVID-19 pandemic can be overlooked.			
D. In terms of not engaging in discrimination, Artificial intelligence systems;			
D1. It will be as sensitive as humans, at the very least			
D2. Artificial intelligence systems, when used for triage purposes, will pose a problem in terms of 'non-discrimination			
E. If you believe there are additional ethical issues that may arise regarding the use of artificial intelligence for triage purposes in emergency services during COVID-19 and similar pandemics, please provide your input			

to collect responses from at least 170 participants to account for the possibility of data loss.

Statistical analysis

SPSS 16 was used for data analysis. Descriptive data were given as mean and standard deviation, and survey responses were given as frequency and percentage. *T*-test was used in the analysis of descriptive data. Categorical data were presented as counts and percentages and compared using Chi-square test or Fisher's exact test as appropriate, with post-hoc Bonferroni adjustments to determine where the difference between groups originated. Statistical significance was accepted as $p < 0.05$.

Results

Our survey was completed by 171 individuals within the specified time frame. Although it was clear at the outset of the survey through our social media posts that the target of the survey was emergency medicine specialists, it was discovered that 4 participants were emergency medicine residents, and as a result, only 167 participants were taken into consideration for evaluation.

An overview of the survey results is shown in the Figure 1.

Participants were composed of 44 females to 123 males with an average age of 38.22 ± 6.79 years, and an average duration of professional experience of 6.79 ± 5.25 years (Table 2).

The responses to questions A1 and A2 were analyzed when evaluating the utility of using artificial intelligence for triage purposes in emergency services during COVID-19 and similar pandemics (such as accurate diagnosis time, fewer complications, etc.). 54.50% of participants believed it would be beneficial for patients, and 70.06% believed it would be beneficial for healthcare professionals.

In section B, participants were asked to assess the use of artificial intelligence in pandemics from the standpoint of responsibility and accountability. Only 12.57% of participants believed that all responsibility lies with emergency medicine specialists, while 23.95% said that only artificial intelligence manufacturers and programmers should bear responsibility. The highest rate of agreement was found for the statement "The responsibility for complications occurring in applications varies depending on their nature" (79.04%). The majority, 63.47%, stated that both parties should be responsible. A demonstration of the answers given by the participants according to their agreement regarding responsibility is presented in Figure 2.

In section C, situations pertaining to the right to privacy and confidentiality were evaluated. 20.36% of participants stated that deep learning-based applications violate privacy, while 23.95% stated that AI-based applications violate privacy. 14.37% of respondents believed that these applications constitute a violation of privacy under all circumstances. Conversely, 61.68% of participants believed that there is no violation of privacy so long as the data is recorded anonymously. In addition, nearly half of the respondents (47.31%) agreed that the ethical aspect of storing patient data in artificial intelligence and similar systems could be disregarded during extraordinary events such as the COVID-19 pandemic.

In section D, opinions on the non-discrimination principle were evaluated. 70.66% of participants believe that AI-based systems would be as sensitive to non-discrimination as humans. However, only 13.37% of respondents agreed that artificial intelligence-based triage would violate the non-discrimination principle.

In section E, respondents were questioned about their perspectives on additional ethical issues not covered by the survey. Seven participants were concerned about how cultural differences might affect patient attitudes toward AI-based applications and their potential repercussions. In addition, one participant expressed concern about the elevated risk of incorrect positive/negative

TABLE 2 The analysis of the participants' descriptive characteristics.

Variable	Result
Age (Years)	38.32 ± 6.20
Gender (*n*, %)	
Female	44 (26.4%)
Male	123 (77.6%)
Duration of experience as an emergency medicine specialist:	
<5 y experience	66 (39.5%)
5–10 y experience	68 (40.7%)
>10y experience	33 (19.8%)

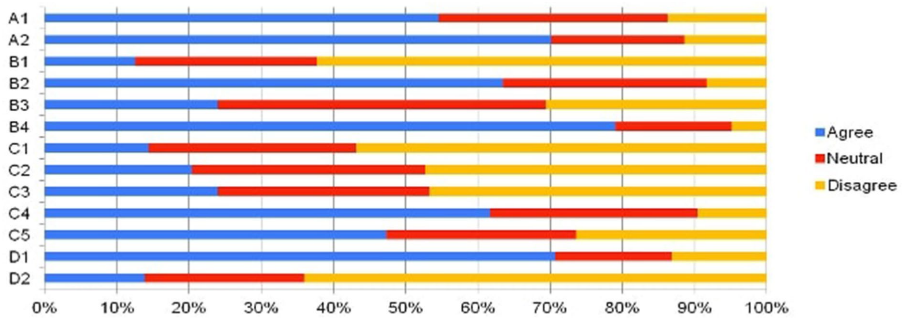


FIGURE 1
Demonstration of questionnaire results.

diagnoses as a result of the difficulties vulnerable groups face in comprehending and expressing themselves.

With the exception of two questions ($p < 0.005$), the majority of the items' responses did not demonstrate any discernible differences based on age, gender, or years of professional experience (Table 3). There was a significant gender difference in responses to the statement that the practitioner should bear sole responsibility for negative outcomes resulting from the use of artificial intelligence in triage (question B1). The proportion of women who disagreed with this statement was higher than that of men (86.3% vs. 53.6%, $p < 0.001$) (Table 4).

Another significant statistical finding was associated with question C2 of the survey, which addressed the ethical aspect of patient data storage in deep learning-based applications, specifically patient privacy. There was a higher rate of disagreement with this statement among emergency medicine specialists with over 10 years of experience compared to other experience duration groups ($p = 0.040$) (Table 4).

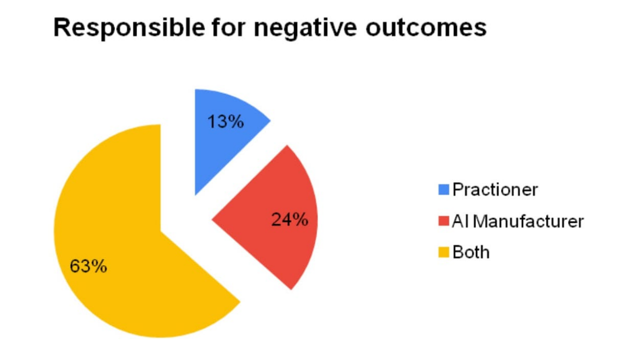


FIGURE 2
Elicitation of participants' perceptions regarding accountability for adverse outcomes arising from the utilization of artificial intelligence (AI). Percentages have been rounded to the nearest number.

Discussion

According to our findings, the majority of emergency medicine specialists in Turkey believe that using AI-based systems for triage in emergency rooms during COVID-19 and other pandemics will benefit both patients and emergency physicians. In addition, participants believe that both the artificial intelligence and the emergency medicine professional should be held accountable for any problems caused by this application, with approximately 80% agreeing that, the responsibility for complications in AI applications varies depending on their nature. A smaller proportion of participants agreed that AI's collection of personal information violates users' privacy, and nearly half said that this issue might be overlooked in extreme circumstances such as a pandemic. In terms of "non-discrimination," most participants believed that artificial intelligence would be just as sensitive to this as humans, if not more so.

Triage is a critical process used in emergency medicine to effectively prioritize and manage the severity and urgency of patients' healthcare needs. Triage is critical in quickly assessing and categorizing individuals based on the severity of their conditions and their chances of survival, especially during medical emergencies or disasters when a large influx of patients requires immediate medical attention. Medical resources can be allocated appropriately by efficiently triaging patients, ensuring that those in critical condition receive prompt care while optimizing the overall allocation of healthcare resources (1).

Triage becomes even more critical in the context of a pandemic, such as COVID-19, because the number of patients seeking medical care may exceed the available resources (2). During a pandemic, the triage process is critical for distinguishing between patients infected with the virus who require immediate medical attention and those who have mild symptoms that can be managed at home. This method ensures that resources, such as hospital beds, medical equipment, and healthcare personnel, are used as efficiently as possible (3).

Furthermore, during a pandemic, healthcare facilities may need to change their triage protocols in order to reduce the risk of infection transmission (4). Patients suspected or confirmed to have COVID-19, for example, may be triaged separately from other patients, and

TABLE 3 Assessing the relationship between participants' demographic characteristics and their answers to questions and statistical results (*p* values).

	A1	A2	B1	B2	B3	B4	C1	C2	C3	C4	C5	D1	D2
Age (Years)	0.192	0.130	0.173	0.718	0.765	0.397	0.555	0.344	0.635	0.574	0.633	0.150	0.818
Gender	0.924	0.159	**<0.001**	0.417	0.595	0.491	0.221	0.743	0.938	0.787	0.534	0.264	0.626
Duration of experience	0.606	0.822	0.683	0.469	0.510	0.876	0.134	**0.040**	0.278	0.167	0.951	0.451	0.775

TABLE 4 Assessment of descriptive items regarding questions/statements (only statistically significant items are presented in this table).

Question no	Descriptives	Agree *n* (%)	Neutral *n* (%)	Disagree *n* (%)	*p*
B1	Female	1 (2.3%)	5 (11.3%)	38 (86.3%)*	<0.001
	Male	20 (16.2%)	37 (30%)	66 (53.6%)	
C2	<5 y experience	17 (25.7%)	19 (28.8%)	30 (45.5%)	0.040
	5–10 y experience	14 (20.6%)	27 (39.7%)	27 (39.7%)	
	>10y experience	3 (9%)	7 (21.2%)	23 (69.6%)*	

No, number; Q, question; Values are presented as number (%). *Indicates the group leading to statistically significant difference.

healthcare personnel may be required to wear personal protective equipment (PPE) to reduce their risk of exposure (6). These precautions are intended to protect both patients and healthcare workers while also slowing the spread of the virus.

Various technologies, such as telephone systems, digital scoring systems, deep learning, and AI-based systems, have been used for triage during pandemics (7–12). During the COVID-19 pandemic, AI programs have been recommended and implemented to improve patient, healthcare worker, and community safety. These AI systems consider descriptive data like age, gender, BMI, medical history, medications, contact history, and risk factors of patients who present to the emergency department. The AI systems generate an output by analyzing the patients' current complaints, physical findings, laboratory tests, and radiological images. The analysis process employs technologies such as algorithms, machine learning, and artificial neural networks to determine probable diagnoses, urgency levels, and the severity of the patients' conditions (13, 14). Subsequently, based on the outputs, patients can be directed to appropriate medical care units, hospitals, or facilities based on their level of urgency. Furthermore, these AI systems can aid in decision-making processes such as categorizing patients for home treatment, emergency department follow-up, or admission to a regular ward or intensive care unit (15–18).

The terms "deep learning" and "artificial intelligence" are frequently used interchangeably, but they are not the same (19). AI systems strive to imitate human learning models and demonstrate human-like intelligence. Deep learning, on the other hand, is concerned with discovering patterns and relationships in large datasets and making inferences. As a result, deep learning is only one technique in the larger field of AI. Deep learning, natural language processing, robotics, and other domains are all part of AI. While AI is used in a variety of fields, deep learning is used specifically for discovering and utilizing patterns and relationships in large datasets (20). Despite briefly mentioning the distinction between AI and deep learning in our survey, we generally preferred to use the term "AI" rather than separate the two terms. Although the use of AI in medicine appears to be promising and beneficial, it is not without ethical concerns. These include, among other things, biases, a lack of transparency, privacy, accountability and responsibility, equity, depersonalization, and autonomy (21, 22). Although this survey could have been designed in a much more comprehensive manner, we focused on the topics of beneficence responsibility and accountability, rights to privacy and confidentiality, and non-discrimination in our study. However, as the scope of survey studies grows and the time required to complete them grows, so does the participation rate. Furthermore, this is the first study to assess emergency medicine specialists' perspectives on AI applications, and it should be viewed as a pilot ethical study focusing on a specific issue rather than a comprehensive ethical study.

Some expert opinions and survey studies have called into question the beneficence and ethical aspects of AI use in various medical fields (23, 24). However, there is currently no article that discusses the ethical implications of AI in the field of Emergency Medicine. Nonetheless, it is worth noting that studies on the ethical implications of AI use in many other areas of medicine have been published. Cobianchi et al. examined the ethical dimension of AI usage in surgical sciences in their study (using a modified Delphi process) and concluded that "the main ethical issues that arise from applying AI to surgery, relate to human agency, accountability for errors, technical robustness, privacy and data governance, transparency, diversity, non-discrimination, and fairness" (22). There are numerous recent studies discussing the ethical dimensions of AI usage in many areas of medicine, including imaging, differential diagnosis, prediction models, and decision-making, and they generally raise similar ethical concerns (25–30).

Unlike previous studies, our research is not a consensus paper reporting expert opinions or a Delphi consensus paper to address experts' ethical concerns. Instead, we asked emergency medicine specialists who are currently or may be using AI for triage purposes during COVID-19 and similar pandemics about the ethical implications of its use. We chose to address the topics of beneficence, responsibility and accountability, privacy and confidentiality rights, and non-discrimination in our study, which focused on a single medical condition and a single purpose. We believe that, as a pilot study in the early stages of the AI era, our research will shed light on future applications. Aside from these, numerous ethical issues concerning various AI usage domains can be discussed (26).

Privacy rights and non-discrimination are prominent ethical debates in literature regarding AI. In medical ethics, the right to privacy includes not only bodily privacy but also that regarding health and personal life. As a result, individuals who are adequately informed have the right to decide how much of their information is shared (31). Within predetermined frameworks, a violation of an individual's privacy rights may be deemed acceptable only when the benefits to the society or third parties outweigh the breach (32). While anonymizing individuals' data before incorporating it into the system can alleviate some privacy concerns, the lack of transparency in how artificial intelligence processes data creates uncertainty about the extent to which individuals can exercise control over their own data (33). In our study, approximately 23% of participants saw retaining and subjecting data to repeated analysis within AI systems as an ethical issue, and a similar percentage saw deep learning-based systems as an ethical issue. Furthermore, 60% of participants believed that as long as the data was collected anonymously, it would not violate their privacy.

In our study, more than 70% of participants believed that AI would be as sensitive to non-discrimination as humans, while only 13% saw AI usage as an ethical concern regarding discrimination. While there is widespread agreement in the literature that AI would be more fair than humans, there are also reservations about the extent to which AI can be fair. When data generated through discriminatory thinking is fed into the system, it has the potential to perpetuate discrimination. Furthermore, the opaque decision-making mechanism of AI, which is based on established algorithms, makes identifying instances of discrimination caused by AI difficult (34, 35).

Our study has some limitations. Firstly, because it is a content-specific study conducted exclusively within the emergency medicine profession, the generalizability of our findings to a broader spectrum of healthcare practitioners may be limited. Specifically, the age distribution of the participating physicians is relatively young, which could potentially introduce bias into our results. While including older emergency physicians could offer a different perspective, it is worth noting that although emergency medicine is not a new specialty in Turkey, the recent surge in the number of graduates in the field may have contributed to the predominance of younger specialists in our study. This demographic trend, to some extent, reflects the current composition of emergency medicine professionals in the country and is a constraint beyond our control. Additionally, our survey concentrated on four specific ethical concerns chosen by the investigators, offering an in-depth exploration of these issues. However, conducting a more extensive survey, such as using the Delphi technique, might have provided a broader ethical perspective. Yet, the practical challenges of recruiting a larger participant pool could have arisen.

Conclusion

According to our findings, emergency medicine specialists in Turkey thought that using AI programs for triage in emergency departments during pandemics could be beneficial, safe, and complication-reducing for patients and healthcare providers. Participants, however, expressed serious ethical concerns about the responsibility and accountability associated with using these systems for the stated purpose. Surprisingly, the majority of participants believed that ethical concerns about data storage and reuse could be overlooked. The perspectives of both the engineers and developers who create AI systems and the potential users, who are healthcare professionals, should be gathered more thoroughly. To develop guidelines, these perspectives should be combined with those of bioethics leaders.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Samsun University Clinical Research Ethics Committee (SÜKAEK) (Approval No. 2022-12-12, 23/11/2022). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

EA: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. AD: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. YY: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. YT: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. MD: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. DT: Conceptualization, Formal analysis, Project administration, Software, Visualization, Writing – original draft. ST: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Open Access funding provided by the Qatar National Library.

Conflict of interest

YY was employed by Department of Emergency Medicine, Hamad Medical Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Hinson JS, Martinez DA, Cabral S, George K, Whalen M, Hansoti B, et al. Triage performance in emergency medicine: a systematic review. *Ann Emerg Med*. (2019) 74:140–52. doi: 10.1016/j.annemergmed.2018.09.022

2. Flaatten H, Van Heerden V, Jung C, Beil M, Leaver S, Rhodes A, et al. The good, the bad and the ugly: pandemic priority decisions and triage. *J Med Ethics*. (2020) 47:e75. doi: 10.1136/medethics-2020-106489

3. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. (2017) 2:230–43. doi: 10.1136/svn-2017-000101

4. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. *Artif Intell Med*. (2009) 46:5–17. doi: 10.1016/j.artmed.2008.07.017

5. Alkahlout BH, Ahmad S. Optimizing patient care in the emergency department: insights from automated alert systems and triage strategies. *Glob Emerg Crit Care*. (2023) 2:71–6. doi: 10.4274/globecc.galenos.2023.25744

6. Neill DB. Using artificial intelligence to improve hospital inpatient care. *IEEE Intell Syst*. (2013) 28:92–5. doi: 10.1109/MIS.2013.51

7. Eppes CS, Garcia PM, Grobman WA. Telephone triage of influenza-like illness during pandemic 2009 H1N1 in an obstetric population. *Am J Obstet Gynecol*. (2012) 207:3–8. doi: 10.1016/j.ajog.2012.02.023

8. Payne R, Darton TC, Greig JM. Systematic telephone triage of possible "swine" influenza leads to potentially serious misdiagnosis of infectious diseases. *J Infect*. (2009) 59:371–2. doi: 10.1016/j.jinf.2009.09.005

9. Lai L, Wittbold KA, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: novel strategies for population health management in response to the COVID-19 pandemic. *Healthc (Amst)*. (2020) 8:100493. doi: 10.1016/j.hjdsi.2020.100493

10. Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun*. (2020) 11:3543. doi: 10.1038/s41467-020-17280-8

11. Chou EH, Wang CH, Hsieh YL, Namazi B, Wolfshohl J, Bhakta T, et al. Clinical features of emergency department patients from early COVID-19 pandemic that predict SARS-CoV-2 infection: machine-learning approach. *West J Emerg Med*. (2021) 22:244–51. doi: 10.5811/westjem.2020.12.49370

12. Soltan AAS, Kouchaki S, Zhu T, Kiyasseh D, Taylor T, Hussain ZB, et al. Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digital Health*. (2021) 3:e78–87. doi: 10.1016/S2589-7500(20)30274-0

13. Nazir T, Mushhood Ur Rehman M, Asghar MR, Kalia JS. Artificial intelligence assisted acute patient journey. *Front Artif Intell*. (2022) 5:962165. doi: 10.3389/frai.2022.962165

14. Arnaud E, Elbattah M, Ammirati C, Dequen G, Ghazali DA. Use of artificial intelligence to manage patient flow in emergency department during the COVID-19 pandemic: a prospective, single-center study. *Int J Environ Res Public Health*. (2022) 19:9667. doi: 10.3390/ijerph19159667

15. Becker J, Decker JA, Römmele C, Kahn M, Messmann H, Wehler M, et al. Artificial intelligence-based detection of pneumonia in chest radiographs. *Diagnostics (Basel)*. (2022) 12:1465. doi: 10.3390/diagnostics12061465

16. Song W, Zhang L, Liu L, Sainlaire M, Karvar M, Kang MJ, et al. Predicting hospitalization of COVID-19 positive patients using clinician-guided machine learning methods. *J Am Med Inform Assoc*. (2022) 29:1661–7. doi: 10.1093/jamia/ocac083

17. Soltan AAS, Yang J, Pattanshetty R, Novak A, Yang Y, Rohanian O, et al. Real-world evaluation of rapid and laboratory-free COVID-19 triage for emergency care: external validation and pilot deployment of artificial intelligence driven screening. *Lancet Digit Health*. (2022) 4:e266–78. doi: 10.1016/S2589-7500(21)00272-7

18. Kapoor A, Kapoor A, Mahajan G. Use of artificial intelligence to triage patients with flu-like symptoms using imaging in non-COVID-19 hospitals during COVID-19 pandemic: an ongoing 8-month experience. *Indian J Radiol Imaging*. (2021) 31:901–9. doi: 10.1055/s-0041-1741103

19. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. (2017) 69:S36–40. doi: 10.1016/j.metabol.2017.01.011

20. Alsuliman T, Humaidan D, Sliman L. Machine learning and artificial intelligence in the service of medicine: necessity or potentiality? *Curr Res Transl Med*. (2020) 68:245–51. doi: 10.1016/j.retram.2020.01.002

21. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front Surg*. (2022) 9:862322. doi: 10.3389/fsurg.2022.862322

22. Cobianchi L, Verde JM, Loftus TJ, Piccolo D, Dal Mas F, Mascagni P, et al. Artificial intelligence and surgery: ethical dilemmas and open issues. *J Am Coll Surg*. (2022) 235:268–75. doi: 10.1097/XCS.0000000000000242

23. Koçer Tulgar Y, Tulgar S, Güven Köse S, Köse HC, Çevik Nasırlıer G, Doğan M, et al. Anesthesiologists' perspective on the use of artificial intelligence in ultrasound-guided regional Anaesthesia in terms of medical ethics and medical education: a survey study. *Eurasian J Med*. (2023). doi: 10.5152/eurasianjmed.2023.22254

24. Bostrom N, Yudkowsky E. (2023). The ethics of artificial intelligence. Available at: http://faculty.smcm.edu/acjamieson/s13/artificialintelligence.pdf

25. Shen FX, Wolf SM, Gonzalez RG, Garwood M. Ethical issues posed by field research using highly portable and cloud-enabled neuroimaging. *Neuron*. (2020) 105:771–5. doi: 10.1016/j.neuron.2020.01.041

26. Couture V, Roy MC, Dez E, Laperle S, Bélisle-Pipon JC. Ethical implications of artificial intelligence in population health and the Public's role in its governance: perspectives from a citizen and expert panel. *J Med Internet Res*. (2023) 25:e44357. doi: 10.2196/44357

27. Lorenzini G, Arbelaez Ossa L, Shaw DM, Elger BS. Artificial intelligence and the doctor-patient relationship expanding the paradigm of shared decision making. *Bioethics*. (2023) 37:424–9. doi: 10.1111/bioe.13158

28. Rolfes V, Bittner U, Gerhards H, Krüssel JS, Fehm T, Ranisch R, et al. Artificial intelligence in reproductive medicine – an ethical perspective. *Geburtshilfe Frauenheilkd*. (2023) 83:106–15. doi: 10.1055/a-1866-2792

29. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. *J Med Internet Res*. (2022) 24:e40238. doi: 10.2196/40238

30. Anshari M, Hamdan M, Ahmad N, Ali E, Haidi H. COVID-19, artificial intelligence, ethical challenges and policy implications. *AI Soc*. (2022) 38:707–20. doi: 10.1007/s00146-022-01471-6

31. Humayun A, Fatima N, Naqqash S, Hussain S, Rasheed A, Imtiaz H, et al. Patients' perception and actual practice of informed consent, privacy and confidentiality in general medical outpatient departments of two tertiary care hospitals of Lahore. *BMC Med Ethics*. (2008) 9:14. doi: 10.1186/1472-6939-9-14

32. Childress JF, Beauchamp TL. Common morality principles in biomedical ethics: responses to critics. *Camb Q Healthc Ethics*. (2022) 31:164–76. doi: 10.1017/S0963180121000566

33. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. (2020) 260:113172. doi: 10.1016/j.socscimed.2020.113172

34. Zuiderveen Borgesius FJ. Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int J Human Rights*. (2020) 24:1572–93. doi: 10.1080/13642987.2020.1743976

35. Cossette-Lefebvre H, Maclure J. AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making. *AI Ethics*. (2022). doi: 10.1007/s43681-022-00233-w

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 31 October 2023
DOI 10.3389/fpubh.2023.1268223

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
David Morgan,
Portland State University, United States
Chengyi Xia,
Tiangong University, China
Junxiang Chen,
Indiana University, United States

*CORRESPONDENCE
Lauren Towler
✉ l.towler@soton.ac.uk

†These authors share first authorship

‡These authors share senior authorship

RECEIVED 27 July 2023
ACCEPTED 16 October 2023
PUBLISHED 31 October 2023

CITATION
Towler L, Bondaronek P, Papakonstantinou T, Amlôt R, Chadborn T, Ainsworth B and Yardley L (2023) Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response: comparing human and machine-assisted topic analysis techniques.
Front. Public Health 11:1268223.
doi: 10.3389/fpubh.2023.1268223

Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response: comparing human and machine-assisted topic analysis techniques

Lauren Towler[1,2*†], Paulina Bondaronek[3,4†], Trisevgeni Papakonstantinou[3,5], Richard Amlôt[6], Tim Chadborn[3], Ben Ainsworth[7,8‡] and Lucy Yardley[1,2‡]

[1]School of Psychology, University of Southampton, Southampton, United Kingdom, [2]School of Psychological Science, University of Bristol, Bristol, United Kingdom, [3]Department of Health and Social Care, Office for Health Improvement and Disparities, London, United Kingdom, [4]Institute for Health Informatics, University College London, London, United Kingdom, [5]Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, London, United Kingdom, [6]Behavioural Science and Insights Unit, UK Health Security Agency, London, United Kingdom, [7]Department of Psychology, University of Bath, Bath, United Kingdom, [8]National Institute for Health Research Biomedical Research Centre, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

Introduction: Machine-assisted topic analysis (MATA) uses artificial intelligence methods to help qualitative researchers analyze large datasets. This is useful for researchers to rapidly update healthcare interventions during changing healthcare contexts, such as a pandemic. We examined the potential to support healthcare interventions by comparing MATA with "human-only" thematic analysis techniques on the same dataset (1,472 user responses from a COVID-19 behavioral intervention).

Methods: In MATA, an unsupervised topic-modeling approach identified latent topics in the text, from which researchers identified broad themes. In human-only codebook analysis, researchers developed an initial codebook based on previous research that was applied to the dataset by the team, who met regularly to discuss and refine the codes. Formal triangulation using a "convergence coding matrix" compared findings between methods, categorizing them as "agreement", "complementary", "dissonant", or "silent".

Results: Human analysis took much longer than MATA (147.5 vs. 40 h). Both methods identified key themes about what users found helpful and unhelpful. Formal triangulation showed both sets of findings were highly similar. The formal triangulation showed high similarity between the findings. All MATA codes were classified as in agreement or complementary to the human themes. When findings differed slightly, this was due to human researcher interpretations or nuance from human-only analysis.

Discussion: Results produced by MATA were similar to human-only thematic analysis, with substantial time savings. For simple analyses that do not require an in-depth or subtle understanding of the data, MATA is a useful tool that can support qualitative researchers to interpret and analyze large datasets quickly. This approach can support intervention development and implementation, such as enabling rapid optimization during public health emergencies.

KEYWORDS

public health, interventions, qualitative analysis, machine learning techniques, triangulation

1. Introduction

Qualitative research plays a vital role in public health, intervention development and implementation research by enabling researchers to develop an informed understanding of the attitudes, perceptions and contextual factors relevant to planning and delivering effective and acceptable health interventions (1, 2). However, most qualitative approaches (such as interviews, focus groups and observation studies) are resource intensive and time-consuming, requiring months or years to collect and analyze rich, in-depth data. Consequently, most qualitative approaches have traditionally been based on studies of relatively small, purposively selected samples (3). Whilst this kind of in-depth approach has enormous benefits in terms of generating nuanced insights for the purpose of theory-building, it is less suitable for some potential applications of qualitative methods. In particular, less resource intensive methods are needed in order to analyze the wealth of qualitative data that can be generated by automated online data collection (for example, of free text responses to population surveys). Whilst computational and automated approaches are commonly used in the field of epidemic modeling to monitor disease spread and adoption of preventative behaviors by members of the public [e.g., (4)], these methods do not provide sufficient insight into perceptions of preventative behaviors and individual decision-making processes that qualitative approaches offer.

Recent advances in technology have facilitated the automatic processing of text-based qualitative datasets, via natural language processing (NLP), a subfield of artificial intelligence. NLP algorithms can quickly produce "triaged" natural text outputs, that have the potential to substantially reduce the amount of text to be examined by research teams whilst remaining meaningful (5). NLP has been applied in several areas of healthcare research: extracting information from electronic healthcare records (6, 7), coding interview transcripts about male health needs (8), or early detection of depression in social networks (9). A direct comparison of an NLP approach which used lexicon-based clustering in WordNet with human-only qualitative analysis analyzed answers from 84 participants to short open-ended text message survey questions (10). They found that NLP generated similar findings although was not of as high quality, and could be used to in combination with human qualitative analysis to provide more detail.

Indeed, the importance of the input of experienced qualitative researchers to NLP-assisted qualitative data analysis must not be overlooked. Findings by Guetterman et al. (10) highlight how experienced qualitative researchers bring knowledge of contextual, theoretical, and sociocultural factors that cannot be replicated by NLP-only approaches. Whilst previous studies show how NLP methods can be used to support deductive approaches where an *a priori* coding framework is in place (11), there is often a need to conduct "bottom-up" inductive and exploratory analyses where ideas are formed from the data itself, particularly when developing new public health interventions or adapting existing interventions to new situations or populations. Inductive qualitative analysis allows researchers to explore relevant issues and topics as guided by members of the relevant population, and generate new ideas in a data-driven way (12, 13). In this project, we therefore aimed to explore the use of a different specific NLP approach which integrates human and exploratory NLP analysis– which we have termed "Machine-Assisted Topic Analysis" (MATA) – to allow expert qualitative researchers to look at large, real-world datasets in a timely manner.

MATA assists qualitative researchers by summarizing major patterns in the text according to generative models of word counts – known as topic models (14). Topic models are able to automatically infer latent topics from text. This means the model assumes that the documents consist of a combination of underlying topics and can be represented as such. Topic models allow for machine-assisted reading of text datasets through creating and extracting the main themes that underlie a corpus and mapping them onto the individual documents. They are particularly useful as tools to analyze large volumes of free-text responses to questions in a data-driven way, in order to summarize the main families of responses. The approach used in this study is based on an application of the Structural Topic Model (14, 15) in particular. The STM is a general framework for topic modeling that is differentiated from other topic modeling methodologies by its ability to enable researchers to include additional variables at the document level, such as the date a document was created or the demographics of the person who created it, as covariates in a topic model. This way the relationships of these variables to specific topics can be estimated and examined or used to run subgroup analyses. Those variables are further used to explain variance in topic prevalence, so affect the frequency with which a topic is discussed. As a result, their inclusion improves inference and qualitative interpretability and also affects the topical content (14). Structural topic models are able to identify patterns, and qualitative researchers can then use the output to extract meaning, interpret and summarize the topics.

Within the context of COVID-19, several NLP researchers have identified NLP as a potentially effective tool for rapid analysis of large-scale text-based datasets in order to meet the rapidly shifting public health needs during a pandemic (11, 16, 17). For example, NLP

Abbreviations: AI, Artificial intelligence; IPA, Interpretative phenomenological analysis; MATA, Machine-assisted topic analysis; NLP, Natural language processing; PBA, Person based approach; STM, Structural topic model; TA, Thematic analysis.

approaches could allow the rapid analysis of views and experiences of public health interventions (such as infection tracking tools, or public health messaging services) via survey response, allowing teams to improve interventions in real-time as issues arise – which can be vital given the rapidly changing context of a worldwide pandemic (3, 18). However, previous comparisons between exploratory NLP methods and human-only qualitative analyses have mostly been conducted on relatively small sample sizes (8, 10). Therefore, there is a need to assess how NLP methods can inductively analyze large datasets for studies with exploratory aims. One such study using a large dataset demonstrated that supervised machine learning approaches could effectively complement human hand-coding (19). The current study builds on Nelson's work by attempting to demonstrate how NLP methods can be applied to 'real world' participant data in a rapid-response situation, providing further evidence for the validity and utility of the method.

Germ Defence is a digital behavior change intervention that aims to improve infection control behaviors during the COVID-19 pandemic (20). In order to remain as effective as possible, Germ Defence was iteratively updated throughout the pandemic, as health guidelines and contextual factors (e.g., virus prevalence, vaccine uptake) changed (18). During the intervention, some website users provided feedback about the content and design, and we used this data to perform separate qualitative analyses using MATA and human-only analysis. We aimed to explore similarities and differences between findings of the two methods, and to compare the person-hours required to conduct each form of analysis, in order to assess the potential value and trustworthiness of MATA for large-scale public health intervention evaluation and optimization.

2. Methods

2.1. Participants

Inclusion criteria were users of the Germ Defence website who were over the age of 18 and able to give informed consent. Between 18th November 2020 until 3rd January 2021, a total of 2,175 people consented to the survey, 1,472 of which responded to at least one open-ended question. During this time, a second national lockdown was in place in the UK, which was replaced by the reintroduction of the tiered system on 2nd December 2020. Data collection ended prior to the third national lockdown on 6th January 2021. Table 1 shows the demographic characteristics of the sample.

2.2. Measures

To gather demographic data, closed questions were asked pertaining to age, sex, ethnicity, education, household size, whether the user or someone else in the household is at increased risk of severe illness if they caught COVID, and whether there could be a current COVID case within the household (experiencing symptoms or contact with confirmed case). Feedback was collected as free-text responses to two questions: "What was helpful about the information on the Germ Defence website?" and "What did you not find helpful about the information on the Germ Defence website?" Responses to these questions provide a rich dataset of recommendations that can

TABLE 1 Demographic characteristics of the sample (N = 1,472).

Demographics	N	%
Who do you live with		
Alone	304	20.7
With children under 16	176	12.0
With family all over 16	889	60.4
With people not related to me	73	5.0
Blank	30	2.0
Increased risk of severe illness (self or household member)		
Yes	861	58.5
No	535	36.3
Blank	76	5.2
Possibility of current COVID-19 infection (self or household member)		
Yes	69	4.7
No	1,335	90.7
Blank	68	4.6
Age		
18–25	10	0.7
26–40	76	5.2
41–60	524	35.6
61–70	471	32.0
70+	324	22.0
Blank	67	4.5
Sex		
Female	972	66.0
Male	423	28.7
Other or prefer to self-describe	4	0.3
Prefer not to say	3	0.2
Blank	70	4.8
Ethnicity		
White	1,331	90.4
Black African	2	0.1
Black Caribbean	5	0.3
Black (other)	2	0.1
Indian	9	0.6
Pakistani	4	0.3
Bangladeshi	1	0.1
Chinese/Southeast Asian	6	0.4
Asian (other)	6	0.4
Other	28	1.9
Prefer not to say	8	0.5
Blank	70	4.8
Education		
Before finishing school	33	2.2
After finishing school	643	43.7
After finishing university	353	24.0
After postgraduate studies	280	19.0
Blank	163	11.1

Participants who selected "Other" categories for ethnicity were able to give an additional open-text response. Most who selected this category were from mixed backgrounds, but some specified themselves as, for example, White Armenian, Turkish/Cypriot, or Nepalese etc.

TABLE 2 Human-only analysis procedure and person-hours.

	Procedure	Hours (total person-hours)
Preparation	Each of the 7 coders were assigned ~210 participants, whose responses were transferred to the NVivo software package. LT set up the initial coding template based on a codebook developed and validated during previous analyses of Germ Defence data (24), previous survey data gathered from website users, and some initial data familiarization. Six VRAs were trained by LT in qualitative coding and using NVivo. This involved giving the VRAs an overview of the qualitative process and its aims, the coding process and the meaning of inductive and deductive coding, and previous qualitative analyses from the Germ Defence project.	25
Coding	Analyzed using a codebook TA approach, template analysis (23). The data were coded deductively onto the thematic codebook, though some inductive codes were integrated into the codebook upon discussion with the team.	95 (13.6 h per coder)
Validity checks	The first 50 survey respondents allocated to each trainee coder (23.81% of average total respondents per coder) were cross-checked, and any discrepancies were discussed in subgroups until agreement was reached, under supervision of LT.	14
Interpretation	LT interpreted the findings and created themes from the coding and discussed with the team. LT presented the results to the wider team, and made any adjustments based on discussion with the coders and wider team.	13.5
Total person hours		147.5

be used to improve the website and guidance provided. The goal of this data collection and analysis was to investigate perceptions of the Germ Defence website, with particular attention to the acceptability of the intervention and the advice it provides.

2.3. Procedure

After they had completed at least one of the two main sections of the intervention (handwashing or reducing illness), visitors to the Germ Defence website received a pop-up asking if they might be interested in taking a survey to help improve the website. The invitation was presented as seeking information on users' views on protecting themselves from Coronavirus, and their thoughts on the Germ Defence website. Users could then follow a link to the study information sheet, consent form, and the online questionnaire hosted on Qualtrics. Ethical approval was granted by the University of Southampton Psychology Ethics Committee (ID: 56445).

2.4. Data analysis

We analyzed the data in two ways; human-only qualitative analysis and MATA.

2.4.1. Human-only qualitative analysis

The human-only analysis was conducted using a codebook thematic analysis (TA) approach using template analysis techniques (21–23) whereby the coding template was applied to the data deductively by several coders, and the unit of analysis was free-text participant response. The coding team was made up of an experienced qualitative researcher and lecturer at the University of Southampton (LT), and a group of 6 voluntary research assistants (VRAs) made up of both undergraduate and postgraduate Psychology students at the Universities of Southampton and Bath.

The initial codebook had been developed through the researchers' (LT) contextual knowledge, involvement in collating feedback for the person-based approach (PBA) development of the Germ Defence intervention, and derived from smaller-scale survey data and formal TA of qualitative interviews with website users (24). Inductive coding was also implemented by the coders where relevant data did not currently fit with existing codes. Any proposed additional inductive codes identified during coding were discussed with the group as soon as possible, so that each coder could keep it in mind for their own coding. However, as the process went on, many of these inductive codes were deemed too thin to remain as standalone codes (for example, "environmental concerns over waste, e.g., disposable masks" or "it's too cold to ventilate"), and so they were merged together or with existing codes. As a result, some of the deductive codes evolved into broader, higher-level codes throughout the process than their original form. This process was done by discussion and interpretation of the code meaning by the team. Any disagreements between the team members throughout this process was resolved by discussion until agreement was reached.

LT then interpreted the shared meaning of the codes within the final framework and created the themes, which were discussed with the team. See Table 2 for further information on how the codebook was developed, and the procedures used in the human analysis. In the MATA, template analysis techniques were also used to analyze the topics generated by the STM, with each topic being the unit of analysis.

2.4.2. Machine-assisted topic analysis

Structured data, such as date, age, sex, education level and ethnicity, were also collected and included in the models as covariates.

2.4.2.1. Preparation

We pre-processed the data using R (version 3.5.2), and cleaned the free text responses using base R functions, the quanteda [version 2.0.1; (25)] and STM [version 1.3.3; (14)] packages. We deleted observations with missing values and duplicate data. The free-text responses were converted into token units using the quanteda package, after punctuation, symbols and numbers were removed. In this instance the tokens were individual words. Data pre-processing was completed by deleting stop words and stemming the tokens. Stemming is the process of reducing words to their root. This acts as a normalization of text data and helps reduce the size of the dictionary which speeds up processing.

TABLE 3 Machine-assisted topic analysis approach and person-hours.

	Procedure	Hours
Preparation	Data cleaning and conversion of data to STM format	8
Coding	The structural topic model is run. The model infers the topics from the corpus of text and maps them back to individual response extracts, which are now automatically grouped into their assigned topics and represented as a distribution of them.	0
Validity checks	Diagnostic analysis and evaluation is conducted of models with 5–40 topics	4
Interpretation	Two coders interpret and apply narrative labels (codes) to the topics (stage 1). This procedure is conducted independently and blindly by the two coders to ensure accuracy and validity. The coders then create broader themes in collaboration through the process of thematic analysis to generate the final template (stage 2)	28 (9 h per coder)
Total person hours		40

2.4.2.2. Coding and validity checks

As a topic modeling method, we implemented the Structural Topic Model (15). Prior to running the models we ran diagnostics to identify the optimal number of topics, according to both the relevant metrics and the aims of the analysis, focusing on the trade-off between semantic coherence and exclusivity [see (15) for a discussion on this method of evaluation]. We tested models with 5–40 topics and differing covariates in terms of semantic coherence score (see (26)), residuals and interpretability by human coders (see Supplementary material 1), separately for each question. Upon visually examining the plots in Supplementary material 2, we identified a Structural Topic Model with 25 topics to be optimal for addressing question A, "What was helpful about the information on the Germ Defence website?" whereas 15 topics were deemed to be optimal for addressing question B, "What did you not find helpful about the information on the Germ Defence website?" In both cases date, age, gender, ethnicity, and level of education were included as covariates. The model automated the equivalent of the coding stage of the analysis by assigning a number of labels to each document, by way of mapping them to topics. The code used for data preparation and modeling is publicly available in the figshare repository: https://figshare.com/articles/dataset/Germ_Defence_-_Machine_Assisted_Topic_Analysis/19514305.

2.4.2.3 Interpretation: qualitative analysis of machine-generated data by trained, supervised coders

The outputs examined consisted of two main elements; the 10 most representative quotes for each topic and two lists of weighted words that constitute the topic. Different types of word weightings were generated with each topic where the following two types were analyzed in subsequent qualitative analysis: (1) Highest Prob (words within each topic with the highest probability) and (2) FREX (words that are both frequent and exclusive, identifying words that distinguish topics). Examples of outputs generated are presented in Supplementary material 3.

In order to analyze the model's output systematically we analyzed it in two stages. In Stage 1, two researchers interpreted the output and agreed upon narrative labels for the topics (henceforth, MATA codes). In the first part of Stage 1 the analysis was blinded. In the second part the two researchers resolved any disagreements in the interpretation through discussion and agreed in the final topic labels. In Stage 2, the researchers analyzed the topics generated by the text analysis and created broader themes. Table 3 provides a breakdown of the steps of the MATA process, along with the person-hours that were spent on each step.

2.4.3. Triangulation

We conducted a formal triangulation in order to compare the results from both approaches. Specifically, we performed a methodological and investigator triangulation, as the results from two different analytical approaches performed by two different analysts were compared (27). Two research teams independently analyzed the Germ Defence data using the two methods described in the previous sections (MATA and human-only TA). A "convergence coding matrix" (28, 29) was created, and two researchers from these separate teams (LT and PB) independently triangulated the findings from both analyses. The codes were then compared with each other and categorized as either; agreement, complementarity, dissonance, or silence (28, 29). Agreement represented conceptual convergence between the analyses, and complementarity referred to a shared meaning or essence between the findings, but some unique nuances were present. Dissonance represented disagreement between the coding, and silence referred to a finding which was present in only one of the analyses. As such, codes were not considered dissonant with each other when they only represented difference of opinion within the sample, and not between the coding from the two methodologies. For example, the code 'clear and simple' from the human analysis was not considered dissonant with 'wordy and repetitive' from the MATA because alternative codes were present which agreed, such as 'information was clear, concise, and easy to understand.' The two analysts then compared and discussed their decisions and reached consensus on the findings.

3. Results

3.1. Person hours

The human qualitative analysis required significantly higher person hours to complete than the MATA (147.5 vs. 40). The only stage which less time in the human analysis than the MATA was the final interpretation stage, likely due to the familiarity with the data gained by coding the data "by hand" and the pre-existing coding template. In the MATA approach, the inference of the topics and the classification component of the analysis was conducted by the machine learning model. In this case, the final interpretation phase consisted of the two stages of generating narrative descriptions of the produced topics and following the process of thematic analysis. This was the first time the human coders came into contact with the data and thus this step was the most time-consuming one in the MATA.

3.2. Primary data analysis

The MATA results were centered on what users found helpful and unhelpful about the Germ Defence website. The themes representing what users found helpful were: 1. *Clear and easy to understand*, 2. P*rovision of new information and reminders*, 3. *Confirming and Reinforcing.* The themes representing what users found unhelpful were: 1. *Repetitive, simplistic, wordy, patronizing*, 2. *Lack of tailoring*, 3. Var*ious issues relating to usability, content and specific features*. For the human analysis, we found 3 main themes: (1) *layout and language style*, (2) *confidence in how to perform the behaviors*, and (3) *reducing all or nothing thinking*. These themes, and how they relate to each other, are presented in section 3.3 Triangulation. As the current study is concerned with the results of the triangulation between the two methods, further information on the results of the separate primary analyses can be found in Supplementary materials 4, 5.

3.2.1. Machine-assisted topic analysis process: inclusion of topics

3.2.1.1. What was helpful about the information on the germ defence website?

Of 25 topics analyzed qualitatively, 22 topics were included in the analysis as they provided substantial insights as expressed by the users' feedback[1]. See Supplementary material 5 for a ranking of the machine-generated topics in terms of prevalence in the corpus for question A.

3.2.1.2. What did you not find helpful about the information on the germ defence website?

Of 15 topics analyzed qualitatively, 13 topics were included in the analysis as they provided substantial insights as expressed by the users' feedback[2]. See Supplementary material 5 for a ranking of the machine-generated topics in terms of prevalence in the corpus for question B.

The MATA codes from both corpora were grouped into major themes representing what users found helpful/unhelpful with the Germ Defence intervention (Figures 1, 2).

3.3. Triangulation

The codes generated from each form of analysis were categorized as either in agreement, or complementary to each other. We found no instances of dissonance or silence within the coding from the two methods. Table 4 presents the full results of the triangulation.

1 The rationale for exclusion of 3 topics from the analysis was:

Topic 4 was deemed incoherent

Topic 11 was described as "Nothing was helpful/Learned nothing new" and hence did not provide a substantial answer to the qualitative question

Topic 23 included mixed issues that were already represented in other themes.

2 The rationale for exclusion of 2 topics from the analysis was:

Topic 13 was deemed incoherent

Topic 15 was described as "Nothing was unhelpful/nothing to dislike" and hence did not provide a substantial answer to the qualitative question.

3.3.1. Instances of agreement

There was a high level of agreement between the findings of the human and MATA analyses, particularly for the themes: *layout and language style* and *confidence in how to perform the behaviors*. All of the codes which made up the *layout and language style* theme from the human analysis were classified as in agreement with the related codes identified in the MATA. Both methods agreed that Germ Defence users found the website clear to use and easy to understand, but there were a few areas requiring improvement. For example, some users felt that the website did not appear "slick" or sophisticated enough, and that the simple language appeared patronizing to some. Some examples of codes classified as in agreement were: "clear and simple" versus "information was clear, concise and easy to understand", and too "simplistic/patronizing" versus "did not provide any new information beyond what is already known and is patronizing".

We also found many instances of agreement between the methods for two of the three codes which made up the theme *confidence in how to perform the behaviors* from the human-only analysis. Both methods agreed that many of the participants felt that the website provided important reminders and reinforcement of the recommended behaviors. For example, for those who were already highly adherent to the behaviors, the website provided assurance that they were doing the right thing and encouragement to continue. For those who experienced difficulty performing the behaviors, the website provided practical guidance and "real-world" examples of how the infection control behaviors could be integrated into users' daily routines. An example of codes classified as in agreement is "clear practical advice and troubleshooting is helpful" from the human-only analysis versus "helpful information users had not thought of before; the case studies were helpful" from the MATA.

Finally, two of the four codes contained within the *reducing all or nothing thinking* theme agreed with codes generated from the MATA. The majority of the agreement here came from finding that some of the behaviors may be more difficult to integrate, particularly for families with young children. Some participants felt that Germ Defence could appear too proscriptive, and placed emphasis on the need to balance the behaviors according to what was deemed practical and necessary for the family to perform to reduce risk. For example, the 'some behaviors are very challenging in certain situations' code from the human-only analysis was classified as in agreement with 'guidance and questions lack consideration for practicalities within families, especially families with young children' from the MATA.

3.3.2. Instances of complementarity

The remaining relationships between the findings of the two methods were judged as complementary and there were no instances of dissonance or silence. Only the theme *reducing all or nothing thinking* contained more codes deemed as complementary than in agreement. Both methods found that users placed emphasis on the need to act according to risk level, and that some of the suggested behaviors could be unrealistic in certain households and/or situations. However, the human analysis placed greater emphasis on the potential mental load of integrating the behaviors, and participants' interpretations of the viral load messages. The viral load messages encouraged some participants by helping them to understand that even small changes (such as implementing some of the behaviors wherever possible and practical, or that they might tailor their behaviors according to risk) can be effective for reducing their risk of

FIGURE 1
What was helpful about the information on the Germ Defence website? Summary of the topics (generated by the model, described by human) and the major themes (generated by human).

catching COVID and/or illness severity. In contrast, believing that they must perform all behaviors perfectly to avoid virus transmission left some participants feeling defeated. The MATA codes did not wholly reflect these interpretations, and so "understanding that small changes matter is motivating" from the human-only analysis was classed as complementary to codes such as: "information on how the virus lives and spreads, along with explanation of the link between amount of viral exposure and severity of illness" from the MATA.

4. Discussion

We aimed to explore the potential value of machine learning analysis techniques to analyze large-scale datasets by conducting a comparison between MATA and traditional thematic codebook analysis using a template approach conducted by humans. We triangulated the results of both forms of analysis in order to highlight the similarities and differences between the two methods, and we compared them by the person-hours needed to complete the analyses.

In regard to the primary data, both analyses found that online public health interventions should be clear and concise. For our participants, a slick and professional appearance conveys trustworthiness, and many felt that a website should be uncomplicated and accessible. However, others felt that it seemed overly simplistic and patronizing, indicating a need for striking balance when designing interventions targeted to a wide audience. Rather than simply stating the recommended behaviors, our participants highlighted the importance of practical information and real-life examples which aim to help website users envision *how* the behaviors can be implemented in their own homes. Having the efficacy of the behaviors confirmed by those perceived to be experts empowered participants to act, and reinforced participants' confidence in their ability to protect themselves and those around them. Finally, our participants indicated that public health interventions should recognize that some of the recommended behaviors can be very challenging in certain situations, and attempting to adhere to all behaviors at all times may not be feasible for many households. Many participants indicated that they would act according to their risk level, and felt that information which appeared overly restrictive and inflexible can leave participants feeling defeated and demotivated. On the other hand, messages which emphasized the concept of viral load helped many participants to understand that making even small changes were worthwhile for reducing viral exposure, and understanding risk reduction as cumulative – rather than absolute – was motivating.

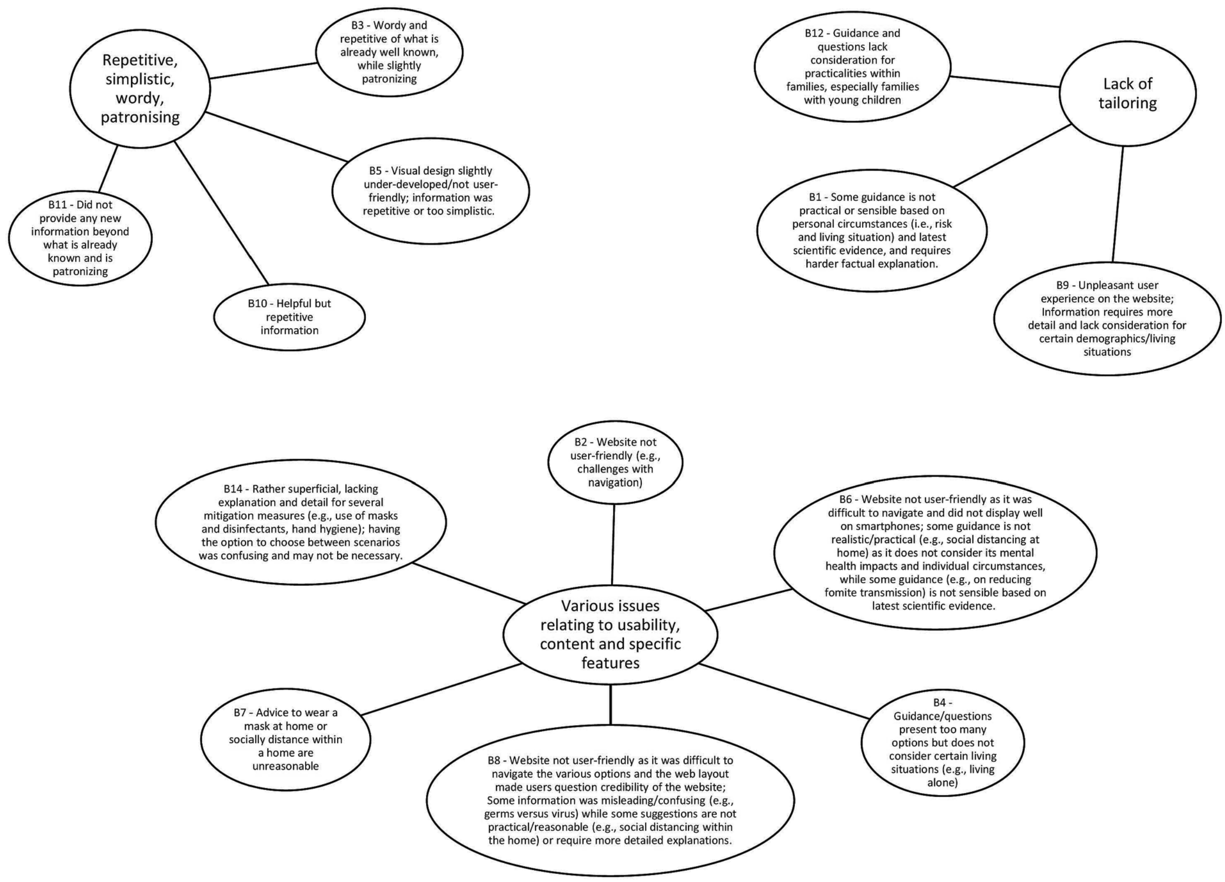


FIGURE 2
What did you not find helpful about the information on the Germ Defence website? Summary of the topics (generated by the model, described by human) and the major themes (generated by human).

As a result of the triangulation between the two methodologies, we found that the results were very similar, with all codes developed from the MATA classified as in agreement or complementary to the codes developed from the human-only analysis. Where the findings were classified as complementary, this was typically due to slightly differing interpretations or nuance which are likely to be due to the human input to the analyses. For example, the investigator leading the human-only analysis (LT) had analyzed previous Germ Defence data, whereas the MATA team had not. It is therefore likely that LT made interpretations based on knowledge gained from previous analyses of Germ Defence data. This particularly seems to be the case for the codes within the *reducing all or nothing thinking* theme, which were more prominent and developed in the human-only analysis by the Germ Defence team. These concepts were salient to the Germ Defence developers because Germ Defence sought to overcome fatalism about infection transmission. Therefore, some of these differences were likely due to investigator difference, and not methodological difference. That said, the codes from the human-analysis were generally more interpretive than the MATA codes. This is different from the findings from another study which compared human analysis with a different NLP approach. Guetterman et al. (10) found that whilst human-only analysis was of higher quality than NLP-only analysis, a combined approach added further conceptual detail and further conclusions than human-only analysis. We did not find this to be the case in the current study, rather, we found that human-only methods yielded similar results to a human-assisted NLP approach.

One potential consideration is that punctuation is removed for the MATA as only words, rather than phrases or sentences, are used as tokens. Due to the purpose of punctuation being to convey and clarify meaning, emphasis, and tone within text, the human coders may have been able to understand nuances within the responses during the early stages of analysis that could have been missed or misattributed by the AI. However, the role that humans play in understanding and interpreting the output of the MATA means that any potential missed meaning should be minimal. Similarly, the topics produced by STM can sometimes be incoherent, or involve multiple seemingly unrelated themes. This would be a major issue if the goal of this method was to conduct an exhaustive and in-depth qualitative analysis of the corpus. However, since the goal of this analysis, and the use case for MATA in general, was to rapidly extract headline insights, this limitation can be mostly overlooked. Nevertheless, researchers should be mindful of these potential issues when they come to interpret the output of the AI.

Due to these considerations, MATA could potentially be seen as a less interpretive method than human-only analysis that is suitable for more descriptive studies of large datasets. Indeed, the concept of information power recommends larger samples for

TABLE 4 Results of the triangulation between the human-only analysis and the MATA.

Human-only themes	Human-only codes	Triangulation with MATA codes	
		Agreement	Complementary
Layout and language style	Clear and simple	A1, A3, A5, A22	
	Not enough information	B9, B11, B14	
	Not streamlined or sophisticated	B5, B2, B6, B8	
	Too repetitive	B3, B5, B10	
	Too simplistic/patronizing	B3, B5, B11, B14	
Confidence in how to perform the behaviors	Clear practical advice and troubleshooting is helpful	A2, A6, A9, A10, A12, A13, A20, A24, A7, A21, A14, A18	B12, B9
	Feeling informed and reinforced by reliable sources is empowering	A12, A13, A16, A20, A7, A14, A25, A19, A17	A15
	Inconsistencies undermine confidence		A20, A17
Reducing all or nothing thinking	Trying to perform all the behaviors is exhausting		B12, B6
	Understanding that small changes matter is motivating		A8, A21, A19
	We should act according to risk	B1	A16, A19
	Some behaviors are very challenging in certain situations	B12, B1, B4, B6, B8	B9

studies with broader, atheoretical, more exploratory aims (30). In order to complete the human-only analysis of a sample of this size, a codebook was created based on previous Germ Defence research, and six research assistants needed to be trained in qualitative analysis. It would not have been feasible to conduct a purely inductive thematic analysis using a large number of coders due to differences in how individuals would interpret and label the data. Other methods of coding large-scale data, such as crowdsourcing though Amazon Mechanical Turk, have been shown to be successful when coding deductively into pre-determined categories (31–33). However, in the absence of these categories, such as in more inductive approaches or studies with more exploratory aims, there have previously been few options available to researchers other than to perform human analyses on limited sample sizes. Approaches such as MATA could be a valuable tool for enabling large-scale sampling for these types of studies.

In the rapidly evolving landscape of artificial intelligence, the emergence and application of Large Language Models (LLMs) like Chat-GPT in qualitative data analysis needs to be considered as a viable alternative approach. Mellon et al. (34) demonstrated how LLMs can accurately replicate human coding of large-scale data when classifying the most important perceived issues in the United Kingdom, such as health and education. However, whilst LLMs offer scalability and efficiency, it is possible that they could inadvertently introduce biases or miss nuanced interpretations if applied uncritically (35), and researchers must ensure they manually validate the output to verify accuracy and quality. Furthermore, Mellon et al. (34) highlight that it is currently unknown whether LLMs can code the sentiment of open-text data, or whether they are capable of coding the data as well as producing a coding scheme in the way that STM does. It is possible that current and future iterations of Chat-GPT could have these capabilities, but further research is required. The current study demonstrates how a MATA approach can integrate qualitative researcher oversight to ensure sentiment is captured.

Therefore, MATA offers researchers a less resource intensive and time-consuming approach to conducting broader exploratory studies within large, nationally representative samples, whilst still ensuring human oversight in the process to accurately capture meaning and sentiment. It could be used to augment approaches which tend to adopt more descriptive aims such as codebook TA, coding reliability TA, and content analysis. For analyses such as reflexive TA or interpretative phenomenological analysis (IPA) where researchers wish to engage with the data on a richly interpretive level, and the researchers' knowledge of the subject matter is considered an important analytic lens, we would not currently consider MATA an appropriate approach based on the current findings.

4.1. Strengths and limitations

The decision to triangulate human qualitative analysis of Germ Defence data with machine learning analysis was made *post hoc*, and as such, both teams worked and made analytical decisions independently from each other. Whilst this could be seen as a limitation of the current study, we believe that the high level of agreement and complementarity between the two analyses demonstrate the trustworthiness of using machine learning techniques to analyze large-scale datasets. Despite the independence of the two teams, the MATA was still able to generate findings very similar to the human analysis. As discussed above, machine learning techniques may be best suited to more descriptive qualitative analyses, and so it is likely that the results were consistent due to the descriptive aims of the human analysis and the similarity between the results would likely not have been as great if compared with a more interpretive analysis.

The sample of participants in the current study was largely homogenous. The majority of participants were white, midlife or older, and at higher risk of severe illness from COVID-19. We are therefore unable to draw conclusions from the current study as to the utility of MATA and NLP methodology for the analysis of more diverse, nationally representative samples. Further research is needed to assess how NLP techniques handle more diverse datasets.

4.2. Conclusion

For studies with more descriptive aims, MATA is a trustworthy and potentially valuable tool to assist researchers analyse large-scale open text data. Previously, qualitative approaches have been limited to small sample sizes by its time-consuming nature. By triangulating the results from a traditional human-only thematic analysis with those from MATA, we have shown that both methods generate comparable findings, whilst MATA has the benefit of being less resource and time intensive. MATA could therefore be used to automate the early familiarization and coding process of more descriptive and less interpretive methods such as codebook analysis or content analysis, especially when the goal is to rapidly extract key topics or concepts from the data for use in a public health emergency. This study contributes to an emerging body of literature into the potential utility of machine learning techniques for use in large-scale qualitative research (5, 8–11).

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.6084/m9.figshare.19514305.

Ethics statement

The studies involving humans were approved by the University of Southampton Psychology Ethics Committee (ID: 56445). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LT: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. PB: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. TP: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing. RA: Conceptualization, Writing – review & editing. TC: Conceptualization, Writing – review & editing. BA: Conceptualization, Writing – original draft, Writing – review & editing. LY: Conceptualization, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was funded by United Kingdom Research and Innovation Medical Research Council (UKRI MRC) Rapid Response Call: UKRI CV220-009. The Germ Defence intervention was hosted by the Lifeguide Team, supported by the NIHR Biomedical Research Centre, University of Southampton. LY is a National Institute for Health Research (NIHR) Senior Investigator and team lead for University of Southampton Biomedical Research Centre. LY is affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Behavioural Science and Evaluation of Interventions at the University of Bristol in partnership with Public Health England (PHE).

Acknowledgments

We would like to thank our voluntary research assistants; Benjamin Gruneberg, Lillian Brady, Georgia Farrance, Lucy Sellors, Kinga Olexa, and Zeena Abdelrazig for their valuable contribution to the coding of the data for the human-only analysis. We would also like to acknowledge Katherine Morton's contribution to the administration of survey, and James Denison-Day for the construction and maintenance of the Germ Defence website.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed are those of the author(s) and not necessarily those of the NIHR, the Department of Health or PHE.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1268223/full#supplementary-material

References

1. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. *Psychiatry Res*. (2019) 280:112516. doi: 10.1016/j.psychres.2019.112516

2. Shuval K, Harker K, Roudsari B, Groce N, Mills B, Siddiqi Z, et al. Is qualitative research second class science? A quantitative longitudinal examination of qualitative research in medical journals. *PLoS One*. (2011) 6:e16937. doi: 10.1371/journal.pone.0016937

3. Vindrola-Padros C, Chisnall G, Cooper S, Dowrick A, Djellouli N, Symmons S, et al. Carrying out rapid qualitative research during a pandemic: emerging lessons from COVID-19. *Qual Health Res*. (2020) 30:2192–204. doi: 10.1177/1049732320951526

4. Xia C, Wang Z, Zheng C, Guo Q, Shi Y, Dehmer M, et al. A new coupled disease-awareness spreading model with mass media on multiplex networks. *Inf Sci*. (2019) 471:185–00. doi: 10.1016/j.ins.2018.08.050

5. Crowston K, Allen E, Heckman R. Using natural language processing technology for qualitative data analysis. *Int J Soc Res Methodol*. (2012) 15:523–43. doi: 10.1080/13645579.2011.625764

6. Ford E, Carroll J, Smith H, Scott D, Cassell J. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. (2016) 23:1007–15. doi: 10.1093/jamia/ocv180

7. Zheng L, Wang Y, Hao S, Shin A, Jin B, Ngo A, et al. Web-based real-time case finding for the population health management of patients with diabetes mellitus: a prospective validation of the natural language processing–based algorithm with statewide electronic medical records. *JMIR Med Inform*. (2016) 4:e37. doi: 10.2196/medinform.6328

8. Leeson W, Resnick A, Alexander D, Rovers J. Natural language processing (NLP) in qualitative public health research: a proof of concept study. *Int J Qual*. (2019) 18:1988702. doi: 10.1177/1609406919887021

9. Cacheda F, Fernandez D, Novoa F, Carneiro V. Early detection of depression: social network analysis and random forest techniques. *J Med Internet Res*. (2019) 21:e12554. doi: 10.2196/12554

10. Guetterman T, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran V. Augmenting qualitative text analysis with natural language processing: methodological study. *J Med Internet Res*. (2018) 20:e231. doi: 10.2196/jmir.9702

11. Lennon RP, Fraleigh R, Van Scoy LJ, Keshaviah A, Hu XC, Snyder BL, et al. Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. Fam med. *Community Health*. (2021) 9:e001287. doi: 10.1136/fmch-2021-001287

12. Braun V, Clarke V. *Successful qualitative research: A practical guide for beginners*. London: Sage Publishing (2013).

13. Greenhalgh T, Taylor R. How to read a paper: papers that go beyond numbers (qualitative research). *BMJ*. (1997) 315:740–3. doi: 10.1136/bmj.315.7110.740

14. Roberts M, Stewart B, Tingley D. STM: an R package for structural topic models. *J Stat Softw*. (2019) 91:1–40. doi: 10.18637/jss.v091.i02

15. Roberts M, Stewart B, Tingley D, Airoldi E. The structural topic model and applied social science. *Neural Inform Process Soc*. (2013)

16. Baclic O, Tunis M, Young K, Doan C, Swerdfeger H, Schonfeld J. Challenges and opportunities for public health made possible by advances in natural language processing. *Can Commun Dis Rep*. (2020) 46:161–8. doi: 10.14745/ccdr.v46i06a02

17. Chang T, DeJonckheere M, Vydiswaran V, Li J, Buis L, Guetterman T. Accelerating mixed methods research with natural language processing of big text data. *J Mix Methods Res*. (2021) 15:398–412. doi: 10.1177/15586898211021196

18. Morton K, Ainsworth B, Miller S, Rice C, Bostock J, Denison-Day J, et al. Adapting behavioral interventions for a changing public health context: a worked example of implementing a digital intervention during a global pandemic using rapid optimisation methods. *Front Public Health*. (2021a) 9:197. doi: 10.3389/fpubh.2021.668197

19. Nelson LK, Burk D, Knudsen M, McCall L. The future of coding: a comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociol Methods Res*. (2021) 50:202–37. doi: 10.1177/0049124118769114

20. Ainsworth B, Miller S, Denison-Day J, Stuart B, Groot J, Rice C, et al. Infection control behavior at home during the COVID-19 pandemic: observational study of a web-based behavioral intervention (germ Defence). *J Med Internet Res*. (2021) 23:e22197. doi: 10.2196/22197

21. Braun V, Clarke V. Reflecting on reflexive thematic analysis. *Qual Res Sport Exerc Health*. (2019) 11:589–97. doi: 10.1080/2159676X.2019.1628806

22. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qual Res Psychol*. (2021) 18:328–52. doi: 10.1080/14780887.2020.1769238

23. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. *Qual Res Psychol*. (2015) 12:202–22. doi: 10.1080/14780887.2014.955224

24. Morton K, Towler L, Groot J, Miller S, Ainsworth B, Denison-Day J, et al. Infection control in the home: a qualitative study exploring perceptions and experiences of adhering to protective behaviours in the home during the COVID-19 pandemic. *BMJ Open*. (2021b) 11:e056161. doi: 10.1136/bmjopen-2021-056161

25. Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, et al. Quanteda: an R package for the quantitative analysis of textual data. *J Open Source Softw*. (2018) 3:774. doi: 10.21105/joss.00774

26. Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. *EMNLP*. (2011) 2011:262–72.

27. Farmer T, Robinson K, Elliott S, Eyles J. Developing and implementing a triangulation protocol for qualitative health research. *Qual Health Res*. (2006) 16:377–94. doi: 10.1177/1049732305285708

28. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. *BMJ*. (2010) 341:c4587. doi: 10.1136/bmj.c4587

29. Tonkin-Crine S, Anthierens S, Hood K, Yardley L, Cals J, Francis N, et al. Discrepancies between qualitative and quantitative evaluation of randomised controlled trial results: achieving clarity through mixed methods triangulation. *Implement Sci*. (2015) 11:66. doi: 10.1186/s13012-016-0436-0

30. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qual. Health Res*. (2016) 26:1753–60. doi: 10.1177/1049732315617444

31. Harris J, Mart A, Moreland-Russell S, Caburnay C. Diabetes topics associated with engagement on twitter. *Prev Chronic Dis*. (2015) 12:402. doi: 10.5888/pcd12.140402

32. Hilton L, Azzam T. Crowdsourcing qualitative thematic analysis. *Am J Eval*. (2019) 40:575–89. doi: 10.1177/1098214019836674

33. Tosti-Kharas J, Conley C. Coding psychological constructs in text using mechanical Turk: a reliable, accurate, and efficient alternative. *Front Psychol*. (2016) 7:741. doi: 10.3389/fpsyg.2016.00741

34. Mellon J, Bailey J, Scott R, Breckwoldt J, Miori M. Does GPT-3 know what the most important issue is? Using large language models to code open-text social survey responses at scale. *Soc Sci Res Netw*. (2023). doi: 10.2139/ssrn.4310154

35. Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? *BMJ*. (2021) 372:n304 doi: 10.1136/bmj.n304

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 01 November 2023
DOI 10.3389/fpubh.2023.1248121

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Tatsawan Timakum,
Chiang Mai Rajabhat University, Thailand
Ravindra Kumar Vemula,
English and Foreign Languages University, India

*CORRESPONDENCE
Rafael Pinto
rafael.pinto@lais.huol.ufrn.br

RECEIVED 26 June 2023
ACCEPTED 16 October 2023
PUBLISHED 01 November 2023

CITATION
Pinto R, Lacerda J, Silva L, Araújo AC, Fontes R, Lima TS, Miranda AE, Sanjuán L, Gonçalo Oliveira H, Atun R and Valentim R (2023) Text mining analysis to understand the impact of online news on public health response: case of syphilis epidemic in Brazil.
Front. Public Health 11:1248121.
doi: 10.3389/fpubh.2023.1248121

Text mining analysis to understand the impact of online news on public health response: case of syphilis epidemic in Brazil

Rafael Pinto[1,2,3]*, Juciano Lacerda[2,4], Lyrene Silva[1], Ana Claudia Araújo[2,4], Raphael Fontes[2], Thaisa Santos Lima[2,5], Angélica E. Miranda[6,7], Lucía Sanjuán[8], Hugo Gonçalo Oliveira[9], Rifat Atun[10,11] and Ricardo Valentim[2,12]

[1]Department of Informatics and Applied Mathematics, Federal University of Rio Grande do Norte, Natal, Brazil, [2]Laboratory of Technological Innovation in Health (LAIS), Federal University of Rio Grande do Norte, Natal, Brazil, [3]Information Systems Coordination, Federal Institute of Rio Grande do Norte, Natal, Brazil, [4]Department of Social Communication, Federal University of Rio Grande do Norte, Natal, Brazil, [5]Federal Senate, Brasília, Brazil, [6]Ministry of Health, Brasília, Brazil, [7]Postgraduate Program in Infectious Diseases, Federal University of Espírito Santo, Vitória, Brazil, [8]Department of Social and Cultural Anthropology, Autonomous University of Barcelona, Barcelona, Spain, [9]Centre for Informatics and Systems of the University of Coimbra (CISUC), Department of Informatics Engineering (DEI), University of Coimbra, Coimbra, Portugal, [10]Health Systems Innovation Lab, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States, [11]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States, [12]Department of Biomedical Engineering, Federal University of Rio Grande do Norte, Natal, Brazil

Background: To effectively combat the rising incidence of syphilis, the Brazilian Ministry of Health (MoH) created a National Rapid Response to Syphilis with actions aimed at bolstering epidemiological surveillance of acquired, congenital syphilis, and syphilis during pregnancy complemented with communication activities to raise population awareness and to increase uptake of testing that targeted mass media outlets from November 2018 to March 2019 throughout Brazil, and mainly areas with high rates of syphilis. This study analyzes the volume and quality of online news content on syphilis in Brazil between 2015 and 2019 and examines its effect on testing.

Methods: The collection and processing of online news were automated by means of a proprietary digital health ecosystem established for the study. We applied text data mining techniques to online news to extract patterns from categories of text. The presence and combination of such categories in collected texts determined the quality of news that were analyzed to classify them as high-, medium-and low-quality news. We examined the correlation between the quality of news and the volume of syphilis testing using Spearman's Rank Correlation Coefficient.

Results: 1,049 web pages were collected using a Google Search API, of which 630 were categorized as earned media. We observed a steady increase in the number of news on syphilis in 2015 ($n = 18$), 2016 ($n = 26$), and 2017 ($n = 42$), with a substantial rise in the number of news in 2018 ($n = 107$) and 2019 ($n = 437$), although the relative proportion of high-quality news remained consistently high (77.6 and 70.5% respectively) and in line with similar years. We found a correlation between news quality and syphilis testing performed in primary health care with an increase of 82.32, 78.13, and 73.20%, respectively, in the three types of treponemal tests used to confirm an infection.

Conclusion: Effective communication strategies that lead to dissemination of high quality of information are important to increase uptake of public health policy actions.

KEYWORDS

communication, mass media, data mining, text extraction, public health, notifiable disease, syphilis

1. Introduction

Syphilis is a major public health problem in Brazil, an upper-middle income country with a unified health system and universal health coverage (1). Figure 1 shows the evolution of syphilis rates from 2010 to 2019. During this period, the incidence rate of congenital syphilis reached, in 2018, 9.0 cases per 1,000 live births, decreasing to 8.2 cases per 1,000 live births in 2019. The detection rate of syphilis in pregnant women reached 21.5 cases per 1,000 live births in 2018, and in 2019 it decreased to 20.8 per 1,000 live births. Acquired syphilis, listed as a notifiable disease in 2010, reached 76.2 cases per 100,000 population in 2018, but reduced to 72.8 cases per 100,000 population in 2019 (2).

In 2016, the Brazilian Ministry of Health (MoH) declared syphilis, a sexually transmitted infection, a public health emergency in Brazil and created a national initiative, "Applied Research for Intelligent Integration to Strengthen Healthcare Networks for a Rapid Response to Syphilis" — also known as the "Syphilis No!" Project (SNP) – to combat syphilis. The Laboratory for Technological Innovation in Health (LAIS) partnered with the MoH and Federal University of Rio Grande do Norte (UFRN) to implement the Syphilis No! Project.

The Syphilis No! Project comprised academic and medical research, as well as educational and health communication activities consisting of corporate and digital communication, advertising, and public relations. Panel 1 describes the core components of the Syphilis No! Project.

The Syphilis No! Project also developed a digital health ecosystem, namely Hermes, with analytics underpinned by computational tools and machine learning that integrated data from multiple sources to monitor project implementation and campaign progress by measuring education and communication activities and through epidemiological surveillance on the number of tests and number of syphilis cases (Panel 2 in Methods section).

In 2018, the 2018–2019 National Campaign to Combat Syphilis (the Syphilis No! Campaign) was launched with the theme "test, treat, and cure" and the motto "remember to take care." Aired between November and December 2018, the campaign was the first to involve national mass media outlets as part of an integrated response with local actions in priority areas, that had high rates of syphilis (3–5). In January to March 2019, the SNP implemented syphilis-related actions through digital social networks. A key theme of the communication campaign developed by the "Syphilis No" project was "Test, Treat and Cure."

During the "Syphilis No!" Campaign, the media played a vital role in enhancing understanding of the disease through messages in several channels. The organizers produced and disseminated a large amount of material throughout Brazil. The campaign broadcasted videos about syphilis via television; radio and streaming platforms delivered pre-recorded audio messages (spots and testimonials); sponsored videos on Youtube channels and pre-recorded audio messages were delivered as ads on Spotify; ads were printed in newspapers and consumer magazines; posters were displayed in shop windows, bus stops, billboards, and other forms of media were implanted in publicly

Panel 1 – Syphilis No! Project

The Syphilis No! Project is a national rapid response strategy to syphilis of the National Health System of Brazil (SUS). The Project is primary health care-oriented and aims to promote collaborative actions between surveillance and care in the territory under an interfederative perspective of technical cooperation in health. Syphilis No! operationalization:

1. Line 1 – Actions of universal scope:

- Purchase and distribution of crystalline and benzathine penicillin;
- Purchase and distribution of rapid syphilis tests;
- Strengthening of laboratory structure for diagnosis;
- Instrumentation of situation rooms in all Brazilian districts and in the Federal District;
- Implementation of national prevention campaigns;
- Development of education and communication tools to be made available to all municipalities;
- Dissemination of strategic information to municipal and district managers to support decision-making;
- Development of studies and research aimed at coping with syphilis.

2. Line 2 – Actions developed mainly by Research and Intervention Supporters on priority areas selected by epidemiological criteria (capitals and strategic municipalities, a total of 100, that in 2015 represented 68.95% of congenital syphilis burden of disease in Brazil) (3).

- Interfederative technical cooperation for implementation of committees (municipal/regional) to investigate mother-to-child transmission of syphilis;
- Cooperation to evaluate actions to confront Syphilis in the municipal and district plans, health programs and management reports of the priority districts and municipalities;
- Interfederative technical cooperation for implementation of syphilis epidemiological surveillance rooms at the municipal level;
- Interfederative technical cooperation to strengthen the health care network and different spaces of care provision for implementation of syphilis care lines (syphilis in pregnant women and sexual partners, children exposed to maternal syphilis, and acquired syphilis), also with intervention in key populations (sex workers, gay persons and men who have sex with men, transgender people);
- Interfederative technical cooperation to strengthen intersectoral actions in the territory, prioritizing social control (induction of agendas involving health, education and social assistance);
- Interfederative technical cooperation to monitor the development of project actions in the situation rooms.

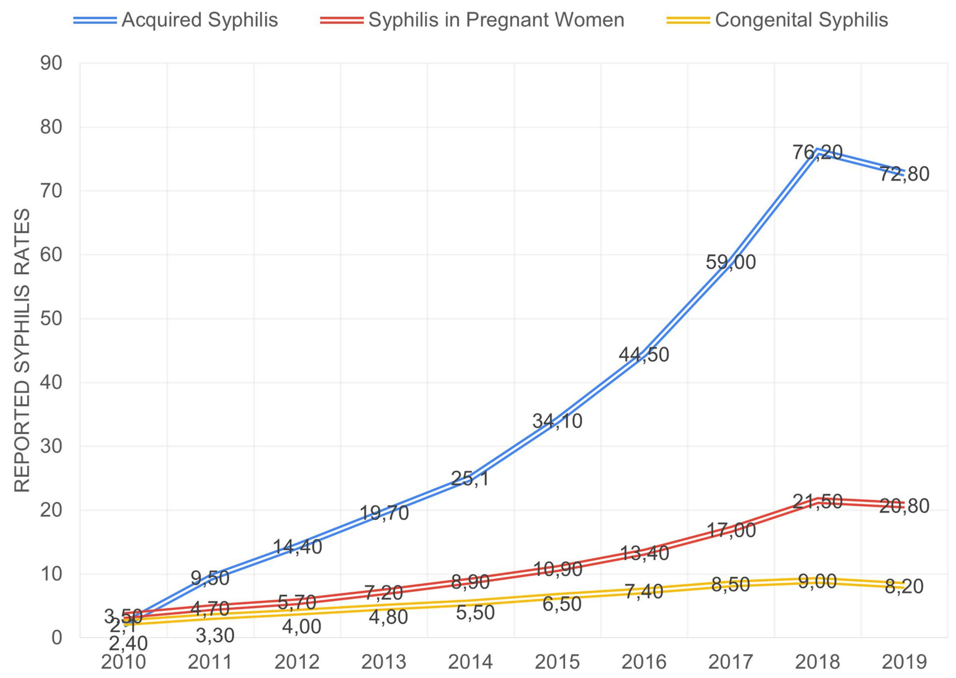


FIGURE 1
The incidence of syphilis in Brazil from 2010 to 2019.

accessible urban spaces; Messages were posted on social networking sites (Facebook, Instagram, and Twitter) by influencers.

The media, especially the press, play a pivotal role in selecting and creating messages to represent facts (6, 7), promoting public health, influencing public opinion, and generating awareness about health issues. Both in content and form, news stories paint a picture of the 'real world,' whose framing, through narrative conventions, can be refined through topics of discussion to social conversations (8). The press and news media report on matters of interest to the public sphere not necessarily to tell people 'how to think,' but 'what to think about' (9).

This paper relies on the premise that a nationwide public communication campaign will likely rely on news coverage (10). Campaigns aim to exert direct influence on those identified as target population by delivering evidence-based health information and addressing disinformation and misinformation. However, experience with political campaigning has shown that capturing the media's attention is the immediate outcome, and "[i]mplicit in this campaign perspective is the idea of the public agenda because control of media setting implies significant influence over the public's agenda" (9).

The prominence of a topic in the media's agenda creates 'salience' (9, 11, 12). In other words, a topic is highlighted in a given period compared to others based on one or more attributes, according to the framing by the media (9, 11, 12) through a gatekeeping process, which selects the facts that become news and are disseminated by media outlets (9, 13).

In addition, in a systematic literature review performed by our research group, we observed a gap in assessing the impact of public health campaigns, regarding the use of online data (i.e., online news) and others user-generated Internet content (14). Thus, this study analyzes online news on syphilis in Brazil by mapping key elements of the news posted online between 2015 and 2019 and measuring their quality. The study covers two time periods: (1) 2015 to 2017, related to communication actions prior to the Syphilis No! Project, and (2) 2018 to 2019, related to the news coverage of communication actions carried out during the Syphilis No! Project.

In this study, we hypothesized that the quantity and quality of news about syphilis disseminated in the form of earned media are significant indicators for measuring the impact of communication actions developed by the Syphilis No! Project throughout 2018 and 2019. Earned media can influence the public agenda and people's decision-making (9, 11, 12) due to the breadth of timely and sound information provided to the general public. Indicators related to earned media coverage could be used to measure the success of a public health campaign (15, 16), to examine the effect of communication actions (17, 18) and to improve public policy and public health interventions.

2. Materials and methods

We conducted a field investigation before characterizing and qualifying news through Hermes. The idea was to outline quality indicators of information about syphilis in the media agenda. This step identifies a type of attribute (9, 11, 12) about the framing of reported news. We determined the indicators, then extracted and performed a content analysis of the web pages selected. The workflow performed in this section is presented in Appendix Figure A1.

We used the Google search when identifying the relevant web pages, using the keywords: 'syphilis,' 'syphilis campaign,' 'AIDS,' 'AIDS campaign,' and 'sexually transmitted diseases.' We narrowed the search to documents written in Brazilian Portuguese. From the search results, two researchers, specialists in public health communication,

TABLE 1 Quantitative distribution of exploratory categories found in the training data from the search results performed by researchers.

Exploratory categories	Generic term	Number of text fragments in the news
Definitions of syphilis	Disease definition	90
Epidemiological data on syphilis	Epidemiological indicators	96
How to prevent syphilis	Prevention	73
How or where to get a syphilis rapid test/diagnosis	Rapid test/diagnosis	76
Consequences of syphilis in key populations and the risk of lethality from tertiary syphilis	Consequences	75
Public communication campaigns	Campaign	65
Effective treatment for syphilis	Treatment	68

manually accessed the most relevant web pages and selected 153 texts, that were characterized as news (19), based on elements that constitute a news piece, namely, the headline, the lead (the paragraph that summarizes the "who, when, where, what, why, and how"), and the body, which further elaborates on the elements mentioned in the lead.

In order to map out relevant features in the news identified, we used thematic analysis and axial coding to identify emerging themes and sub themes (20). Through indexicality, that is, the process of analyzing text fragments that constituted the news that were analyzed, we categorized, by induction, news about syphilis into the following exploratory categories (or quality indicators): definitions of syphilis; epidemiological data on syphilis; how to prevent syphilis; how or where to get a syphilis rapid test/diagnosis; consequences of syphilis in key populations (such as pregnant and infants) and risk of lethality due to tertiary syphilis; effective treatments to cure syphilis; and public communication campaigns (details provided in the Appendix Topic A2).

This analysis resulted in seven generic news categories: Disease Definition; Epidemiological Indicators; Prevention; Rapid Test/Diagnosis; Consequences; Campaign, and Treatment. Table 1 provides an overview of the exploratory categories and their quantitative distribution of text fragments across the body of the 153 news analyzed. These text fragments will be defined as 'training data' and after homogenization and standardization, they will serve to verify if the most important words of each category are contained in the online news.

In addition, we designed an online survey on the identified exploratory categories and recruited seven specialists in communication and public health to analyze the importance of their presence in news items as a way of qualifying texts about health problems. The evaluators had experience in the following areas: federal health management; state health management; primary care professional or municipal management; management in the Health Council; university research; communication and health; and journalism. Each question individually evaluated the importance of the categories through a balanced Likert Scale as follows: not important; slightly important; moderately important; very important; and extremely important.

Figure 2 highlights the varying levels of importance assigned to different categories related to the news items. Categories such as prevention, treatment, and consequences are consistently rated as highly important, while others, such as disease definition and rapid test/diagnosis, also hold significant relevance. This assessment supports our findings regarding the importance of qualifying health news using the defined criteria.

Once the exploratory categories and training data were defined, we proceeded to the second step of the workflow, which consisted of identifying the volume and quality of online news about syphilis through Hermes.

Based on a multidimensional analysis framework, such an ecosystem monitors the reach of public health actions related to campaign activities, education, communication, and epidemiological surveillance (4). It used computational tools and machine learning to monitor progress with Syphilis No! project and assist managers and decision-makers in evaluating the effects of public health policy interventions and communication campaigns. Panel 2 describes the elements, the computational tools and machine learning approaches used for analysis in the Hermes ecosystem with more details provided in the Appendix.

Panel 2: Hermes – A Digital Ecosystem to Assess Public Health Policies

A software ecosystem refers to a collection of software products that have some given degree of symbiotic relationships (21), used by a set of actors on top of a common technological platform (22). A software ecosystem must be seen from multiple perspectives, such as technology, communication, and epidemiological surveillance.

Hermes focuses on bringing knowledge to stakeholders through the heterogeneous data collection, processing, integration, and visualization of this information through a single platform, using computational tools, machine learning techniques, and statistical analysis that enable public health policy evaluation.

Features:

Actor-Centered

It recognizes the different actors, e.g., stakeholders, decision-makers, communication campaign analysts, and financial and public managers.

Data-Driven

Enables collecting heterogeneous data and processing it through computational techniques using machine learning. Facilitates data interpretation, transforming it into feasible information for assessment. Strengthens decision-making based on prediction and understanding of a problem.

Heterogeneous Data Sources and Data integration

Hermes collects and integrates several data sources, such as Google Search, Google Trends, epidemiological surveillance databases of the Ministry of Health, databases of Massive Open Online Courses (MOOCs) collected from SUS' Virtual Learning Environment (AVASUS); social networks (Instagram, Twitter, Facebook, and YouTube); in addition to allowing manual entry or upload of files comprising data from Public Health Campaigns. It ensures consistency, quality, and completeness for correlation and causality checking.

Automated Processes

It automates processes based on machine learning algorithms, natural language processing, and statistical analysis: (i) text mining and feature extraction; (ii) sentiment analysis; (iii) clustering; (iv) time series decomposition; (v) interrupted time-series segmented regression analysis; and (vi) coefficient correlation analysis between variables of interest.

FIGURE 2
Findings of the analysis on the significance of quality criteria for evaluating texts related to health issue.

TABLE 2 Web pages categorized after main text content extraction.

Type	Rationale	Number
Page not found / Content not available	Page no longer exists, or content is exclusive to members.	42
Search results	Page content results from a search within the website.	69
Non-text content	Page does not include text content (e.g., podcasts, images, videos, PDF, or PPT files).	123
Scientific research	The main content is part of a scientific research summary, such as an article, poster, thesis, or dissertation.	185
News	The main content has characteristic features of a news report.	630

In this study, from January 2015 to December 2019, Hermes collected online content that was Google indexed and included the term 'syphilis' in Brazilian Portuguese web pages. It performed this task by using a Google Search API that allows retrieving, in JSON format, a set of websites indexed by keyword searched for (23). The fields returned by the automatized search results included: the date when the news was indexed, URL, title, and a snippet of the news. 1,049 web pages were found in this period using these parameters.

Hermes has a content extraction module that uses the Newspaper Python library (24) and extracted the following fields from each collected URL: title, publication date, keywords, running text, and abstract.

At this point, it was necessary to identify if these 1,049 web pages were, in fact, news. Thus, we defined a typology that made it possible to identify which web pages featured online news. A manual identification was carried out and 630 web pages were defined as news (Table 2).

To substantiate and build the automated extraction of features found in the online news through Hermes, we used the text mining technique, otherwise known as Text Data Mining, which calculates the weight of a keyword (score) in a vector model (matrix) (25).

We checked unigrams (individual terms), bigrams (two consecutive terms), and trigrams (sequences of three terms) through a comparative analysis to ascertain which keywords combination could be the most suitable for extracting characteristic features included in news. We chose to use bigrams since they provide a richer contextual representation than unigrams, as they capture the proximity relationship between adjacent words and have a smaller dimensionality than the trigrams. It allowed a more refined analysis of news texts' specific characteristics and contexts. For example, a bigram like "syphilis campaign" can convey more accurate and meaningful information than the isolated terms "syphilis" or "campaign." The choice of bigrams is directly aligned with the study's objectives, which aimed to identify specific characteristics (exploratory categories) of news items. By capturing meaningful relationships between adjacent words, bigrams allowed for extracting more accurate and relevant characteristics, providing valuable insights into news content.

The following procedures were executed in order to homogenize and standardize the bigrams: (i) changing the capitalization of words

TABLE 3 Main bigrams found for each exploratory category in the training sample.

Category	Bigrams
Disease definition	Bacteria cure; sexually transmitted; transmitted caused; curable treatment; simple treatment; manifest infection; sexually disease; called syphilis
Epidemiological indicators	Year age; health increase; new cases; cases type; congenital period; types syphilis; syphilis parents; thousand people; last bulletin
Prevention	Condom distribution; better condom; through use; prevention use; doctors reinforce; need condoms; reinforce need; use condoms
Rapid test / Diagnosis	Serologic test; rapid test; syphilis result; test detected; easily detected; rapid used; used diagnosis; result minutes
Consequences	Case not; baby development; cause complications; lead death; genital sores; genital rashes; cause abortion
Campaign	Syphilis campaign; no syphilis; municipal campaign; activities integrate; campaign ending; campaign intended
Treatment	With penicillin; penicillin treatment; based antibiotics; base treatment; antibiotic treatment; treated syphilis

to lowercase, (ii) deleting special characters, (iii) removing all punctuation, (iv) removing extra spaces, (v) removing accents and numbers, and (vi) deleting "stop words" (words that had no meaning, such as adverbs and prepositions). Finally, the sentences were converted into bigrams, which comprise a dictionary of words representing each group (the training sample and the online news about syphilis). Table 3 provides examples of the main bigrams found for each exploratory category in the training sample.

This process yielded a matrix, with each row corresponding to a news item and each column to the bigrams. The cells refer to the score of bigram's relevance in the news, and they were estimated through Term Frequency – Inverse Document Frequency (TF–IDF) (26). This technique statistically determines the importance of a word in a document corpus relative to other texts within the same database. The weight of a word for such a document is contingent on the number of times it appears in it, but is offset by the frequencies of the words in the other documents within the same database (27). So, words that are common in every document, such as *this*, *what*, and *if*, rank low even though they may appear many times, since they do not mean much to that document in particular.

The TF–IDF used to fit these data is straightforward.

$$\mathrm{TF-IDF}(t,d,D) = \mathrm{TF}(t,d)^{*}\ \mathrm{IDF}(t,d,D)$$

Where the Term-Frequency (TF) measures up to how many times (freq) the word t exists in the document d. This frequency is calculated as follows:

$$\mathrm{TF}(t,d) = \mathrm{freq}(t,d)$$

For example, in this collection of documents: "this is the first document.," "this document is the second document." and "Is this the first document?," the TF of the word document is 1 for the first document, 2 for the second document and 1 again for the third document.

The IDF is used to measure the t score from the frequency in d and in the collection of documents. The IDF is defined by the log between the total of documents (N) and the frequency of documents d where the term t occurs (dft). This frequency is calculated as follows:

$$IDF(t,d,D) = \log\left(\frac{N}{dft}\right)$$

Using the same example, the IDF of the *t document* is log(3/3), log(1), thus the IDF of the *t document* is 0. Therefore, the final score of *t* is the weight resulted by frequency the *t* in the *d* (TF) and the inverse document frequency (IDF).

Subsequently, all scores of bigrams found within news' characteristic features were summed, then a final score was defined. Finally, considering the results, the Hermes ecosystem, in an automated process, determined the acceptance threshold for category identification for each of the seven categories identified, based on the category's average score.

The acceptance threshold is responsible for guaranteeing that one news holds the minimum score for the presence of features to be identified. Thus, we verified whether or not the score applied is greater than or equal to the acceptance threshold. If so, the category identified characterized such a news. This procedure was used for all 630 news extracted and for all seven categories we defined, in which a news can include none, one, or more exploratory categories.

Quality parameters of news were determined based on the presence and combination of categories in them, as follows: low quality (0–2 categories), medium quality (3–4), high quality (5 or more).

Using Spearman's Rank Correlation Coefficient (28), we examined the association between the number of news over the period January 2015 to December 2019, grouped by their respective quality levels (low, medium, and high), and the number of serology tests for syphilis diagnosis performed in primary health care in Brazil. The data on number of tests were obtained through the Outpatient Information System (Sistema de Informação Ambulatorial, SIA) of the Unified Health System (Sistema Único de Saúde, SUS), available on the MoH webpage.[1]

SIA/SUS allows for the retrieval of test results based on monthly and yearly quantities, as follows: (i) Treponemal test for syphilis detection, (ii) Fluorescent Treponemal Antibody-Absorption (FTA-ABS) IgG test for syphilis diagnosis, (iii) Fluorescent Treponemal Antibody-Absorption (FTA-ABS) IgM test for syphilis diagnosis, (iv) Nontreponemal test for syphilis detection, (v) Nontreponemal test for detecting syphilis in pregnant women, (vi) Rapid Syphilis Test, (vii) Rapid syphilis test for detecting the infection in pregnant women or fathers/partners. The collected data for testing includes the period from 2015 to 2019.

1 https://datasus.saude.gov.br/acesso-a-informacao/producao-ambulatorial-sia-sus/ (accessed on 25 November 2020).

TABLE 4 Distribution of exploratory categories found in analyzed news on syphilis per year, from 2015 to 2019.

Exploratory category	2015	2016	2017	2018	2019
Number of online news item about syphilis per year	18 (2.86%)	26 (4.15%)	42 (6.67%)	107 (16.98%)	437 (69.37%)
Campaign	11	16	32	78	308
Consequences	15	21	41	91	352
Disease definition	11	14	26	67	241
Epidemiological indicators	13	11	29	55	243
Prevention	15	21	35	86	360
Rapid test/Diagnosis	13	25	39	92	382
Treatment	16	22	34	88	348
None of the categories	1	1	0	11	22
Total number of news	630				

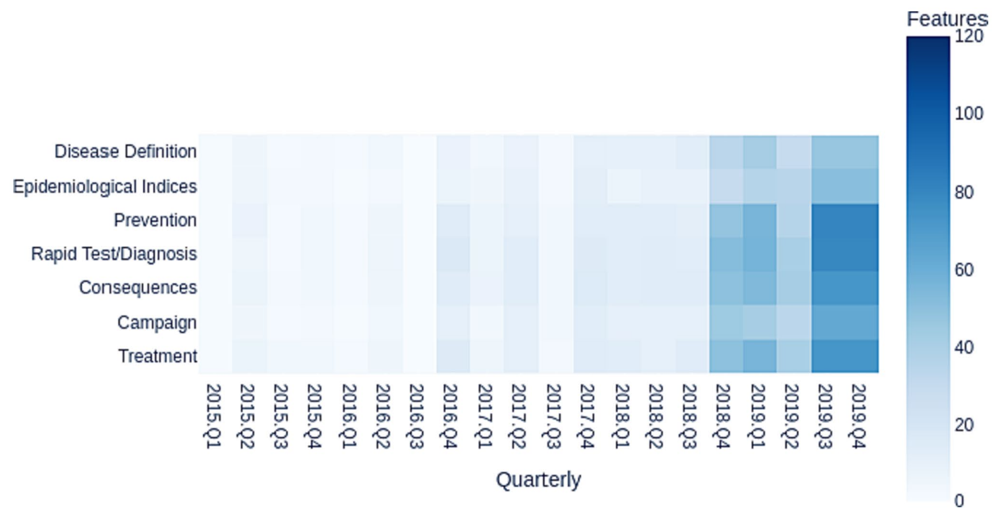


FIGURE 3
Heat map of categories by quarter from 2015 to 2019.

We also repeated the analysis using Pearson's and Kendall's coefficients, but Spearman's Rank Correlation Coefficient method (also referred to as Spearman's rho) revealed better results. Perhaps because (i) data are generally not distributed across the two variables, (ii) there is a monotonic relationship among data, and (iii) both variables are ordinal (29).

3. Results

A total of 1,049 web pages were gathered through the utilization of the Google Search API. Among these, 630 pages were classified as earned media. An upward trend in the quantity of syphilis-related news item was observed over the years, with 18 articles in 2015, 26 in 2016, and 42 in 2017. However, a significant surge in news coverage occurred in 2018 (107 articles) and 2019 (437 articles). Despite this increase, the proportion of high-quality news remained consistently high, with 77.6 and 70.5% respectively, aligning with previous years. Our findings indicated a correlation between news quality and the performance of syphilis testing in primary healthcare settings, demonstrating an increase of 82.32, 78.13, and 73.20%, respectively, across the three types of treponemal tests used to confirm an infection.

Table 4 provides the total number of news mapped by year, along with exploratory categories found. The number of news increased gradually from 18 in 2015 to 26 in 2016 and 42 in 2017, then rose substantially to 107 in 2018. In 2019 the number of news increased to 437 – a figure higher than the sum of the number of news in the four preceding years.

Figure 3 shows the mapping of categories by quarter from 2015 to 2019. Each cell represents the number of news with respective categories. In 2015 the highest number of news was in Q2 (April–June), citing the terms "prevention," "consequences," and "treatment." The periods of highest intensity in 2015 is related to the festive seasons in Brazil (June festivals) and the month of October, when the Ministry of Health dedicates to discussing Sexually Transmitted Infections (STIs) and disseminates data from the epidemiological bulletin in Brazil. In 2016 and 2017 the highest number of news was in Q4 (October–December), citing the terms "campaign," "consequences," "test/diagnosis," and "treatment," coinciding with the National Day to Combat Syphilis and Congenital Syphilis in October. In the rest of the

TABLE 5 Number of news each year categorized by quality.

Quality level	2015		2016		2017		2018		2019	
	n	% of total	*n*	% of total	*n*	% of total	*n*	% of total	*n*	% of total
High	13	72.2%	19	73.1%	33	78.6%	83	77.6%	308	70.5%
Medium	2	–	2	–	6	–	10	–	71	–
Low	3	–	5	–	3	–	14	–	58	–
Total	18	–	26	–	42	–	107	–	437	–

TABLE 6 Correlation between the number of tests and the number of news according to their quality.

Types of syphilis tests	News quality		
	High	Medium	Low
Treponemal test for syphilis detection	73.20%	67.15%	62.78%
Nontreponemal test for syphilis detection	69.15%	62.04%	58.66%
FTA-ABS IgG test for syphilis diagnosis	78.13%	62.12%	64.58%
FTA-ABS IgM test for syphilis diagnosis	82.32%	68.84%	68.31%
Nontreponemal test for detecting syphilis in pregnant women	64.92%	58.76%	50.85%
Rapid syphilis test	58.98%	46.31%	42.64%
Rapid syphilis test for detecting the infection in pregnant women or fathers/partners	65.33%	56.27%	52.07%

year these categories were hardly ever or never mentioned in the news. As of 2016 Q4, the patterns begin to gain greater intensity due to the Ministry of Health's declaration regarding the syphilis epidemic in Brazil. In this way, the theme gained more space in the Brazilian media agenda than in previous years.

In 2017, the news focused on the impact of the declaration of the syphilis epidemic in Brazil made at the end of 2016. However, the campaigns were still focused on congenital syphilis, produced with low investment, and disseminated mainly on social networks. As of 2018, SNP actions were intensifying, holding seminars with state and municipal entities in all regions of Brazil on the strategy to combat syphilis. Then the national campaign was launched, maintaining greater journalistic coverage in the period.

The "Syphilis No!" Project was part of this strategy to change course in the fight against syphilis in Brazil, involving more resources for publicity campaigns, including intelligence actions that mobilized states and municipalities. Qualified actors selected in each territory acted in this project, which made up a support network for research and project actions, as detailed in Panel 1.

In 2018 and 2019, while there was a preponderance of news in Q4 of each year, and in particular October, there were news throughout the year. The majority of news mentioned "campaign," "consequences," "rapid test/diagnosis," "treatment," and "prevention."

Table 5 shows the number of news categorized by quality. While the volume of news increased substantially in 2018 and 2019, the relative proportion of high-quality news remained consistently high (77.6 and 70.5% respectively) and in line with similar years.

There was a correlation between the quality of news about syphilis and the number of serology tests performed in primary health care for diagnosing syphilis as shown in Table 6. The rapid tests for syphilis in pregnant women or fathers/partners showed a 65.33% correlation with high-quality news, while rapid tests in the general population had a 58.98% correlation. Nontreponemal tests for syphilis showed a 69.15% correlation with the high-quality news and treponemal tests (confirmatory tests for syphilis) showed 82.32, 78.13, and 73.20% over the same period.

4. Discussion

Implementing effective communication strategies to facilitate the widespread circulation of accurate and reliable information is crucial in order to enhance the dissemination of public health interventions. Our findings demonstrate a significant surge in news coverage through earned media as well as increased testing, indicating a positive correlation between the two. This highlights the potential of continuous and comprehensive national mass media outlets as an instrument for promoting public policies addressing health crises.

The number of news on syphilis rose rapidly following the implementation of the communication strategy as part of the Syphilis No! Project. While there was a cumulative total of 86 news on syphilis in 2015 to 2017, the number of news rose to 107 in 2018 (154.8% increase compared to 2017) and 308 in 2019 (308.4% increase compared to 2018).

In 2010, syphilis became a notifiable disease in Brazil and was declared a public health emergency in 2016. However, these major public health policies did not lead to a change in the number of news or an increase in testing and diagnosis (2, 30).

While in May 2015, the MoH conducted a campaign on mother-to-child transmission (MTCT) for Mother's Day and in 2016 and 2017 ran communication campaigns on the National Day to fight Against Syphilis and Congenital Syphilis, these were one-off events (31) and had no meaningful effect on news and testing.

The launch of the communication actions of the Syphilis No! Project in February 2018 coinciding with the Brazilian Carnival was a turning point. A campaign on social networks with local and state initiatives, followed. In March 2018, the SNP launched intervention agendas across the country through the Ministry of Health, with an emphasis in 100 priority municipalities (3, 4, 32). In November 2018, the national campaign "Remember to take care of yourself" (2018–2019) was launched. The communication campaign meant that actions were not have meant that communication was no longer limited to one-off annual events, such as the National Day to fight Against Syphilis and Congenital Syphilis, but were spread throughout the year (31) involving a myriad of products and communication actions, including in 'paid and owned media' that helped to increase news in earned media.

Carrying out media campaigns in support of the adoption and dissemination of public health interventions is a necessary condition to generate in society a regime of attention and visibility for a problem such as syphilis. The media can generate a legitimation process and make the problem visible on a national scale. However, to raise awareness and change habits in the population, developing a set of articulated actions in the territories is necessary.

Notably, public policies to combat syphilis had no significant effect between 2010 and 2016, clearly observed by the increase in cases and the low amount of news on the subject in recent years (2015–2017). However, as of 2018, it is possible to observe the influence of communication efforts through the support of the theme in media coverage.

It is important to emphasize that the SNP intervention actions go beyond the communication area since actions developed mainly by Research and Intervention Supporters in priority cities with a high rate of cases of congenital syphilis helped to: strengthen the health care network and the different care spaces for the implementation of syphilis care lines; implement syphilis epidemiological surveillance rooms at the municipal level; evaluate actions to combat syphilis at municipal and district levels. The Research and Intervention Supporters whose actions helped improve coordination of communication efforts and the messaging among state and local health offices and the local press, acting as an "opinion leader," to establish a "two-step flow of communication" (33) to ensure high-quality news. The communication actions led by SNP produced greater "resonance," with local and national press (17).

In addition, actions of a universal scope were also carried out throughout the territory, such as: the purchase and distribution of crystalline and benzathine penicillin; the purchase and distribution of rapid syphilis tests; reinforcement of the laboratory structure for diagnosis; and the instrumentation of situation rooms in all Brazilian districts and the Federal District.

The SNP included various actions that enhanced the impact of the communication campaign and ensured that the topic remained prominent on the agenda-setting it apart from previous years.

McCombs (9) argues that the agenda-setting process depends not only on the time period of media exposure a topic has but also on the potential correspondence to the audience's need for orientation on the topic. The results reveal a substantial increase in news in earned media and in testing, with a correlation between the two, suggesting that a sustained communication campaign could be a powerful tool for promoting public policies to tackle health crises.

The Ministry of Health recorded the highest number of syphilis tests performed per 1,000 population in 2018 (2,1 million) and 2019 (2,5 million), compared to 1,4 million in 2017. While in 2010–2018, the number of cases of acquired syphilis, syphilis during pregnancy, and congenital syphilis increased substantially, but fell in 2019 (4, 32, 34).

Digital information and communications technologies in health and the application of computational methods based on artificial intelligence can be used to develop predictive analytics to inform real-time response to effectively manage public health outbreaks and crises. The Hermes digital ecosystem developed for the project, played a critical role in monitoring campaign progress, regularly capturing data registering, and processing of information related to communication actions, health system interventions (testing) and epidemiological parameters (number of cases of acquired syphilis, syphilis in pregnant women and congenital syphilis). The use of computational methods that enabled analysis of data of heterogeneous nature to examine the public health response and its results in real time throughout the country to provide a powerful tool in planning and monitoring of the public health campaign for syphilis and can be transferred to other public health challenges.

Nowadays, for an individual to perceive a subject as relevant, such subject needs to be highlighted in the media with a particular frequency, being highlighted in their agenda. Thus, if a subject gains greater exposure in the media for some time, it becomes seen as important by the public. For example, until 2017, syphilis was not highlighted in the Brazilian media. Therefore, it was considered a neglected disease.

As of 2018, a set of systematic and strategic communication actions has been developed, supported by technological, financial and intellectual resources. The actions guided the priority target audiences, health managers, and professionals nationwide. These actions intensively positioned syphilis in the Brazilian media agenda from 2018 onwards, drawing public attention. The lack of qualified information, as it is a neglected disease, generated a feeling of uncertainty in public, who started to seek more information and seek health units and learn about the forms of diagnosis and treatment. Notably, an effective and ongoing communication campaign can promote public policies and provide efficient responses to health crises.

There are potential limitations of our study. The first one relates to the completeness of the news collection used as a dataset, an external threat mitigated by choosing one of the largest existing content indexers, namely Google Search. However, it would be a mistake to assume that Hermes can retrieve every existing online news through the Google Search API. In addition, Hermes used filters for: (i) language, which narrows the search to documents written in Brazilian Portuguese, and (ii) geolocation, which limits the search results to documents originated from Brazil. That may constitute a barrier insofar as the geolocation parameter checks the domain (URL) and the geographical location of the Web server's IP address. Future works may add other search engines, such as Bing and Yahoo, to expand the search result for news items on the Internet. The second limitation relates to the search results related to the 42 missing pages Google indexed. If they had been incorporated into the analysis, they would likely increase the number of resulting news. However, when analyzing the date of news of missing web pages, we observed they had the same proportionality of 2015 ($n = 1$), 2016 ($n = 3$), 2017 ($n = 1$), 2018 ($n = 15$), and 2019 ($n = 22$) results.

Notwithstanding limitation, the study reveals the effective application of a digital health system that incorporates all the elements of a complex public health campaign that included a communication campaign, education, health system interventions, training, expanded access to testing and treatment.

The ability to explore online news through machine learning has aroused the interest of parallel study groups, bringing new insights for stakeholders to analyze public health campaigns from different perspectives, such as sentiment analysis techniques (35) and similarity analysis (36).

5. Conclusion

This study from Brazil, an upper-middle income country, led by a multidisciplinary group of researchers involving public health specialists, clinicians, experts in computational science, data scientists, educationalists, communication experts and marketing experts reveals the successful implementation of public health actions with a communication campaign that led to major increases in online news related to syphilis and the shift in the media landscape and the public health response after the Syphilis No! Project interventions.

The Hermes ecosystem was able to effectively capture the number and frequency of news stories before and after the Syphilis No! Project, effectively classify news according to informational categories and identify high-, medium-, and low-quality news, and to examine the relationship between the communication campaign and public health results to orient more effective and targeted health policies and interventions to manage the syphilis epidemic in Brazil.

The study reveals the utility of integrated digital health information systems in guiding public health policies and actions to ensure effective responses to public health challenges across all country income categories.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

RP, JL, AA, LSi, and RV: conceptualization. RP, RF, JL, AA, LSi, and RV: methodology. RP, RF, and HG: software, formal analysis, and data curation. RP, JL, LSi, AA, RF, TL, AM, LSa, HG, RA, and RV: validation and writing – review and editing. RP, JL, AA, RF, LSi, and RV: investigation. RP, LSi, and RV: resources. RP, JL, AA, RF, and TL: writing – original draft preparation. RP and RF: visualization. JL, LSi, RV, and RA: supervision. RV: project administration and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by a grant to the Syphilis No! Project from Brazilian Ministry of Health (Project Number: 54/2017). The funders had no role in study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1248121/full#supplementary-material

References

1. Atun R, de Andrade LOM, Almeida G, Cotlear D, Dmytraczenko T, Frenz P, et al. Health-system reform and universal health coverage in Latin America. *Lancet*. (2015) 385:1230–47. doi: 10.1016/S0140-6736(14)61646-9

2. da Saúde Secretaria de Vigilância em Saúde BM. Boletim Epidemiológico – Sífilis. Brasília (DF). (2020). Available at: http://www.aids.gov.br/pt-br/pub/2020/boletim-sifilis-2020

3. Pinto R, Valentim R, Fernandes da Silva L, Fontoura de Souza G, Góis Farias de Moura Santos Lima T, Pereira de Oliveira CA, et al. Use of interrupted time series analysis in understanding the course of the congenital syphilis epidemic in Brazil. *Lancet Region. Health Americas*. (2022) 7:100163. doi: 10.1016/j.lana.2021.100163

4. de Morais Pinto R, de Medeiros Valentim RA, Fernandes da Silva L, Góis Farias de Moura Santos Lima T, Kumar V, Pereira de Oliveira CA, et cc. Analyzing the reach of public health campaigns based on multidimensional aspects: the case of the syphilis epidemic in Brazil. *BMC Public Health*. (2021) 21:1632–13. doi: 10.1186/s12889-021-11588-w

5. Lacerda JS, Muneiro LC, Oliveira JRM, de Araújo KS, de Araújo ACC, Paiva JCL, et al. Sentidos produzidos sobre campanhas de prevenção da sífilis: algumas perspectivas sobre possíveis mediações. In *Comunicação e Mediações: novas perspectivas*. Eds. E. Trindade, L. V. B. Maluly, M. A. Pavan and M. L. Fernandes. ECA-USP: São Paulo (2021). 208–26. doi: 10.11606/9786588640272

6. Schudson M. Approaches to the sociology of news In: J Curran and D Hesmondhalgh, editors. *Media and society*. New York: Bloomsbury, W.W. Norton & Company (2019). 139–66.

7. Lippmann W, Curtis M. *Public opinion*. New York: Routledge (2017).

8. Schudson M. *A política da forma narrativa: a emergência das convenções noticiosas na imprensa e na televisão. Jornalismo: questões, teorias e "estórias"* Lisboa: Vega. (1993); p. 278–293.

9. McCombs M. *A teoria da agenda: a mídia e a opinião pública*. Petrópolis, RJ: Vozes (2009).

10. Porto MP. Lutando contra a AIDS entre meninas adolescentes: os efeitos da Campanha de Carnaval de 2003 do Ministério da Saúde do Brasil. *Cad Saude Publica*. (2005) 21:1234–43. doi: 10.1590/S0102-311X2005000400026

11. Agenda MCM. Setting, Media effects on In: JD Wright, editor. *International encyclopedia of the social & behavioral Sciences*. *2nd* ed. Oxford: Elsevier (2015). 351–6.

12. Silva JAB. A expansão da teoria do agenda-setting em sistemas informativos da Web. *Galáxia*. (2014) 14:262–73. doi: 10.1590/1982-25542014216757

13. Ferreira GB. Gatekeeping changes in the new media age: the internet, values and practices of journalism. *Research*. (2018) 14:486–505. doi: 10.25200/BJR.v14n2.2018.1026

14. Pinto RM, Silva LF, Valentim RAM, Kumar V, de Gusmão CMG, de Oliveira CA, et al. Systematic review on information technology approaches to evaluate the impact of public health campaigns: real cases and possible directions. *Front Public Health*. (2022) 9:715403. doi: 10.3389/fpubh.2021.715403

15. Lazarin LR, Slongo LA. Mídia Paga ou Espontânea? Percepções de profissionais de marketing e comunicação sobre a importância de anúncios publicitários e notícias. *Rev Bras Gestão e Inovação*. (2015) 3:1–24. doi: 10.18226/23190639.v3n2.01

16. Neves LFF, Stasiak D, de Castro Basile R. Release e geração de mídia espontânea: um estudo de caso do Projeto Visibilidade UFG1. In: Proceedings of the 40o Brazilian congress of communication sciences, São Paulo, Brasil; (2017). p. 1–15.

17. Pessoni A. C. C. Releases sobre saúde nas assessorias de imprensa das administrações públicas do ABC: produção e tendências. In: Proceedings of the 37o Brazilian congress of communication sciences; (2014). p. 1–14.

18. Harris P, Kent J, Sainsbury P, Marie-Thow A, Baum F, Friel S, et al. Creating "healthy built environment" legislation in Australia; a policy analysis. *Health Promot Int*. (2018) 33:1090–100. doi: 10.1093/heapro/dax055

19. Qing S, Ying Z, Pengzhou Z. Research review on key techniques of topic-based news elements extraction. 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS). Wuhan, China: IEEE (2017). p. 585–590. doi: 10.1109/ICIS.2017.7960060

20. Orozco G, González R. *Una coartada metodológica. Abordajes cualitativos en la investigación en comunicación, medios y audiencias*. México: Tintable (2011).

21. Messerschmitt DG, Szyperski C. *Software ecosystem: understanding an indispensable technology and industry*. Cambridge: MIT Press (2003).

22. Manikas K, Hansen KM. Software ecosystems – a systematic literature review. *J Syst Softw*. (2013) 86:1294–306. doi: 10.1016/j.jss.2012.12.026

23. Gavali R. Discovery service for engineering and technology literature through google custom search: a case study. *DESIDOC J. Library Inform Technol*. (2015) 35:417–21. doi: 10.14429/djlit.35.6.8825

24. Ou-Yang L. Newspaper: article scraping & curation. *Python Library Retrieved*. (2013). Available at: https://newspaper.readthedocs.io/en/latest/.

25. Feldman R, Sanger J. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge: Cambridge University Press (2007).

26. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. *Inf Process Manag*. (1988) 24:513–23. doi: 10.1016/0306-4573(88)90021-0

27. Hong TP, Lin CW, Yang KT, Wang SL. Using TF-IDF to hide sensitive itemsets. *Appl Intell*. (2013) 38:502–10. doi: 10.1007/s10489-012-0377-5

28. Kumar A, Abirami S. Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method. *Inf Sci*. (2018) 460:23–41. doi: 10.1016/j.ins.2018.05.003

29. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*. (2011) 30:87–93. doi: 10.2478/v10117-011-0021-1

30. da Saúde Secretaria de Vigilância em Saúde BM. *Strategic actions agenda for reducing syphilis in Brazil*. Brasília (DF): Ministério da Saúde - Brasília (2017) Available at: http://www.aids.gov.br/pt-br/pub/2017/agenda-de-acoes-estrategicas-para-reducao-da-sifilis-no-brasil.

31. Lacerda JdS, Muneiro LC, Oliveira Junior MS, Araujo KS, Paiva JCL. *Campanha de combate à sífilis 2018–2019: das condições de produção às estratégias comunicativas*. São Paulo: ECA-USP. (2020); p. 162–176.

32. De Andrade IGM, De Medeiros Valentim RA, De Oliveira CAP. The influence of the No Syphilis Project on congenital syphilis admissions between 2018 and 2019. *Brazilian J Sex Trans Dis*. (2020) 32:1–6. Available at: https://bjstd.org/revista/article/view/892.

33. Sharma N, Nahak D, Mohan F. Public health communication by health workers during COVID-19: a study of Jalandhar District of Punjab. *J Contemp Issues Bus Govern*. (2021) 27:2515–22.

34. Valentim RAM, Caldeira-Silva GJP, da Silva RD, Albuquerque GA, de Andrade IGM, Sales-Moioli AIL, et al. Stochastic petri net model describing the relationship between reported maternal and congenital syphilis cases in Brazil. *BMC Med Inform Decis Mak*. (2022) 22:40. doi: 10.1186/s12911-022-01773-1

35. Marques T, Cezário S, Lacerda J, Pinto R, Silva L, Santana O, et al. Sentiment analysis in understanding the potential of online news in the public health crisis response. *Int J Environ Res Public Health*. (2022) 19:16801. doi: 10.3390/ijerph192416801

36. Cezario S, Marques T, Pinto R, Lacerda J, Silva L, Santos Lima T, et al. Similarity analysis in understanding online news in response to public health crisis. *Int J Environ Res Public Health*. (2022) 19:17049. doi: 10.3390/ijerph192417049

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 07 November 2023
DOI 10.3389/fpubh.2023.1277389

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Dorit Nitzan,
Ben-Gurion University of the Negev, Israel
Laura Uccella,
Lugano Regional Hospital, Switzerland

*CORRESPONDENCE
Jiang Xiaolian
✉ jiang_xiaolian@126.com

RECEIVED 14 August 2023
ACCEPTED 16 October 2023
PUBLISHED 07 November 2023

CITATION
Lihua M, Xiaolian J, Song W and Ning J (2023) Application of health self-management intervention program for metabolic syndrome patients in the bereaved population following the Wenchuan earthquake.
Front. Public Health 11:1277389.
doi: 10.3389/fpubh.2023.1277389

Application of health self-management intervention program for metabolic syndrome patients in the bereaved population following the Wenchuan earthquake

Ma Lihua[1,2], Jiang Xiaolian[1]*, Wang Song[1] and Jiang Ning[1]

[1]West China School of Nursing/West China Hospital, Sichuan University, Chengdu, China, [2]The First Hospital of Lanzhou University, Gansu, China

Background: The destructive Wenchuan earthquake has led to approximately 800,000 people being bereaved. In the previous cross-sectional study, we explored the long-term incidence of Metabolic Syndrome (MS) and studied its influencing factors among the bereaved population 12 years after the Wenchuan earthquake. Chronic disease self-management has become a recognized public health service. Studies have shown that demographic and genetic factors, stress, geographical environment, society, culture, dietary habits, lifestyle, and other aspects influence MS. Due to the Wenchuan earthquake being a serious stress event, the implementation of targeted interventions should be discussed further.

Objectives: To verify the effect of applying a self-management intervention program for patients with MS among the bereaved population following the Wenchuan earthquake.

Design: A randomized controlled trial (RCT) design was adopted.

Participants: A total of 132 bereaved patients with MS following the Wenchuan earthquake constituted the sample.

Methods: The study was based on the Cognitive–Phenomenological–Transaction, Chronic Disease Self-Management Program, and Patient Empowerment Conceptual Model, which combined with the latest evidence-based guidelines, were used to systematically evaluate cross-sectional results of this study that were used to construct a stress management-based health self-management intervention program and MS health self-management manual for bereaved patients with MS following the Wenchuan earthquake. In addition, we revised and completed a health self-management intervention program and health self-management manual for patients with MS by using the expert consultation method. General data were collected prior to intervention (T0). We collected the patients' MS disease-related physiological indicators before intervention (T0), after intervention (T1), and 2 months after intervention (T2). EipData3.1 software was used to input data in duplex and duplicate, and SPSS22.0 software was used for statistical analysis.

Results: The variance analysis showed that the total score of healthy self-management behavior and the score of diet management, exercise management, drug management, and emotional management have intergroup effects, time effects, and group–time interaction effects ($p < 0.05$). When the differences between groups were further compared, we found that the total score and the score of six dimensions (excluding disease self-monitoring management) were higher than those of the control groups at T1 and T2, and the differences were statistically significant ($p < 0.05$).

Conclusion: The intervention program of healthy self-management for patients with MS who come from bereaved families following the Wenchuan earthquake can effectively improve patients' health self-management behaviors.

KEYWORDS

metabolic syndrome, bereavement, self-management, intervention, RCT - randomized controlled trial

1. Introduction

Earthquake is a natural disaster with great destructive power (1). China is a developing country with frequent earthquakes, and it is also one of the countries with the most casualties due to earthquakes (2). Wenchuan earthquake in 2008 was not only the most destructive earthquake in China after the Tangshan earthquake in 1976, but also it is one of the most destructive earthquakes in the world in the first decade of the new century (3). In the earthquake, 69,227 people were killed, 374,643 people were injured, and 17,923 people were missing, bringing direct economic losses of 845.1 billion yuan to the Chinese people. According to statistics, Wenchuan earthquake caused a total of more than 80,000 people missing or died. According to the calculation of about 10 relatives of each deceased, the earthquake caused a total of about 800,000 people bereaved (4). Bereaved people experience property loss, burial of themselves/others, injury, disability, death of relatives, interruption of medical and health services, changes in production activities and lifestyles, as well as the pain of losing relatives and the destruction of family structures and functions, and many other earthquake-related disasters. If the population is in a state of continuous stress response for a long time, it will often induce a series of neuroendocrine reactions, mainly manifested as sympathetic nerve excitation and increased secretion of hypothalamic–pituitary–adrenal cortex (5), which makes corticotropin-releasing hormone Elevated levels, the increase of these hormones is closely related to body obesity and elevated blood sugar, blood pressure, blood lipids, etc. adverse effects. At present, most of the research on the health problems of the people after the earthquake in my country is limited to physical trauma and mental and psychological aspects. More than 80% of the researches focus on 3 months to 2 years after the earthquake drastically reduced (6).

Metabolic syndrome (MS) is a group of clinical syndromes mainly manifested by metabolic disorders such as central obesity (visceral obesity), insulin resistance, glucose metabolism, lipid metabolism and hypertension (7). MS is one of the escalating public health problems in most countries and regions in the world, with its prevalence ranging from 13.1 to 43.6%, and showing an increasing trend year by year (8). Studies have found that MS is not only a risk factor for cardiovascular disease (CVD), diabetes and chronic kidney disease (9–11), but also significantly increases prostate cancer, breast cancer, leukemia, pancreatic cancer, colorectal cancer, and liver cancer. And other cancer patients, resulting in a 33% increase in overall cancer mortality (12). MS is closely related to a variety of chronic diseases, causing each other to form a vicious circle, which seriously damages people's health, reduces the quality of life, and brings a heavy economic burden of disease to families and society.

MS is a process of accumulation of chronic damage to body functions, and is the result of decompensation of the body in response to various negative factors. It is a preventable, controllable, reversible and treatable disease. Relevant studies have found that cognitive behavioral therapy, exercise, lifestyle management, self-management education, and coping style guidance can alleviate patients' anxiety, depression and other negative emotions, master chronic disease knowledge and improve self-efficacy, thereby improving patients' blood sugar, blood lipid levels (13–15). A meta-analysis of 8 randomized controlled trials showed that lifestyle interventions can effectively slow the progression of MS disease, improve waist circumference (Waist Circumference, WC), blood pressure, fasting plasma glucose (Fasting Plasma Glucose, FPG) and triglycerides (Triglyceride, TG) (16).

The previous cross section explored the long-term incidence of MS and studied its influencing factors among the bereaved population 12 years after the Wenchuan earthquake. The research based on influencing factors and constructed a intervention program, which is a multi-ethnic culturally adaptable health self-management. This research provides a reliable basis and methodological reference for the systematic development of the health management of bereaved population in China, as well as established relevant guidelines and consensus especially for the MS patients in multi-ethnic gathering areas after disasters. The incidence of MS among the bereaved in the Wenchuan earthquake was 18.5% in cross section. However, there is no relevant research on the dynamic change of MS incidence in bereaved groups after disasters, so the incidence of MS in bereaved groups in Wenchuan earthquake cannot be compared with that in the same population under the same environment. Some scholars adopted the multistage and multilayer cluster sampling method in Sichuan and took CDS as the diagnostic standard. Up to October 2007, a total of 3,511 people were surveyed, and the prevalence of MS was 11.5% (male 14.3%, female 9.4%). The survey time was the closest to the prevalence of MS in Sichuan before the Wenchuan earthquake in 2008. It is suggested that the incidence rate in this study is much higher than that in the same area before the earthquake.

2. Methods

2.1. Design

A randomized controlled trial (RCT) design was adopted.

2.2. Research objects

2.2.1. Target groups

A cross-sectional study in Yingxiu Town and Beichuan County, the severely stricken area of the Wenchuan earthquake, enrolled confirmed MS patients who were bereaved by the earthquake.

2.2.2. Sampling standards

Inclusion criteria: ① MS patients diagnosed in a cross-sectional study; ② 18 years old and above; ③ permanent population (residential time more than 6 months per year); ④ ability to communicate normally and have basic literacy skills (*researchers are able to read Chinese characters and understand their meaning accurately*); and ⑤ provide informed consent and participate voluntarily.

Exclusion criteria: ① pregnant women; ② hearing impairment; ③ inability to communicate normally; ④ having cognitive impairment or mental illness.

2.3. Participants

In the first stage, 8 rural communities with similar economic conditions were selected as the target communities for sampling to facilitate the sampling. In order to avoid contamination, 6 communities in Yingxiu Town (*It was 11 kilometers from the epicenter*); 2 communities in Beichuan County (*It is about 132 km from the epicenter, but suffered great losses*) were the control communities.

In the second stage, paper strips with the names of the 6 village committees in Yingxiu Town were placed into an opaque container, and then they were taken twice in order. The two communities of the intervention group selected. There were 118 MS patients in the two villages.

In the third stage, the intervention group and the control group were selected from the 4 research communities selected in the first and second stages. The random sampling method was performed using Stata/MP14.0 software to generate a random number generator according to the ratio of experimental group: control group = 1:1 for randomized sample extraction.

Blind method: data collectors, data entry and statistical analysis personnel were blinded.

2.4. Intervention content and methods

Patients in the control group received routine chronic disease management in village clinics/township health centers. After baseline data collection, patients who were interested in participating in the study after the last data collection were informed that they could get a copy of the "Handbook" and corresponding health consultations for free. On the basis of routine management in the control group, patients in the intervention group received MS health self-management education intervention, mainly including the following:

1) Group teaching: a total of 8 times, once a week, for 8 consecutive weeks. The specific content and arrangement are shown in Table 1. The patients and family members were required to participate together to preview the relevant parts of the self-study manual before listening to the class. Teaching method: group multimedia lectures, 10–15 people in each group, 3 ~ 5 people discussed and shared after the lecture, and individual tutoring was conducted and questions answered according to the needs of the intervention subjects.
2) Log records: MS health self-management logbooks were distributed to guide patients to record daily so that patients could reflect on self-management, and it was also convenient for researchers to find problems and provide timely feedback.
3) WeChat online help/telephone follow-up: WeChat online help was used to solve the problems of the intervening subjects online at any time during the course and within 4 weeks after the end of the course. At the same time, the intervention subjects received 4 weeks of telephone/face-to-face follow-up once a week at the end of the course. The telephone/face-to-face follow-up was carried out according to the follow-up record sheet. The content mainly included assessing the patient's self-management practice at various stages, checking the completion of expected management goals, giving patients self-management support and encouragement, and helping patients solve the practice process difficulties and problems encountered.

2.5. Intervention effect evaluation indicators and evaluation tools

2.5.1. Evaluation time

In this study, the intervention evaluation time was set at baseline (T_0), at the end of the intervention (8 weeks of teaching plus 4 weeks of follow-up, T_1), and 2 months after the end of the intervention (T_2) to evaluate the effect of the intervention and to verify the metabolic syndrome at different follow-up times.

2.5.2. Evaluation tools and methods

1. General information survey form.

① Demographic data: gender, age, marital status, education level, religious beliefs, family monthly income *per capita*, medical payment method, occupation, etc.

② Earthquake traumatic experience: self-buried/injured/disability in the earthquake, family property loss, witnessing the burial/injury/death of others in earthquake, loss of specific relatives, number of bereavements, and remarriage/rebirth.

③ Disease-related information: family genetic history, etc.

2. Self-management Behavior Scale for Patients with MS.

10.3389/fpubh.2023.1277389

TABLE 1 Health self-management intervention program for patients with metabolic syndrome.

Subjects	Purpose	Contents	Methods
1. Project start 2. Transfer of MS patients diagnosed for the first time	1. Understand the purpose and significance of the project. 2. Establish a partnership with the target of intervention. 3. Help patients with MS diagnosed for the first time transfer to doctors in village clinics or township health centers.	1. Distribute the "Manual on the Health Self-management of MS Patients in the Wenchuan Earthquake Bereavement" (hereinafter referred to as the "Manual"). 2. Have intervention objects introduce themselves and become familiar with each other. 3. Introduce project. 4. Provide feedback of the baseline assessment results of various metabolic indices of MS patients. 5. Help patients with MS diagnosed for the first time transfer to doctors in village clinics or township health centers. 6. Establish a WeChat group to answer questions online and provide instant help.	1. The researcher introduces himself. 2. Intervention objects introduce themselves and meet each other. 3. Project introduction, emphasizing the significance and importance of the project. 4. Distribute the "Manual" to explain the usage and requirements for pre-class preparation and after-class review. 5. Feedback the baseline assessment results of various metabolic indices of MS patients. 6. Help patients with MS diagnosed for the first time transfer to doctors in village clinics or township health centers. 7. Establish a WeChat group, inform the purpose of the group, and let participants know that if they have any questions, they can ask online at any time.
Lesson 1 Overview of MS disease-related knowledge and health self-management	1. Master MS disease related knowledge, MS risk factors and prevention. 2. Recognize the importance of MS health self-management, six management strategies (stress/emotion, diet, daily life, exercise/weight, drug management and self-monitoring). 3. Develop self-management goals for MS health.	1. Feedback on qualitative interviews and baseline assessment of patients' knowledge of MS prevention and treatment. 2. Study the first part to the third part of the "Handbook." 3. Explain the importance of MS health self-management. 4. Formulate health self-management goals. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Feedback on the qualitative interview of MS prevention knowledge mastery and baseline "MS Prevention Knowledge Scale" evaluation results. 2. PPT explains the first to third parts of the "Handbook" (the definition of metabolic syndrome, diagnostic criteria and hazards; MS risk factors; MS prevention and treatment). 3. Explain the importance of health self-management for MS patients. 4. Jointly develop health self-management goals with the intervention targets. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.
Lesson 2 Stress/emotion management 1 (Daily life)	1. Understand the source, classification and stress response of daily life stress. 2. Recognize the importance of stress management. 3. Master the daily stress management strategy. 4. Master common pressure relaxation techniques.	1. Review the essentials of MS disease knowledge. 2. Feedback cross-sectional research on the psychological influencing factors of MS, the influencing factors of emotional management in qualitative interviews and the misunderstandings, and the baseline assessment results of emotional management in "Health Self-Management Behavior of MS Patients." 3. Study the content of daily life stress management in Part 4, Section 5, "Stress/Emotion Management" of the "Handbook." 4. Demonstrate common training techniques for stress relaxation. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Review the essentials of MS disease knowledge. 2. Provide feedback on cross-sectional research on the psychological influencing factors of MS disease, qualitative interview emotional management influencing factors and misunderstandings; and feedback on the baseline evaluation results of emotional management in "Healthy Self-Management Behavior of MS Patients." 3. PPT explains the "Manual" Part 4, Section 5 "Stress/Emotion Management" Daily Life Stress/Emotion Management: ① Correctly understand stress and stress response: the source of daily stress (life, environment, society, etc.) and stress response (physiology, psychology, behavior); acceptance of stress/emotion exists objectively ② Correctly manage stress and stress response: the importance of stress/emotion management to health; adjustment of bad cognition; establish a positive and healthy outlook on life and values; vent bad emotions; actively cope with stress ③ Actively seek social support and use social resources: seek relatives, friends, psychologists, etc. to express their pressure/emotions; use social resources to express emotions and obtain information, financial support, etc. ④ Common relaxation skills training: breathing training, meditation exercises and progressive muscle relaxation training 4. Audio/video demonstration of stress relaxation training (breathing training, meditation exercises and progressive muscle relaxation training) methods. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.

(Continued)

10.3389/fpubh.2023.1277389

TABLE 1 (Continued)

Subjects	Purpose	Contents	Methods
Lesson 3 Stress/emotion management 2 (Bereavement crowd)	1. Recognize the impact of earthquake stress on the health of bereaved families. 2. Understand grief and grief reactions. 3. Master stress/emotion management strategies (grief coping, family management and posttraumatic growth strategies). 4. Develop stress self-management plans and goals.	1. Review the essentials of daily stress/emotion management. 2. Study the content of stress/emotion management for bereavement groups in Part 4, Section 5, "Stress/Emotion Management" of the "Handbook." 3. Develop stress management plans and goals. 4. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Review the essentials of daily stress/emotion management. 2. PPT explains the "Manual" Part IV, Section 5 "Stress/Emotion Management" Pressure/Emotion Management of Bereavement Groups: ① Relevant knowledge: bereavement grief and grief reaction; the impact of earthquake stress on the health of bereaved people. ② Stress/emotion management strategy □Sorrow coping strategies: accept facts, think positively (combining Qiang and Tibetan religious beliefs); seek emotional companionship, etc. □Family business management: reorganization/remarriage family business strategy; family relationship improvement strategy; effective family support and utilization strategy; list of tasks for intervention targets □Growth management: Cognitive reconstruction, optimistic face, accomplish things that can bring a sense of accomplishment to oneself, learn to appreciate life, increase interpersonal communication, increase new possibilities (interests, job opportunities, etc. combined with local development) and enhance personal strength (strong, independent, brave) 3. Group discussion: stress/emotion management challenges and problems. 4. Experience sharing: stress/emotion management experience. 5. Develop stress/emotion management plans and goals with the intervention target. 6. Intervention objects evaluate and feedback on the effect of this lecture.
Lesson 4 diet manage	1. Master the purpose, methods and content of diet management. 2. Develop a diet management plan.	1. Review the essentials of stress/emotion management. 2. Feedback qualitative research on some of the influencing factors and misunderstandings of diet self-management, the baseline assessment results of dietary management in "Health Self-Management Behavior of MS Patients." 3. Study "Diet Management" in Part 4, Section 1 of "Handbook." 4. Formulate diet management plans and goals. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Review the essentials of stress/emotion management. 2. Feedback on the qualitative interviews on some of the influencing factors and misunderstandings of diet management, and the baseline assessment results of diet management in "Health Self-Management Behavior of MS Patients." 3. PPT explains the contents of Section 1 of Part Four of the "Handbook" (Diet Management Principles, Three Major Nutrients, Diabetes and Hypertension Dietary Points, Tips for Dietary Cooking). 4. Group discussion: problems and solutions in self-management of diet. 5. Experience sharing: experience in self-management of diet. 6. Develop diet self-management plans and goals with the intervention targets. 7. Evaluation and feedback on the effect of this lecture by the intervention objects.

(Continued)

TABLE 1 (Continued)

Subjects	Purpose	Contents	Methods
Lesson 5 1. Daily life management 2. Exercise/weight management	1. Master the purpose, methods, content and precautions of daily life, exercise/weight management. 2. Develop daily life and exercise management plans.	1. Review the main points of diet management. 2. Provide feedback on cross-sectional research on the impact of daily life (smoking, drinking, sleep quality) and physical exercise on MS, qualitative research on the weak links and misunderstandings of daily life management; the baseline of daily life and exercise management in “Health Self-Management Behavior of MS Patients” evaluation result. 3. Study “Manual” Part 4, Section 2 “Daily Life Management” and Section 3 “Exercise/Weight Management.” 4. Develop self-management plans and goals for smoking, drinking, sleep and exercise. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Review the main points of diet management. 2. Provide feedback on cross-sectional research on the impact of daily life (smoking, drinking, sleep quality) and physical exercise on MS, share the results of qualitative research results in patients with misunderstandings about drinking and physical exercise. 3. PPT explains the contents of Section 2 and Section 3 of the fourth part of the “Handbook”: daily life management and sports management: □ The hazards of smoking (including snuff), the relationship between smoking and metabolic syndrome, the benefits of quitting, tips for quitting, how to deal with difficulties in quitting □ The harm of drinking, the relationship between drinking and metabolic syndrome, the misunderstanding of drinking to lower blood sugar, the contraindications of drinking and drugs, and the coping strategies of drinking; tips for improving sleep quality □ Importance and significance of exercise/weight management, calculation of normal weight range, exercise intensity evaluation method, exercise method, exercise precautions, prevention and coping strategies of exercise hypoglycaemia, exercise misunderstanding 4. Group discussion: problems and solutions in daily life management and sports management. 5. Experience sharing: daily life and exercise self-management experience. 6. Work with the intervention subjects to develop self-management plans and goals for sports and daily life. 7. Intervention objects evaluate and feedback on the effect of this lecture.
Lesson 6 Drug management	1. Recognize the importance of drug management for MS control. 2. Master the names, usage, effects and precautions of commonly used antihypertensive drugs, lipid-lowering drugs, and hypoglycaemic drugs. 3. Master the method of insulin injection. 4. Develop a drug management plan.	1. Review the essentials of daily life management and exercise management. 2. Provide feedback on the evaluation results of baseline drug management of intervention subjects, weaknesses and misunderstandings of qualitative research drug management. 3. Study “Drug Management” in Part IV, Section 4 of the “Handbook.” 4. Formulate drug management plans and goals. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Review the essentials of daily life management and exercise/weight management of MS patients. 2. Provide feedback on the results of baseline drug management and qualitative research on weak links and misunderstandings in drug management. 3. PPT explains the “Drug Management” in Part IV, Section 4 of the “Handbook”: □ Importance of drug management □ The names, usage, effects and precautions of commonly used antihypertensive drugs, lipid-lowering drugs, and hypoglycaemic drugs □ Misunderstanding of drug management □ Medication management strategy □ Demonstrate insulin injection method 4. Group discussion: problems and solutions in drug management. 5. Experience sharing: drug management experience. 6. Jointly formulate drug management plans and goals with the intervention targets. 7. Evaluation and feedback on the effect of this lecture by the intervention objects.

(Continued)

10.3389/fpubh.2023.1277389

TABLE 1 (Continued)

Subjects	Purpose	Contents	Methods
Lesson 7 Self-monitoring (1) (theory)	1. Recognize the importance of metabolic indicators and monitoring of complications. 2. Understand the time, significance and control objectives of blood glucose, blood pressure, weight and waist measurement.	1. Review the main points of medication management. 2. Provide feedback on the self-monitoring baseline assessment results and qualitative research on the weaknesses and misunderstandings of self-monitoring. 3. Study the self-monitoring in Section 6 of Part IV of the "Handbook."	1. Review the main points of medication management. 2. Provide feedback on self-monitoring baseline assessment results, qualitative research on the weaknesses and misunderstandings of self-monitoring. 3. PPT explains the content of "Self-Monitoring" in Part 4, Section 6 of the "Handbook": ◇ □ Importance of monitoring of MS metabolic indicators ◇ □ MS control target value of each metabolic index ◇ □ Monitoring time point and significance of blood sugar, blood pressure, etc. ◇ □ The difference between blood glucose monitoring and HbA1c monitoring ◇ □ Blood glucose, blood pressure, waist circumference and weight monitoring frequency and precautions Importance, content and specific monitoring items of MS complications monitoring
Lesson 8 Self-monitoring (2) (Measurement methods)	1. Master the measurement methods of each component of MS. 2. Master the use of health management log records.	1. Review the relevant knowledge of MS self-monitoring. 2. Demonstrate the measurement method of each component of MS. 3. Explain how to use the health self-management log record sheet. 4. Develop self-monitoring plans and goals. 5. Evaluation and feedback on the effect of this lecture by the intervention objects.	1. Review the essentials of MS self-monitoring knowledge. 2. Demonstrate how to measure blood sugar, blood pressure, weight and waist circumference. 3. Explain how to use the health self-management log record sheet. 4. Group discussion: self-monitoring problems and solutions. 5. Experience sharing: self-monitoring experience. 6. Jointly develop self-monitoring plans and goals with the intervention targets. 7. Evaluation and feedback on the effect of this lecture by the intervention objects.
Intervention follow-up 1 month	1. Check the management plan and the completion of management objectives. 2. Support and encourage intervention targets to implement management strategies. 3. Help the intervening object to solve the difficulties and problems in the implementation of self-management.	1. Check the implementation of the self-management plan and the achievement of management objectives. 2. Help the intervening object to solve the difficulties and problems encountered in the implementation of self-management. 3. Encourage and support patients to implement management strategies.	According to the follow-up record sheet, follow up once a week by phone or face-to-face; combined with online communication and help of WeChat group.

10.3389/fpubh.2023.1277389

The scale is based on the chronic disease self-management theory (17). Identification of metabolic syndrome using phenotypes consisting of triglyceride levels with anthropometric indices in Korean adults [J]. In 2019, Dr. Ni Zhihong from West China Hospital of Sichuan University developed a "Metabolic Synthesis" based on the compilation of the chronic disease education and research center of Stanford University and other scholars (18). Self-management Behavior Scale, which can comprehensively evaluate the MS self-management behavior of community residents, including 7 dimensions (diet management, exercise/weight management, daily life style management, medication management, emotional management, self-monitoring), 36 items, using Likert 5-level scoring method, the option scores are 1 = never, 2 = occasionally, 3 = sometimes, 4 = frequently, 5 = always, the total score is 180 points, the more the score is High indicates that the patient's self-management behavior is better. The Cronbach's α coefficient of the scale is 0.868, and the test–retest reliability is 0.957.

2.5.3. Pre-test

To determine the performance of the MS patients' health self-management research tool and the feasibility of the program, the researchers used Stata. The MP14.0 software random number generator randomly selected 30 MS patients and randomly assigned them to the control group (n = 15) and the self-management intervention group (n = 15). The control group received chronic disease management in conventional village clinics/township hospitals, and the self-management education group received MS health self-management interventions and conventional village clinic chronic disease management. Patient-related data were collected before and after the intervention. The results show that the overall feasibility of the research program is good, the intervention process (8 weeks of teaching plus 4 weeks of follow-up) and the process of data collection are culturally adaptable and easy to communicate, and the questionnaire expression is clear and easy to understand. In addition, after collecting baseline data, 15 patients in the control group were re-evaluated by the MS Patient Self-Management Behavior Scale at 2 week intervals. The test–retest reliability, Cronbach's coefficient and test–retest reliability of the scale met the statistical requirements.

2.5.4. Data collection

1. Data collectors: 2 local medical undergraduates in fourth grade were recruited and unified training was conducted (1 student was responsible for patient physiological index measurement, and the other student was responsible for filling the patient questionnaire). Measurement methods and precautions were explained to ensure the accuracy of data collection.
2. Data collection location: ① questionnaire data were collected at home; ② basic disease data (anthropometric and physiological data) were measured on an empty stomach in each village clinic/township health center.
3. How to fill in the questionnaire: the respondents should have filled in the questionnaire by themselves. To ensure the accuracy and completeness of the information, when the questionnaire was collected, it was checked on the spot for any missing or questionable items. The respondents were asked to fill in and verify the questionnaire and then take it back after verification.
4. Collection method of disease-related physiological indicators: same as the current situation investigation part.
5. Data collection time: a formal intervention trial was conducted from March to July 2021, and relevant data of patients at baseline (T_0), end of intervention (T_1) and 2 months after the end of intervention (T_2) were collected.

2.5.5. Statistical analysis

Statistical description: Counting data are described by frequency and composition ratio; normal distribution of measurement data is described by mean ± standard deviation, and skew distribution is described by median and interquartile range.

Statistical inference: SPSS 22.0 software was used to perform statistical analysis on the data. According to the level of $\alpha = 0.05$, the p value is a two-sided probability for statistical inference. ① Comparison of enumeration data between groups: chi-square test or Fisher's exact probability method was used. ② Comparison of measurement data between groups: comparison of means of normal distribution: if the variances were uniform, the group t test was performed, and if the variances were not uniform, the t' test or Wilcoxon rank sum test was performed. Wilcoxon rank sum test was performed for comparison of nonnormal distribution means. ③ Intragroup time effect and between-group effect comparison of each index: repeated measurement data analysis of variance or generalized mixed benefit model analysis was performed.

2.6. Ethical issues

This study strictly follows the biomedical ethics code (No: 965). The research plan was sent to the Ethics Committee of West China Hospital of Sichuan University for approval.

3. Results

3.1. Baseline situation of study subjects

3.1.1. Information collection

A total of 330 MS patients who were bereaved in the Wenchuan earthquake were established based on the diagnostic criteria of the "Guidelines for the Prevention and Treatment of Type 2 Diabetes in China (2017)" (19). All study subjects completed the T_0 baseline data collection; T_1 collected a total of 127 valid data points, with a loss to follow-up rate of 3.78%; T_2 collected a total of 124 valid data points, with a loss to follow-up rate of 2.27%; the total loss to follow-up rate was 6.06%.

3.1.2. General information of research objects

After normality testing, the ages of the control group and the intervention group did not follow a normal distribution (Kolmogorov–Smirnov test, $p < 0.05$), and the 25th percentile, median, and 75th percentile were used to describe the two age distribution characteristics of the groups. In addition, the demographic data and the remaining variables of the earthquake trauma experience between the two groups were count data, which were described by frequency and composition ratio. According to the data characteristics, the statistical method used Pearson's chi-square test or Fisher's exact test. There was no significant difference between the basic demographic data and earthquake trauma experience data of the two groups of patients ($p > 0.05$) see Table 2.

The average age of the 132 survey subjects included in this section was 51.85 ± 12.19 years old; there were more women than men, accounting for 56.1%. The education level was mainly elementary school and below, accounting for 67.4%. The proportion of married people (including remarried) was 63.6% (12.9%). The occupation was mainly agriculture, accounting for 79.5%. The ethnic group was mainly Han nationality, accounting for 66.7%. The living status was mainly not living alone, accounting for 93.2%. The medical payment was mainly based on the new rural cooperative medical system, with a proportion of 65.9%. The family monthly income *per capita* was mainly 1,000–3,000 yuan/month/person, accounting for 64.4%. The specific demographic data of the intervention group and the control group are shown in Table 2.

In the earthquake trauma experience of the study subjects, 34.1% were buried themselves during the earthquake, 35.6% were injured themselves, 3.8% were disabled, 38.6% experienced severe or above property losses caused by the earthquake, and 85.3% were injured during the earthquake. The survey respondents witnessed others being buried, 84.1% witnessed others being injured, and 71.2% witnessed the death of others. The specific trauma experience data of the intervention group and the control group are shown in Table 3.

3.2. The impact of MS self-management on patients' health self-management behavior and disease-related metabolic indicators

The total scores of health self-management behaviors and disease-related metabolic indicators at T_0, T_1, and T_2 of the two groups of patients are shown in Tables 4 and 5. The total scores of patients' health self-management behaviors at three time points and disease-related metabolic indicators were analyzed for variance analysis and sphere test (Mauchly method) of repeated measurement data. The results of the sphere test showed $p > 0.05$, indicating that the data met the conditions of the sphere test. One-way analysis of variance was performed.

The results of the analysis of variance of repeated measurement data show the following:

1. intervention methods had statistically significant differences in the scores of self-management behavior and diet management, exercise management, emotional management, medication management and other lifestyle management ($p < 0.05$) (see Table 4).

Different intervention methods had statistical significance in body weight and WC ($p < 0.05$), but no statistical significance in blood pressure, FPG, and blood lipids ($p > 0.05$). The impact of community chronic disease management on the weight and WC of MS patients bereaved in the Wenchuan earthquake had an inter-group effect.

2. Time effect: The differences in self-management behavior and scores of various dimensions at different follow-up time points were statistically significant ($p < 0.05$), showing a significant time effect (see Table 4). The differences in body weight, WC, BMI, and SBP at different follow-up time points were statistically significant ($p < 0.05$), showing a significant time effect (see Table 5).
3. Interaction effect between intervention and time: There was an interaction effect between different intervention methods and different follow-up times ($p < 0.05$). Through the interactive contour map (Figures 1 and 2), it can be seen that with the passage of follow-up time, the two groups of patients' self-management behaviors and the trend of changes in the scores of each dimension are different (see Table 4, Figure 1).

Through the interactive contour map (see Figure 2), it can be seen that with the passage of follow-up time, the two groups of patients have different trends in weight, WC, BMI, blood pressure, FPG, and blood lipids.

4. Differences between groups: *t* tests were performed on the self-management behavior and scores of each dimension of the two groups of patients at each follow-up point. The results showed that at T0, the difference in scores between the two groups was not statistically significant ($p > 0.05$). The scores of the other dimensions at T1 and T2 were higher than those of the control group, and the differences were statistically significant ($p < 0.05$), see Table 4.

T-test was performed on the weight, WC, BMI, blood pressure, FPG and blood lipids of the two groups at each follow-up time point. The results showed: ① At T_0, T_1, and T_2 three time points, there was a difference in BMI between the two groups. There was no statistical significance ($p > 0.05$); ②At T_0, there was no significant difference in body weight and WC between the two groups ($p > 0.05$). ③There was no statistically significant difference in blood pressure and blood lipids between the two groups at the three time points ($p > 0.05$). ④At T_0 and T_1, there was no statistically significant difference in FPG between the two groups of patients ($p > 0.05$); at T_2, the difference in FPG between the two groups was statistically significant ($p < 0.05$) (see Table 5).

4. Discussion

The results of this study showed that the total score of self-management behavior in the control group at T_2 did not change much from the baseline, and the total score increased by an average of 1.18 points from the baseline, while the intervention group increased the self-management behavior at T_1 and T_2 (Figures 1, 2).

The results of this study are basically consistent with the relevant literature reports retrieved at home and abroad, and both indicate that patients can effectively improve their health self-management behaviors when they receive health self-management interventions. Such as Wang Yasha (20), Vanessa et al. (21), Marks et al. (22), Shin et al. (23) and Zhai Yingfen et al. (24).

The possible reasons for the expected results of this study are summarized into the following aspects. ① The study is designed on the

10.3389/fpubh.2023.1277389

TABLE 2 Baseline characteristics of general data of research subjects ($n = 132$).

Index	Number (%)		χ^2/U	p
	Intervention group (n = 66)	Control group (n = 66)		
Age (years old)	51.2	52.5	Z = 1.100	0.271#
Median (*P25,P75*)	(40.8，64.2)	(45.8,60.2)		
Gender				
Male	30 (45.5)	28 (42.4)	1.939	0.191*
Female	36 (54.5)	38 (57.6)		
Education				
Elementary school and below	42 (63.6)	47 (70.3)	1.011	0.799*
junior high school	18 (27.3)	14 (21.1)		
High school / technical secondary school and above	6 (9.1)	5 (7.7)		
Marital status				
Unmarried	3 (4.5)	1 (1.5)	5.125	0.401**
Married	53 (80.4)	53 (80.4)		
Divorced	2 (3.0)	1 (1.5)		
Widowed	8 (12.1)	11 (16.6)		
Occupation				
Non-farmers	17 (25.8)	10 (15.2)	2.281	0.195*
Farmer	49 (74.2)	56 (84.8)		
Ethnic				
Han	40 (60.6)	48 (72.7)	2.505	0.286**
Qiang	22 (33.3)	14 (21.2)		
Tibetan	4 (6.1)	4 (6.1)		
Residence status				
Live alone	6 (9.1)	3 (4.5)	1.073	0.492*
Not living alone	60 (90.9)	63 (95.5)		
Medical expenses types				
Rural residents' medical insurance	27 (40.9)	18 (27.3)	2.731	0.098*
Rural cooperative medical system	39 (59.1)	48 (72.7)		
Household monthly income *per capita*				
No fixed income	4 (6.1)	13 (19.7)	6.277	0.099**
≤ 1,000	6 (9.1)	5 (7.6)		
1,001 ~ 3,000	44 (66.7)	41 (62.1)		
3,001 ~ 5,000	12 (18.1)	7 (10.6)		

Remark:*, Pearson Chi-square test；**， Fisher's Exact test.

basis of three theoretical frameworks combined with evidence-based guidelines, previous cross-sectional studies and qualitative interview results, and fully excavated MS self-management intervention methods are highly targeted. ② The design of the intervention plan of this study is based on the teaching of MS disease-related knowledge, starting with the patient's cause, earthquake trauma/stress management, and sharing successful examples of communication with patients and empowering them. At the same time, it combines follow-up and strengthened interventions to health beliefs, produce a conscious and proactive attitude, and promote healthy behavior. ③ The researchers fully contacted and understanded the patients' eating habits and cultural customs from current situation investigations and qualitative interviews. Solving the difficulties, needs and challenges encountered in the implementation of patients' self-management behaviors is very important for improving patients' health self-management behaviors, which is consistent with the research reports of Aktas and other scholars (25). ④ During the teaching intervention period, due to the impact of the epidemic, the hospital's treatment

TABLE 3 Study subject's earthquake trauma experience ($n = 132$).

Index	Category	Number (%)		χ^2	p
		Intervention group (n = 66)	Control group (n = 66)		
I was buried in the earthquake	Yes	26 (39.4)	19 (28.8)	1.652	0.199*
	No	40 (60.6)	47 (71.2)		
I was injured in the earthquake	Yes	28 (42.4)	19 (28.8)	2.676	0.102*
	No	38 (57.6)	47 (71.2)		
I was disabled during the earthquake	Yes	3 (4.5)	2 (3.0)	0.208	0.648*
	No	63 (95.5)	64 (97.0)		
Family members died or whereabouts are unknown in the earthquake	1 person	24 (36.4)	30 (45.5)	2.090	0.352**
	2 person	29 (43.9)	21 (31.8)		
	3 people and above	13 (19.7)	15 (22.7)		
Property loss	none	2 (3.0)	4 (6.1)	6.999	0.136**
	Mild	5 (7.6)	12 (18.2)		
	Moderate	28 (42.4)	30 (45.5)		
	severe	20 (30.3)	10 (15.1)		
	Extremely severe	11 (16.7)	10 (15.1)		
Witnessed others being buried by the earthquake	Yes	35 (53.0)	42 (63.6)	1.527	0.217*
	No	31 (47.0)	24 (36.4)		
Witnessing others injured by the earthquake	Yes	52 (78.8)	59 (89.4)	2.775	0.096*
	No	14 (21.2)	7 (10.6)		
Witnessed the death of others by the earthquake	Yes	49 (74.2)	45 (68.2)	0.591	0.442*
	No	17 (25.8)	21 (31.8)		

Remark:*，Pearson Chi-square test；**，Fisher's Exact test.

TABLE 4 Comparison of the scores of healthy self-management behaviors between the two groups ($\bar{\chi} \pm s$，$n = 132$).

Index	groups	T_0	T_1	T_2	$F_{inter\text{-}group}$ (p)	F_{time} (p)	$F_{Interactive}$ (p)
Total score	Control	85.05 ± 4.06	85.61 ± 3.91	86.23 ± 3.74			
	intervention	84.14 ± 5.07	90.20 ± 4.99	93.74 ± 4.64			
	t	−1.136	5.874	6.605	23.912	121.617	80.945
	p	0.258	<0.001	<0.001	(<0.001)	(<0.001)	(<0.001)
Diet managem-ent	Control	31.74 ± 3.06	32.39 ± 2.91	32.35 ± 2.89			
	intervention	31.27 ± 3.02	33.83 ± 3.07	34.52 ± 2.95			
	t	−0.887	2.757	4.261	42.979	4.418	20.912
	p	0.376	0.007	<0.001	(<0.001)	(0.039)	(<0.001)
Exercise managem-ent	Control	7.30 ± 1.43	8.14 ± 1.10	8.52 ± 1.14			
	intervention	7.39 ± 1.33	7.61 ± 1.18	7.73 ± 1.00			
	t	−0.377	2.652	4.219	5.213	2.652	4.219
	p	0.707	0.009	<0.001	(0.026)	(0.009)	(<0.001)
Emotional managem-ent	Control	13.24 ± 1.71	13.14 ± 1.41	13.12 ± 1.53			
	intervention	13.02 ± 1.84	13.74 ± 1.94	14.26 ± 1.26			
	t	−0.734	2.052	4.637	4.155	15.109	27.944
	P	0.464	0.042	<0.001	(0.046)	(<0.001)	(<0.001)

(Continued)

TABLE 4 (Continued)

Index	groups	T_0	T_1	T_2	$F_{inter-group}$ (p)	F_{time} (p)	$F_{Interactive}$ (p)
Drug managem-ent	Control	8.00 ± 1.41	8.08 ± 1.31	7.64 ± 1.43			
	intervention	7.92 ± 1.37	8.67 ± 1.50	9.09 ± 1.46			
	t	−0.312	2.404	5.769	9.319	7.047	23.193
	p	0.755	0.018	<0.001	(0.003)	(0.002)	(<0.001)
Disease self-monit-oring	Control	9.67 ± 1.63	9.68 ± 1.64	9.85 ± 1.38			
	intervention	9.65 ± 1.65	9.88 ± 1.61	10.39 ± 1.35			
	t	−0.053	0.480	2.286	0.669	15.738	5.535
	p	0.958	0.632	0.024	(0.416)	(<0.001)	(0.006)
Other lifestyle	Control	8.91 ± 1.37	8.83 ± 1.26	9.20 ± 1.14			
	intervention	8.86 ± 1.58	9.71 ± 1.44	10.09 ± 1.59			
	t	−0.176	3.726	3.704	7.259	25.630	15.554
	p	0.861	<0.001	<0.001	(0.009)	(<0.001)	(<0.001)
Communication with physicians	Control	6.09 ± 1.21	5.88 ± 1.17	6.35 ± 0.86			
	intervention	6.11 ± 1.12	6.29 ± 1.17	6.88 ± 1.10			
	t	0.074	2.005	3.197	3.481	23.762	6.885
	p	0.941	0.047	0.002	(0.067)	(<0.001)	(0.002)

Remarks: interaction, intervention × time.

TABLE 5 Comparison of metabolic indicators between the two groups ($\bar{\chi} \pm s$, $n = 132$).

Index	groups	T_0	T_1	T_2	$F_{inter-group}$ (p)	F_{time} (p)	$F_{Interactive}$ (p)
Weight	Control	73.06 ± 10.39	73.42 ± 10.11	73.74 ± 10.05			
	intervention	71.02 ± 9.45	69.50 ± 9.53	66.76 ± 7.96			
	t	−1.182	−2.294	−4.425	11.433	35.395	79.499
	p	0.239	0.023	<0.001	(0.001)	(<0.001)	(<0.001)
WC	Control	92.48 ± 8.64	92.24 ± 8.37	92.52 ± 8.05			
	intervention	91.67 ± 6.59	89.58 ± 5.76	88.62 ± 5.51			
	t	−0.611	−2.131	−3.240	4.096	28.988	33.742
	p	0.542	0.035	0.002	(0.047)	(<0.001)	(<0.001)
BMI	Control	25.91 ± 2.68	26.11 ± 2.62	26.18 ± 2.68			
	intervention	26.39 ± 2.21	25.80 ± 2.17	24.79 ± 2.06			
	t	1.130	−0.722	3.341	1.041	35.505	71.261
	p	0.260	0.472	<0.001	(0.311)	(<0.001)	(<0.001)
SBP	Control	138.03 ± 17.45	136.26 ± 17.76	137.47 ± 17.03			
	intervention	141.85 ± 20.88	138.76 ± 20.44	137.05 ± 17.35			
	t	1.140	0.750	−0.142	0.403	9.926	4.841
	p	0.257	0.455	0.888	(0.528)	(<0.001)	(0.009)
DBP	Control	83.52 ± 10.00	83.30 ± 9.65	84.48 ± 8.76			
	intervention	84.95 ± 12.05	84.48 ± 10.65	83.94 ± 9.94			
	t	0.746	0.668	−0.334	0.185	1.340	8.408
	p	0.457	0.505	0.739	(0.668)	(0.269)	(0.001)
FPG	Control	6.24 ± 1.78	6.27 ± 1.43	6.31 ± 1.36			
	intervention	6.12 ± 1.46	5.94 ± 1.17	5.83 ± 0.87			
	t	−0.427	−1.428	2.400	1.557	1.125	4.106
	p	0.670	0.156	0.018	(0.217)	(0.328)	(0.019)

(Continued)

TABLE 5 (Continued)

Index	groups	T_0	T_1	T_2	$F_{inter-group}$ (p)	F_{time} (p)	$F_{Interactive}$ (p)
TC	Control	4.30 ± 0.79	4.33 ± 0.79	4.38 ± 0.75			
	intervention	4.49 ± 0.93	4.38 ± 0.89	4.25 ± 0.81			
	t	1.223	0.341	−0.958	0.067	1.823	8.086
	p	0.223	0.733	0.340	(0.797)	(0.170)	(0.001)
TG	Control	2.26 ± 1.17	2.25 ± 1.08	2.30 ± 0.97			
	intervention	2.21 ± 0.95	2.14 ± 0.91	2.08 ± 0.86			
	t	−0.261	−0.608	−1.361	0.456	1.614	3.571
	p	0.795	0.544	0.176	(0.502)	(0.203)	(0.031)
LDL-C	Control	2.92 ± 0.75	2.92 ± 0.76	2.95 ± 0.76			
	intervention	2.93 ± 0.68	2.89 ± 0.61	2.85 ± 0.60			
	t	0.111	−0.280	−0.774	0.146	0.269	0.875
	p	0.912	0.780	0.440	0.704)	(0.765)	(0.419)
HDL-C	Control	1.27 ± 0.27	1.26 ± 0.32	1.26 ± 0.30			
	intervention	1.30 ± 0.30	1.33 ± 0.30	1.36 ± 0.30			
	t	0.537	1.129	1.837	1.737	2.076	3.530
	p	0.592	0.261	0.068	(0.192)	(0.134)	(0.035)

Remarks: interaction, intervention × time.

process was accompanied by QR code scanning, body temperature measuring, nucleic acid testing, and patients and companions filling out the epidemiological questionnaire.

The results of this study showed that the weight of the subjects in the control group did not change much from the baseline at T_2, and their weight increased by 0.68Kg on average from the baseline; while the subjects in the intervention group were significantly lower than the baseline at T_1 and T_2. The average weight loss was 4.26Kg, which was 6.0% less than the baseline weight. According to the evidence-based guidelines recommendation standard (19).

There are similarities and differences in the four results of blood lipids (TC, TG, HDL-C and LDL-C) reported by scholars Shengnan (26), Senhai et al. (27) and Yasha (20).

Analysis of the four blood lipids (TC, TG, HDL-C and LDL-C) between the two groups of patients in this study were not statistically significant at the two time points of T_1 and T_2. The possible reasons for the failure to achieve the expected results are: ① During the entire period of the intervention, the study was in the epidemic stage. It was forbidden to gather more 10 people, and collective entertainment activities (playing mahjong, dancing square dance) were canceled, which may be more idle than in previous years. ② The intervention of this study started in March. The climate became warmer and darker. Patients usually exercise moderately after eating. The work and rest habits of these two rural residents may be difficult to change within 5 months.

4.1. Struggles and limitations

Studies on the long-term impact of earthquake stress on the health of bereaved families are rarely reported worldwide. Due to some objective conditions in this study, the metabolic outcome indicators of this intervention study did not include HbA1c. The FPG results at the three time points of T_0, T_1, and T_2 are used as indicators to judge whether the patient's blood glucose has improved, rather than the average blood glucose HbA1c of 2 to 3 months, which may be biased.

This study selected three detection time points: baseline (T_0), end of intervention (T_1), and 2 months after the end of intervention (T_2) on an evidence-based basis. The long-term effects of health self-management interventions need to be further explored. Small numbers examined and the consequent difficult generalisability of results.

5. Conclusion

The intervention program of healthy self-management for MS patients from bereaved families in the Wenchuan earthquake can effectively improve patients' health self-management behaviors and patients' weight, WC, BMI, and FPG in the short term. However, the effects of improving the patient's blood lipid and blood pressure levels are uncertain and need further verification.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethics Committee of West China Hospital of Sichuan University for approval(No:965). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal

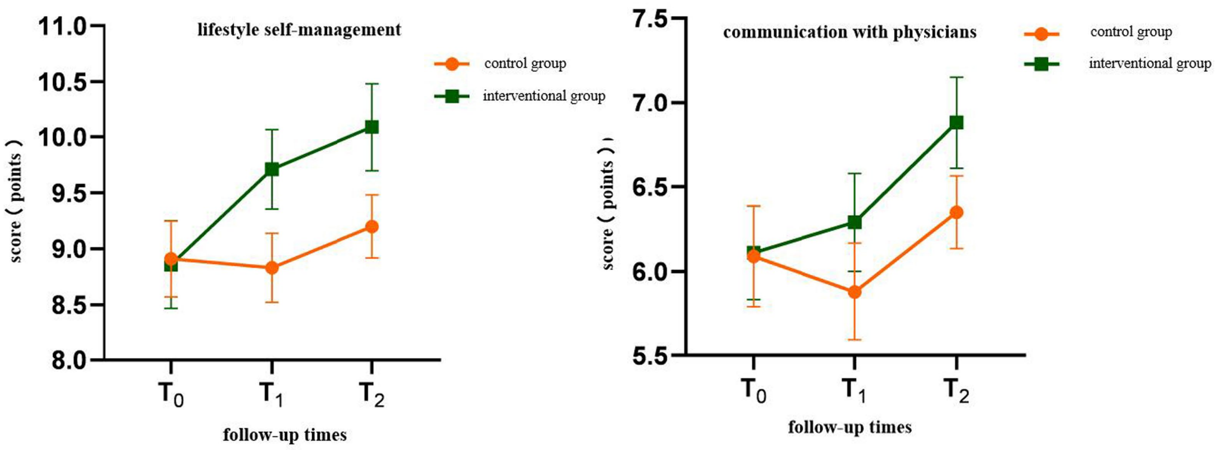


FIGURE 1
Interactive contour map of health self-management behavior score Interactive contour map of health self-management behavior score.

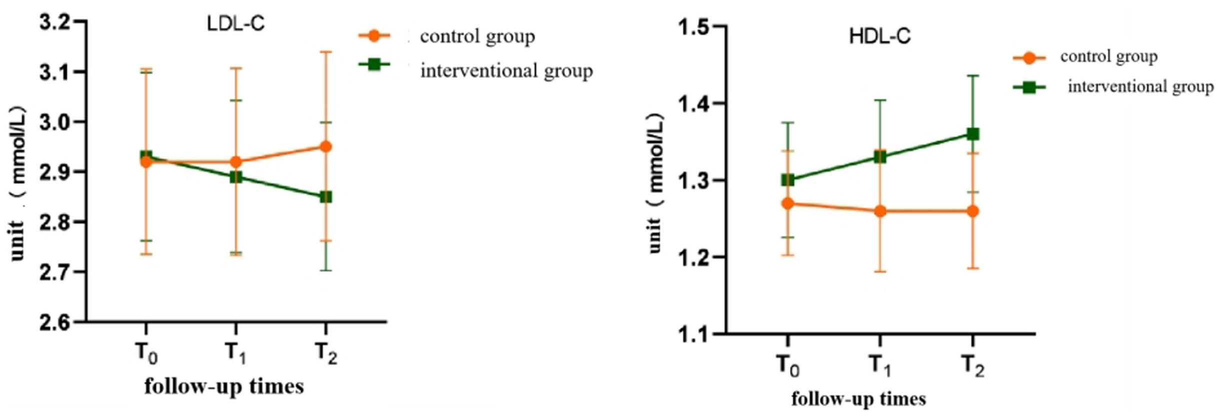


FIGURE 2
Interactive contour map of Metabolic Index.

guardians/next of kin. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

Author contributions

ML: Conceptualization, Writing – original draft. JX: Software, Writing – review & editing. WS: Supervision, Writing – original draft. JN: Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research granted from the Natural Science Foundation of Gansu Province (No. 21JR1RA106), National Natural Science Foundation of China (No.72264022) and the First Hospital of Lanzhou University doctoral research start-up fund (ldyyn2022-97).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Carmassi C, Dell'Oste V, Bertelloni CA. Disrupted rhythmicity and vegetative functions relate to PTSD and gender in earthquake survivors. *Front Psychiatry*. (2020) 16:492006. doi: 10.3389/fpsyt.2020.492006

2. Rafiey H, Alipour F, LeBeau R. Prevalence and determinants of PTSD 3 years after an earthquake in Iran. *Community Ment Health J*. (2019) 55:542–7. doi: 10.1007/s10597-019-00384-x

3. State Council Information Office of the People's Republic of China, (2010), *Wenchuan earthquake*, Available at: http://www.scio.gov.cn/zhzc/6/2/Document/1003461/1003461.htm, (Accessed: October 14, 2009).

4. Xinman D. *A study on the bereavement, anxiety, depression and its influencing factors of the bereaved one year after the Wenchuan earthquake*. China: Sichuan University (2010).

5. Qianjin J. *Medical psychology*. Beijing: People's Health Publishing House (2000).

6. Yongxiang M. *Study on the effect of health self-management intervention for hypertension patients in Banfang community*. China: Sichuan University (2010).

7. Zhang G, Cui P, Jin W. Changes in hydrological behaviours triggered by earthquake disturbance in a mountainous watershed. *Sci Total Environ*. (2021) 18:1–6. doi: 10.1016/j.scitotenv.2020.143349

8. Athyros VG, Bouloukos VI, Pehlivanidis AN, Papageorgiou AA, Dionysopoulou SG, Symeonidis AN, et al. The prevalence of the metabolic syndrome in Greece: the MetS-Greece multicentre study. *Diabetes Obes Metab*. (2005) 7:397–405. doi: 10.1111/j.1463-1326.2004.00409.x

9. Kohsari M, Moradinazar M, Rahimi Z. Liver enzymes and their association with some Cardiometabolic diseases: evidence from a large Kurdish cohort. *Biomed Res Int*. (2021) 2021:1–8. doi: 10.1155/2021/5584452

10. Nehus E. Obesity and chronic kidney disease. *Curr Opin Pediatr*. (2018) 30:241–6. doi: 10.1097/MOP.0000000000000586

11. Xie K, Bao L, Jiang X. The association of metabolic syndrome components and chronic kidney disease in patients with hypertension. *Lipids Health Dis*. (2019) 18:229. doi: 10.1186/s12944-019-1121-5

12. Gathirua-Mwangi WG, Monahan PO, Murage MJ, Zhang J. Metabolic syndrome and total cancer mortality in the third national health and nutrition examination survey. *Cancer Causes Control*. (2017) 28:127–36. doi: 10.1007/s10552-016-0843-1

13. Cooney LG, Milman LW, Hantsoo L, Kornfield S, Sammel MD, Allison KC, et al. Cognitive-behavioral therapy improves weight loss and quality of life in women with polycystic ovary syndrome: a pilot randomized clinical trial. *Fertil Steril*. (2018) 110:161–171.e1. doi: 10.1016/j.fertnstert.2018.03.028

14. Shomaker LB, Kelly NR, Radin RM, Cassidy OL, Shank LM, Brady SM, et al. Prevention of insulin resistance in adolescents at risk for type 2 diabetes with depressive symptoms: 1-year follow-up of a randomized trial. *Depress Anxiety*. (2017) 34:866–76. doi: 10.1002/da.22617

15. Williams J, Stubbs B, Gaughran F, Craig T. 'Walk this Way' – a pilot of a health coaching intervention to reduce sedentary behaviour and increase low intensity exercise in people with serious mental illness: study protocol for a randomised controlled trial. *Trials*. (2016) 17:594. doi: 10.1186/s13063-016-1660-2

16. Shan Q, Shengli . A diagnostic test of the generalized anxiety scale for generalized anxiety disorder screening in the outpatient department of psychology in general hospitals. *Chin J Ment Health*. (2015) 29:939–44.

17. Lee BJ, Kim JY. Identification of metabolic syndrome using phenotypes consisting of triglyceride levels with anthropometric indices in Korean adults. *BMC Endocr Disord*. (2020) 20:29. doi: 10.1186/s12902-020-0510-0

18. Zhihong N. *A study on health self-management intervention in Xinjiang Kazakh patients with metabolic syndrome*. China: Sichuan University (2019).

19. Chinese Experts'. Consensus on Diagnosis and Treatment of Cardiovascular Disease with Insomnia (2017).

20. Yasha W, Xiao X, Tao F. Study on the effect of comprehensive intervention in patients with metabolic syndrome. *Chin J Dis Control*. (2017) 21:784–8.

21. Bullón-Vela V, Abete I, Tur JA, Pintó X, Corbella E, Martínez-González MA. Influence of lifestyle factors and staple foods from the Mediterranean diet on non-alcoholic fatty liver disease among older individuals with metabolic syndrome features. *Nutrition*. (2020) 71:110620–8. doi: 10.1016/j.nut.2019.110620

22. Marks S. Culturally sensitive education can decrease Hispanic workers' risk of metabolic syndrome[J]. *Workplace Health Saf*. (2016) 64:543–9. doi: 10.1177/2165079916634712

23. Shin N-M, Choi J, Cho I, Park BJ. Self-management program for heart healthy behavior among middle and old-aged Korean women at risk for metabolic syndrome. *J Cardiovasc Nurs*. (2017) 32:E8–E16. doi: 10.1097/JCN.0000000000000406

24. Zhai Yingfen X, Hongyan C. The effect of comprehensive nursing intervention on medication compliance, sleep quality and quality of life in patients with metabolic syndrome. *China Med Herald*. (2015) 12:142–5.

25. Aktas MF, Mahler A, Hamn M, Perger G, Simon F, Westenhöfer J. Lifestyle interventions in Muslim patients with metabolic syndrome-a feasibility study. *Eur J Clin Nutr*. (2018) 73:18–26. doi: 10.1038/s41430-018-0371-z

26. Shengnan M. *Community metabolic syndrome intervention research based on chronic disease management model*. China: Zhejiang University (2017).

27. Senhai Y, Feixia P, Hangjie G. Evaluation of the effect of intensive lifestyle intervention in rural metabolic syndrome patients. *Prev Med*. (2017) 29:1193–8.

frontiers | Frontiers in Public Health

TYPE Systematic Review
PUBLISHED 09 November 2023
DOI 10.3389/fpubh.2023.1279718

Check for updates

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Hun Lee,
University of Ulsan, Republic of Korea
Amir Faisal,
Sumatra Institute of Technology, Indonesia

*CORRESPONDENCE
Xiaobo Xia
✉ xbxia21@163.com
Jian Jiang
✉ jiangjianxy@126.com

†These authors have contributed equally to this work

RECEIVED 18 August 2023
ACCEPTED 18 October 2023
PUBLISHED 09 November 2023

CITATION
Zhou Y, Dai M, Sun L, Tang X, Zhou L, Tang Z, Jiang J and Xia X (2023) The accuracy of intraocular lens power calculation formulas based on artificial intelligence in highly myopic eyes: a systematic review and network meta-analysis. *Front. Public Health* 11:1279718. doi: 10.3389/fpubh.2023.1279718

The accuracy of intraocular lens power calculation formulas based on artificial intelligence in highly myopic eyes: a systematic review and network meta-analysis

Yi Zhou[1,2,3†], Minhui Dai[1,2,3†], Lingyu Sun[1,2,3], Xiangyi Tang[1,2,3], Ling Zhou[1,2,3], Zhiyao Tang[4], Jian Jiang[1,2,3]* and Xiaobo Xia[1,2,3]*

[1]Eye Center of Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]Hunan Key Laboratory of Ophthalmology, Central South University, Changsha, Hunan, China, [3]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China, [4]Xiangya School of Nursing, Central South University, Changsha, Hunan, China

Objective: To systematically compare and rank the accuracy of AI-based intraocular lens (IOL) power calculation formulas and traditional IOL formulas in highly myopic eyes.

Methods: We screened PubMed, Web of Science, Embase, and Cochrane Library databases for studies published from inception to April 2023. The following outcome data were collected: mean absolute error (MAE), percentage of eyes with a refractive prediction error (PE) within ±0.25, ±0.50, and ±1.00 diopters (D), and median absolute error (MedAE). The network meta-analysis was conducted by R 4.3.0 and STATA 17.0.

Results: Twelve studies involving 2,430 adult myopic eyes (with axial lengths >26.0 mm) that underwent uncomplicated cataract surgery with mono-focal IOL implantation were included. The network meta-analysis of 21 formulas showed that the top three AI-based formulas, as per the surface under the cumulative ranking curve (SUCRA) values, were XGBoost, Hill-RBF, and Kane. The three formulas had the lowest MedAE and were more accurate than traditional vergence formulas, such as SRK/T, Holladay 1, Holladay 2, Haigis, and Hoffer Q regarding MAE, percentage of eyes with PE within ±0.25, ±0.50, and ±1.00 D.

Conclusions: The top AI-based formulas for calculating IOL power in highly myopic eyes were XGBoost, Hill-RBF, and Kane. They were significantly more accurate than traditional vergence formulas and ranked better than formulas with Wang–Koch AL modifications or newer generations of formulas such as Barrett and Olsen.

Systematic review registration: https://www.crd.york.ac.uk/PROSPERO/, identifier: CRD42022335969.

KEYWORDS

intraocular lens, formulas, high myopia, artificial intelligence, prediction error

1. Introduction

Myopia is a common refractive error that affects a significant proportion of the world population. The global prevalence of myopia was estimated to be 22.9% in 2000 and is projected to increase to 49.8% by 2050 (1). High myopia, defined as a refractive error of −6.00 diopters (D), is associated with axial lengths >26.0 mm (2). Myopia has been shown to be a risk factor for the development of cataracts, particularly nuclear cataracts and posterior subcapsular cataracts (3). Traditionally, vergence formulas such as SRK/T, Holladay 1 and 2, and Hoffer Q have been commonly used (4). However, these traditional formulas tend to result in hyperopic surprises, leading surgeons to empirically aim for a myopic target (5–7). Highly myopic eyes have complex structural changes such as zonular weakness (8), increases in anterior chamber depth (ACD) (9), premature vitreous degeneration, and posterior scleral staphyloma (6), which reduces the predictive accuracies of existing formulas. The Wang—Koch (WK) axial length (AL) adjustment (6, 10) and newer generations of formulas such as Barrett (available at: https://calc.apacrs.org/barrett_universal2105/) and Olsen (11) were developed to address these issues.

In recent years, artificial intelligence (AI) technology has been adopted to improve the accuracy and precision of IOL power calculations in myopic eyes. The Hill-radial basis function (RBF) formula (available at: http://rbfcalculator.com/online/index.html) and Kane formula (available at: www.iolformula.com) are gaining increasing popularity. Both formulas were developed and validated using large datasets. They used machine learning algorithms based on several patient-specific factors, including AL, keratometry (K), and lens thickness (LT). Other AI-based formulas, such as Emmetropia Verifying Optical (EVO) and Ladas Super Formula, have also been developed (11, 12).

Recent studies have compared the accuracy of AI-based formulas, traditional vergence formulas, newer generations of vergence formulas, and formulas with Wang–Koch adjustments. However, due to the large number of formula types, the process of recalculating IOL power using all the methods was time-consuming, and few studies have performed comprehensive comparisons between formulas. Our network meta-analysis, therefore, aims to comprehensively compare and rank the formulas in myopic patients who underwent cataract surgery. The findings of the present study will provide valuable clinical guidance for selecting the appropriate IOL formulas for myopic eyes.

2. Materials and methods

The present study was registered at Prospero (CRD42022335969, https://www.crd.york.ac.uk/PROSPERO/).

2.1. Search strategy and selection criteria

Two investigators (YZ and LS) searched PubMed, Web of Science, Embase, and Cochrane Library for studies published from their inception to 5 April 2023. The search terms used for searching the clinical condition are as follows: "myopia," "long axial length," "long AL," "long eye," "intraocular lens," and "IOL." The two investigators independently evaluated the title and abstract of all the identified studies. Additionally, we manually examined the reference lists of clinical trials, related meta-analyses, and systematic reviews to identify relevant studies.

Studies were retained if they met the following inclusion criteria: (1) focused on individuals with ocular AL longer than 26.0 mm; (2) included eyes with uncomplicated cataract surgery with in-the-bag fixated mono-focal IOL implantation; and (3) used at least two of the selected IOL power calculation formulas. Articles were excluded if they (1) used no AI formula; (2) included patients under 18 years; (3) had a history of other ocular diseases, eye surgery, or trauma; (4) included toric, multi-focal, piggyback, or not in-the-bag fixated IOL implantation; (5) included astigmatism correction using femtosecond laser-assisted cataract surgery; (6) did not provide any of the outcome data (MAE ± SD, percentage of eyes with a refractive PE within ±0.25, ±0.50, and ±1.00 D, MedAE); (7) measured optical biometry using approaches other than Lenstar, IOL Master, or Pentacam; and (8) were review articles or discussion papers, conference abstracts, or studies done on animals.

2.2. Data collection and processing

Two authors (MD and XT) extracted the following outcome data independently: (1) The percentage of eyes with a refractive prediction error (PE) within ±0.50 and ±1.00 diopters (D), (2) mean absolute error (MAE), and (3) median absolute error (MedAE) in refractive prediction. Participant and intervention characteristics were also extracted. For data that were missing or could not be directly obtained, we contacted the authors or used the WebPlotDigitizer tool (https://automeris.io/WebPlotDigitizer/) to read data from figures.

The percentage of eyes with PE within ±0.25, ±0.50, and ±1.00 D was dichotomous data. Thus, a binomial model was applied, and odds ratio (OR) with 95% CIs was calculated for the relative effect. The MAE was continuous data. Thus, a continuous model was applied, and a mean difference with 95% CIs was calculated for the relative effect. It is notable that MedAE was not suitable for the meta-analysis; therefore, only descriptive analyses were performed.

2.3. Quality assessment

Two authors (LZ and ZT) assessed the risk of bias in the included studies following the guidance of the quality appraisal tool for case series studies using a modified Delphi technique developed by the Institute of Health Economics (13). The following eight domains in the included studies were evaluated: study objective, study population, intervention and co-intervention, outcome measure, statistical analysis, results and conclusions, competing interests and sources of support, and new item. The clarity of each item in the eight domains was classified as "Yes," "No," and "Unclear/Partly stated."

2.4. Publication bias detection

To assess the publication bias across studies, a graphic tool was developed by Chai (14). The code was integrated into an R package *netmeta*. The command *funnel()* generated a funnel plot to visualize publication bias across included studies. The obvious publication was presented as an asymmetric distribution of comparison-adjusted funnel plots.

2.5. Sensitivity analysis and inconsistency assessment

A sensitivity analysis was performed by repeating the network meta-analysis with the previously excluded high-risk studies. If the result was significantly influenced, the inconsistency between direct and indirect comparisons was assessed using the node-splitting approach (15), which differentiates direct and indirect evidence on a particular comparison and the design-by-treatment interaction models, assuming consistency throughout the entire network. A p-value < 0.05 was considered an inconsistency.

2.6. Surface under the cumulative ranking curve (SUCRA)

The probability of interventions at each ranking could be evaluated by SUCRA (16). The SUCRA value of each formula was assessed for the following primary outcomes: the percentage of eyes with a refractive PE within ±0.50 and ±1.00 D, MAE ± SD, and MedAE. A SUCRA value ranges from 0 to 100%, with a value closer to 100% indicating a higher likelihood that a formula is in the top rank. A SUCRA ranking figure was presented to report the SUCRA value for respective outcomes.

2.7. Subgroup analysis

To further compare the accuracy of AI-based formulas and conventional formulas, we performed subgroup analysis stratified by ALs (26.0–28.0, 28.0–30.0, and ≥30.0 mm) in studies where subgroup stratification with the same criteria was also conducted. Eyes with ALs >28.0 mm were defined as extremely myopic eyes. The MAE was compared in each subgroup using the evaluation metrics described above.

2.8. Statistical analysis

Network meta-analyses were performed using a random-effects model. All analyses were conducted using R 4.3.0 and STATA 17.0 for statistical analyses. The R packages *gemtc, ggplot2, netmeta, and ggrepel* were used for analysis, data output, and visualization.

3. Results

3.1. Study selection

The literature search strategy is presented in Supplementary Table S1. After removing duplications, 871 articles were identified from the literature search. Twenty-four full-text articles were further screened for eligibility. The preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram is shown in Supplementary Figure S1. The final inclusion of this systematic review consisted of 12 studies involving 2,430 adult myopic eyes that underwent uncomplicated cataract surgery with in-the-bag fixated mono-focal IOL implantation.

3.2. Study characteristics and network geometry

A summary of all eligible studies is shown in Supplementary Table S2 (17–28). The included AI formulas were Kane, Ladas super formula, Hill-RBF Version 2.0 and 3.0, XGBoost, K6, and Olsen. The included traditional formulas (based on vergence or ray-tracing) were Barrett Universal II, SRK/T, Holladay 1, Holladay 2, Hoffer Q, Haigis, Emmetropia Verifying Optical (EVO), and Olsen. If Wang–Koch (WK) adjustment was applied, the formula was analyzed as an independent formula. Table 1 shows the brief description and abbreviations for the formulas. The number of studies and eyes involved in each formula is shown in Supplementary Table S3.

The number of formulas involved in the studies ranged from 2 to 11. Of the included studies, 10 (83.3%) were from China, 1 (8.3%) recruited participants from countries in Europe, and 1 (8.3%) from Australia.

3.3. Risk of bias

The risk of bias from within the included articles is shown in Supplementary Table S4. All studies gained "Yes" in "study objective," "outcome measures," "statistic analysis," and "competing interests and sources of support." In the domain of "study population," all 12 studies obtained at least three "Yes" responses. In "results and conclusions," seven out of 12 studies gained over three "Yes" remarks. All the studies were retrospective designs. Overall, all studies gained at least 13 "Yes" responses among 20 items and were regarded as high quality.

Publication bias across studies was evaluated by funnel plot shown in Supplementary Figures S2–S5. The estimates of all the comparisons were symmetrically distributed in comparison-adjusted funnel plots, suggesting no publication bias across studies.

3.4. Mean absolute error in refractive prediction

Mixed comparisons for MAE between AI-based formulas and traditional formulas are presented in Figure 1A. XGBoost formula was superior to Hoffer Q, and Kane was superior to SRK/T in terms of MAE. Most AI-based formulas, except for Kane, showed lower errors in refractive prediction when compared to Holladay 1. Hill-RBF Version 2 and 3 and Kane formulas showed lower errors when compared to Haigis. All AI-based formulas did not outperform Barrett, which is representative of the newer generation of traditional formulas. However, when Wang–Koch adjustment was applied to SRK/T, Haigis, and Holladay 1 formulas, there was no significant difference between traditional and AI-based formulas.

TABLE 1 Brief description of the formulas included in the network meta-analysis.

Formula	Classification	Year	Brief description
Kane	AI-based	2016	Based on AL, K, ACD, LT (optional), CCT (optional), Gender, A constant, and post-operative refractive target. Blended approach (AI, regression, and vergence)
Ladas super formula	AI-based	2015	Applies most ideal calculations from other formulas (SRK/T, Hoffer Q, Holladay 1, Holladay, Haigis, etc.)
Hill-RBF			Based on AL, K, ACD, LT (optional), WTW (optional), CCT (optional), A-constant, and post-operative refractive target
Version 2.0	AI-based	2018	Excludes out-of-bounds values Might be significantly influenced by LT
Version 3.0	AI-based	2020	
All	AI-based	2018	Version 2.0 with out-of-bounds values
XGBoost	AI-based	2020	Based on the XGBoost machine learning regression technique. Incorporates several clinical features, and the BUII formula results Targets highly or extremely myopic eyes
K6	AI-based	2020	Transforms the optical biometer's AL to be the distance from the anterior cornea to the retinal pigment epithelium Uses a proprietary estimated lens position calculation based on post-operative measurement of 245 eyes
FullMonte IOL	AI-based		Uses a Monte Carlo Markov Chain simulator to produce its refractive predictions
Olsen	Traditional	2014	Ray-tracing Based on ACD, LT, and post-operative refractive target
Barrett Universal II	Traditional	2010	Based on AL, K, ACD, LT (optional), WTW (optional), LF/DF, A-constant, and post-operative refractive target The formula is not publicly available
SRK/T			
SRK/T	Traditional	1990	Based on AL, K, A-constant, and post-operative refractive target
SRK/T_WK	Traditional		SRK/T formula with WK adjustment
SRK/T_MWK	Traditional		SRK/T formula with modified WK adjustment
Holladay 1			
Holladay 1	Traditional	1988	Based on AL, K, SF, and post-operative refractive target
Holladay 1_WK	Traditional		Holladay 1 formula with WK adjustment
Holladay 1_MWK	Traditional		Holladay 1 formula with modified WK adjustment
Holladay 2	Traditional	1995	Based on AL, K, ACD, LT, WTW, CCT, Age, A-constant/ACD/SF, and post-operative refractive target
Hoffer Q	Traditional	1993	Based on AL, K, pACD, and post-operative refractive target
Haigis			
Haigis	Traditional	2004	Based on AL, K, ACD, three constants (a0; a1, which is associated with measured ACD; and a2, which is associated with measured AL), and post-operative refractive target
Haigis_WK	Traditional		Haigis formula with WK adjustment
Emmetropia Verifying Optical (EVO)	Traditional	2019	Based on AL, K, ACD, A-constant, LT (optional), CCT (optional), corneal refractive LVC status, and post-operative refractive target

3.5. Percentage of eyes with a refractive PE within ±0.25D, ±0.50, and ±1.00 D

Mixed comparisons for the percentage of eyes with a PE within ±0.25, ±0.50, and ±1.00 D are presented in Figure 1B. In terms of the percentage of eyes with a PE within ±0.25 D (% PE within ±0.25 D), Kane was superior to Haigis, Hoffer Q, Holladay 1, and Holladay 2. Hill-RBF was better than Haigis, Hoffer Q, Holladay 1, Holladay 2, and SRK/T. However, if out-of-bounds were not excluded when applying Hill-RBF, the formula did not outperform SRK/T or Haigis. Ladas super formula showed the same percentage of eyes with a PE within ±0.25 D as all the traditional formulas. The XGBoost method was superior to most traditional formulas except for Wang–Koch adjusted formulas of newer generations.

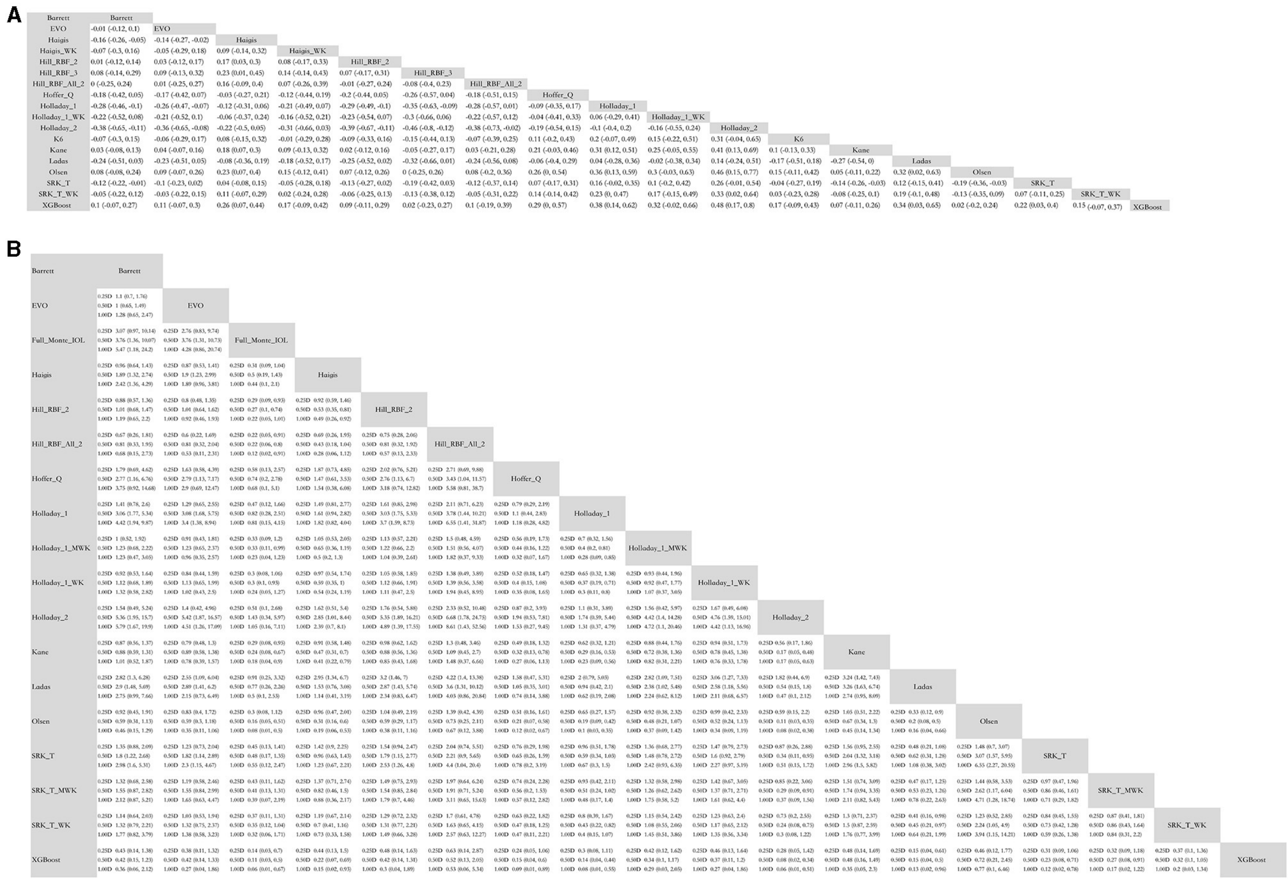


FIGURE 1
Mixed comparison of AI-based formulas and traditional formulas. **(A)** Mean absolute error. **(B)** Percentage of eyes with predictive error within ±0.25D, ±0.50 D, and ±1.00 D. WK, Wang–Koch AL adjustment; MWK, modified Wang–Koch AL adjustment; EVO, Emmetropia Verifying Optical; Barrett, Barrett Universal II; Ladas, Ladas super formula; RBF, radial basis function.

Similarly, in terms of percentage PE within ±0.50 D, Kane and XGBoost were better than Haigis, Hoffer Q, Holladay 1, Holladay 2, and SRK/T. Hill-RBF was better than Hoffer Q, Holladay 1, Holladay 2, and SRK/T. However, Ladas super formula and FullMonte IOL formula were comparable to traditional ones.

Regarding the percentage of eyes with PE within ±1.00 D, XGBoost was superior to Haigis, Hoffer Q, Holladay 1, Holladay 2, and SRK/T. Kane and Hill-RBF were better than Haigis, Holladay 1, Holladay 2, and SRK/T. Again, Ladas super formula and FullMonte IOL formula were comparable to traditional ones.

It is notable that, in the percentage of eyes with PE within ±0.25, ±0.50, and ±1.00 D, AI-based formulas were comparable to newer generations of traditional vergence formulas or Wang–Koch adjusted formulas.

3.6. Median absolute error (MedAE) in refractive prediction

Figure 2 and Supplementary Table S5 show the analysis and formula ranking results for MedAE, and there were 12 studies in which 21 formulas were involved. We found that the XGBoost, Hill-BRF, and Kane formulas had lower MedAE (0.2730, 0.2730, and 0.2730, respectively).

3.7. SUCRA ranking of all outcomes

The SUCRA values provided the probabilities of AI-based or traditional formulas at each ranking and are presented in Figure 3 and Supplementary Tables S6–S9. The probabilities of each formula being the best were also plotted.

For obtaining the minimal MAE, XGBoost, Hill-RBF Version 3.0, and Olsen ranked as the three best (Figure 3A). Holladay 2 ranked the worst. However, XGBoost did not show significant superiority to Hill-RBF Version 3.0 and Olsen [Hill_RBF_3 vs. XGBoost = 0.02 (−0.23, 0.27) vs. Olsen vs. XGBoost = 0.02 (−0.2, 0.24); Figure 1A]. The probabilities of XGBoost, Hill-RBF Version 3.0, and Olsen being the best were 0.30880, 0.26040, and 0.21345, respectively.

For the percentage of eyes with a PE within ±0.25 D, ±0.50 D, and ±1.00 D, XGBoost, Hill-RBF Version 2.0, and Kane were the best ranking AI-based formulas (Figure 3B). Similarly, there was no significant difference between each of the three formulas. Among

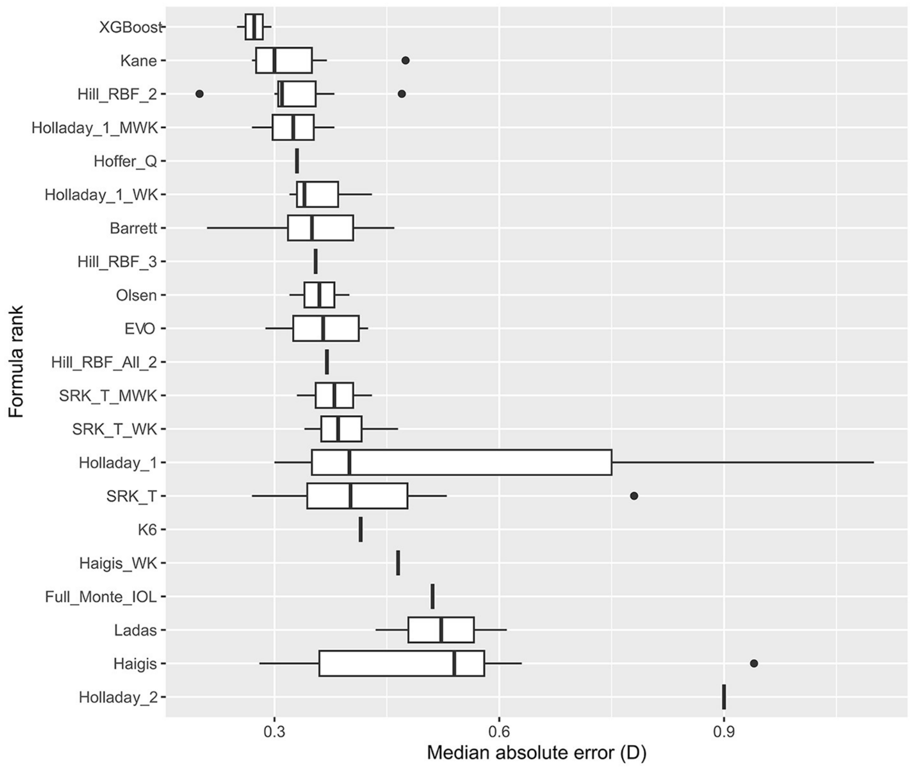


FIGURE 2
Formula rank in median absolute error (MedAE). WK, Wang–Koch AL adjustment; MWK, modified Wang–Koch AL adjustment; EVO, Emmetropia Verifying Optical; Barrett, Barrett Universal II; Ladas, Ladas super formula; RBF, radial basis function.

all the formulas, Ladas had the lowest probability in the percentage of eyes with PE within ±0.25 D and ±1.00 D (both 0.000125), while Holladay 2 had the lowest probability in percentage PE within ±0.50 D (0.000000).

3.8. Inconsistency analysis

To detect the inconsistency within networks, the node-splitting approach was applied. No significant consistency ($p > 0.05$) was observed in terms of MAE or percentage of eyes with PE within ±0.25 D, ±0.50 D, and ±1.00 D (Supplementary Tables S10–S13). Significant consistency ($p > 0.05$) was detected in the analyses above.

3.9. Subgroup analysis

Six of the studies underwent subgroup analysis using the previously described criteria (stratifying ALs into three subgroups: 26.0–28.0, 28.0–30.0, and ≥30.0 mm, Supplementary Tables S14–S16). Three studies involving 11 formulas and 381 eyes were included for subgroup analysis because they provided comprehensive MAE values, SD values, and sample sizes required for network meta-analysis using continuous data.

The mixed comparisons of the AI-based and conventional formulas for each subgroup are presented in Figure 4. In eyes with ALs between 26.0–28.0 and 28.0–30.0 mm, all formulas were comparable to each other. In extremely myopic eyes with an AL ≥30.0 mm, the XGBoost formula was significantly more accurate than Haigis [MAE decreased by 0.39 (0.05, 0.73)] and SRK/T [MAE decreased by 0.37 (0.01, 0.74)], and Hill-RBF 3.0 was significantly more accurate than Haigis [MAE decreased by 0.38 (0.02, 0.74)]. Other formulas were comparable to each other in the subgroup with ALs >30.0 mm.

The SUCRA values and the probabilities of each formula being the best are provided in Supplementary Table S17. In the 26.0–28.0 mm subgroup, XGBoost (SUCRA = 0.79861), Hill-RBF 2.0 (SUCRA = 0.756005), and Hill-RBF 3.0 (SUCRA = 0.717455) were the top three formulas. In the 28.0–30.0 mm subgroup, Hill-RBF 3.0 (SUCRA = 0.644415), XGBoost (SUCRA = 0.599635), and Kane (SUCRA = 0.583095) were the top three formulas. In the ≥30.0 mm subgroup, XGBoost (SUCRA = 0.88663), Hill-RBF 3.0 (SUCRA = 0.855355), and Hill-RBF 2.0 (SUCRA = 0.580605) were the top three formulas.

4. Discussion

The present study is the first network meta-analysis to evaluate the accuracy of AI-based formulas for IOL power calculation in myopic eyes with axial lengths >26.0 mm. To clearly discuss the characteristics of each formula, we divided the 21 formulas into the following four types: (1) AI-based formulas; (2) newer

FIGURE 3
Surface under the cumulative ranking curve ranking plot. SUCRA, surface under the cumulative ranking curve; PrBest, probabilities of being the best; PE, predictive error; D, diopter; WK, Wang–Koch AL adjustment; MWK, modified Wang–Koch AL adjustment; EVO, Emmetropia Verifying Optical; Barrett, Barrett Universal II; Ladas, Ladas super formula; RBF, radial basis function.

generation of traditional formulas (such as Barrett and Olsen); (3) traditional vergence formulas (such as SRK/T, Hoffer Q, Holladay 1, and Holladay 2); and (4) traditional vergence formulas with AL adjustment (such as SRK/T_WK or SRK/T_MWK). By analyzing MAE and the percentage of eyes with ±0.25 D, ±0.50 D, and ±1.00 D of prediction error, we demonstrated that XGBoost, Kane, and Hill-RBF were the most accurate AI-based formulas.

The XGBoost formula was designed exclusively for myopic eyes (19, 26). It was developed and validated using data from 1,450 patients. The subgroup analysis in our study showed that, in eyes with an AL ≥30.0 mm, the XGBoost formula was significantly more accurate than traditional ones (such as Haigis and SRK/T). This advantage might result from the study design where the average AL of the recruited patients was >29.00 mm, and extremely high myopic eyes were taken into consideration. Additionally, the XGBoost formula included cases where <-2.5 D myopic refractive targets were scheduled (19). It was, therefore, suggestive that the XGBoost formula might be more reliable for IOL power prediction in highly or extremely myopic eyes compared to other AI-based formulas. Most recently, the Zhu–Lu formula (https://HM-ZLF.com/), developed by the same team that developed using XGBoost and support vector regression (SVR) algorithms, demonstrated improved and stable accuracy compared to other formulas (29).

Hill-RBF was the first IOL power calculation method based purely on artificial intelligence and was installed on Lenstar (HaagStreit, Switzerland) (30). Hill-RBF 2.0 is based on more than 12,000 eyes and can calculate IOL power for a target different than zero. It was based on AL, K, ACD, and LT [CCT, LT, and CD are optional (31, 32)]. The Hill-RBF 2.0 was limited due to the "out-of-bounds" warnings. Hill-RBF 3.0 formula, an improvement from its 2.0 version, utilized pattern recognition and employed an advanced method of data interpolation. However, there has

Barrett	Barrett										
EVO	26.0 - 28.0 mm: -0.05 (-0.21, 0.1) 28.0 - 30.0 mm: -0.21 (-0.51, 0.16) ≥30.0 mm: -0.02 (-0.22, 0.22)	EVO									
Haigis	26.0 - 28.0 mm: 0 (-0.2, 0.22) 28.0 - 30.0 mm: -0.12 (-0.57, 0.33) ≥30.0 mm: -0.21 (-0.55, 0.12)	26.0 - 28.0 mm: 0.05 (-0.16, 0.28) 28.0 - 30.0 mm: 0.08 (-0.4, 0.52) ≥30.0 mm: -0.2 (-0.55, 0.13)	Haigis								
Haigis_WK	26.0 - 28.0 mm: -0.13 (-0.34, 0.1) 28.0 - 30.0 mm: -0.09 (-0.52, 0.38) ≥30.0 mm: 0.04 (-0.27, 0.34)	26.0 - 28.0 mm: -0.08 (-0.3, 0.16) 28.0 - 30.0 mm: 0.12 (-0.37, 0.57) ≥30.0 mm: 0.05 (-0.28, 0.35)	26.0 - 28.0 mm: -0.13 (-0.38, 0.13) 28.0 - 30.0 mm: 0.03 (-0.48, 0.56) ≥30.0 mm: 0.25 (-0.14, 0.63)	Haigis_WK							
Hill_RBF_2	26.0 - 28.0 mm: 0.04 (-0.14, 0.24) 28.0 - 30.0 mm: 0.08 (-0.31, 0.52) ≥30.0 mm: 0.03 (-0.26, 0.32)	26.0 - 28.0 mm: 0.1 (-0.09, 0.3) 28.0 - 30.0 mm: 0.3 (-0.14, 0.7) ≥30.0 mm: 0.05 (-0.25, 0.32)	26.0 - 28.0 mm: 0.04 (-0.22, 0.3) 28.0 - 30.0 mm: 0.21 (-0.34, 0.78) ≥30.0 mm: 0.24 (-0.15, 0.65)	26.0 - 28.0 mm: 0.17 (-0.1, 0.43) 28.0 - 30.0 mm: 0.18 (-0.38, 0.74) ≥30.0 mm: -0.01 (-0.38, 0.38)	Hill_RBF_2						
Hill_RBF_3	26.0 - 28.0 mm: 0.04 (-0.16, 0.25) 28.0 - 30.0 mm: 0.04 (-0.39, 0.49) ≥30.0 mm: 0.17 (-0.13, 0.45)	26.0 - 28.0 mm: 0.09 (-0.11, 0.31) 28.0 - 30.0 mm: 0.24 (-0.24, 0.69) ≥30.0 mm: 0.18 (-0.14, 0.46)	26.0 - 28.0 mm: 0.04 (-0.2, 0.27) 28.0 - 30.0 mm: 0.16 (-0.36, 0.68) **≥30.0 mm: 0.38 (0.02, 0.74)**	26.0 - 28.0 mm: 0.16 (-0.08, 0.42) 28.0 - 30.0 mm: 0.12 (-0.38, 0.65) ≥30.0 mm: 0.13 (-0.21, 0.47)	26.0 - 28.0 mm: -0.01 (-0.26, 0.25) 28.0 - 30.0 mm: -0.05 (-0.61, 0.5) ≥30.0 mm: 0.14 (-0.24, 0.49)	Hill_RBF_3					
K6	26.0 - 28.0 mm: -0.15 (-0.38, 0.09) 28.0 - 30.0 mm: -0.12 (-0.56, 0.34) ≥30.0 mm: 0.07 (-0.23, 0.36)	26.0 - 28.0 mm: -0.1 (-0.33, 0.15) 28.0 - 30.0 mm: 0.09 (-0.4, 0.54) ≥30.0 mm: 0.08 (-0.24, 0.37)	26.0 - 28.0 mm: -0.15 (-0.41, 0.11) 28.0 - 30.0 mm: 0.01 (-0.51, 0.54) ≥30.0 mm: 0.28 (-0.09, 0.65)	26.0 - 28.0 mm: -0.02 (-0.29, 0.25) 28.0 - 30.0 mm: -0.03 (-0.57, 0.5) ≥30.0 mm: 0.03 (-0.32, 0.39)	26.0 - 28.0 mm: -0.19 (-0.47, 0.08) 28.0 - 30.0 mm: -0.2 (-0.77, 0.34) ≥30.0 mm: 0.04 (-0.34, 0.4)	26.0 - 28.0 mm: -0.19 (-0.44, 0.07) 28.0 - 30.0 mm: -0.15 (-0.67, 0.36) ≥30.0 mm: -0.1 (-0.43, 0.23)	K6				
Kane	26.0 - 28.0 mm: -0.03 (-0.17, 0.14) 28.0 - 30.0 mm: 0 (-0.32, 0.34) ≥30.0 mm: -0.04 (-0.28, 0.16)	26.0 - 28.0 mm: 0.02 (-0.12, 0.21) 28.0 - 30.0 mm: 0.21 (-0.15, 0.53) ≥30.0 mm: -0.02 (-0.28, 0.17)	26.0 - 28.0 mm: -0.03 (-0.24, 0.2) 28.0 - 30.0 mm: 0.12 (-0.34, 0.59) ≥30.0 mm: 0.17 (-0.18, 0.49)	26.0 - 28.0 mm: 0.1 (-0.12, 0.34) 28.0 - 30.0 mm: 0.09 (-0.38, 0.56) ≥30.0 mm: -0.08 (-0.41, 0.23)	26.0 - 28.0 mm: -0.07 (-0.25, 0.13) 28.0 - 30.0 mm: -0.08 (-0.52, 0.32) ≥30.0 mm: -0.07 (-0.38, 0.2)	26.0 - 28.0 mm: -0.06 (-0.27, 0.16) 28.0 - 30.0 mm: -0.04 (-0.5, 0.42) ≥30.0 mm: -0.2 (-0.51, 0.07)	26.0 - 28.0 mm: 0.12 (-0.1, 0.37) 28.0 - 30.0 mm: 0.12 (-0.35, 0.59) ≥30.0 mm: -0.11 (-0.43, 0.18)	Kane			
SRK_T	26.0 - 28.0 mm: -0.01 (-0.22, 0.2) 28.0 - 30.0 mm: 0 (-0.43, 0.44) ≥30.0 mm: -0.19 (-0.55, 0.15)	26.0 - 28.0 mm: 0.04 (-0.17, 0.26) 28.0 - 30.0 mm: 0.21 (-0.26, 0.65) ≥30.0 mm: 0.08 (-0.24, 0.37)	26.0 - 28.0 mm: -0.02 (-0.25, 0.22) 28.0 - 30.0 mm: 0.12 (-0.38, 0.64) ≥30.0 mm: 0.02 (-0.41, 0.44)	26.0 - 28.0 mm: 0.11 (-0.13, 0.36) 28.0 - 30.0 mm: 0.09 (-0.43, 0.61) ≥30.0 mm: -0.23 (-0.63, 0.17)	26.0 - 28.0 mm: -0.06 (-0.32, 0.2) 28.0 - 30.0 mm: -0.09 (-0.64, 0.45) ≥30.0 mm: -0.22 (-0.65, 0.19)	26.0 - 28.0 mm: -0.05 (-0.29, 0.18) 28.0 - 30.0 mm: -0.03 (-0.54, 0.47) ≥30.0 mm: -0.36 (-0.74, 0.01)	26.0 - 28.0 mm: 0.13 (-0.12, 0.39) 28.0 - 30.0 mm: 0.12 (-0.39, 0.63) ≥30.0 mm: -0.26 (-0.66, 0.13)	26.0 - 28.0 mm: 0.01 (-0.21, 0.22) 28.0 - 30.0 mm: 0 (-0.45, 0.45) ≥30.0 mm: -0.15 (-0.5, 0.21)	SRK_T		
SRK_T_WK	26.0 - 28.0 mm: -0.04 (-0.24, 0.17) 28.0 - 30.0 mm: -0.07 (-0.5, 0.4) ≥30.0 mm: -0.06 (-0.37, 0.25)	26.0 - 28.0 mm: 0.01 (-0.19, 0.23) 28.0 - 30.0 mm: 0.14 (-0.35, 0.59) ≥30.0 mm: -0.04 (-0.38, 0.27)	26.0 - 28.0 mm: -0.04 (-0.28, 0.2) 28.0 - 30.0 mm: 0.06 (-0.46, 0.57) ≥30.0 mm: 0.15 (-0.23, 0.54)	26.0 - 28.0 mm: 0.09 (-0.16, 0.33) 28.0 - 30.0 mm: 0.02 (-0.49, 0.55) ≥30.0 mm: -0.09 (-0.45, 0.27)	26.0 - 28.0 mm: -0.09 (-0.34, 0.17) 28.0 - 30.0 mm: -0.16 (-0.72, 0.41) ≥30.0 mm: -0.08 (-0.47, 0.3)	26.0 - 28.0 mm: -0.08 (-0.31, 0.16) 28.0 - 30.0 mm: -0.1 (-0.62, 0.41) ≥30.0 mm: -0.22 (-0.56, 0.13)	26.0 - 28.0 mm: 0.11 (-0.14, 0.37) 28.0 - 30.0 mm: 0.05 (-0.46, 0.58) ≥30.0 mm: -0.12 (-0.48, 0.23)	26.0 - 28.0 mm: -0.02 (-0.24, 0.19) 28.0 - 30.0 mm: -0.07 (-0.53, 0.4) ≥30.0 mm: -0.02 (-0.32, 0.32)	26.0 - 28.0 mm: -0.03 (-0.27, 0.21) 28.0 - 30.0 mm: -0.07 (-0.58, 0.44) ≥30.0 mm: 0.14 (-0.26, 0.55)	SRK_T_WK	
XGBoost	26.0 - 28.0 mm: 0.06 (-0.1, 0.22) 28.0 - 30.0 mm: 0.01 (-0.32, 0.35) ≥30.0 mm: 0.18 (-0.04, 0.39)	26.0 - 28.0 mm: 0.11 (-0.07, 0.3) 28.0 - 30.0 mm: 0.22 (-0.22, 0.6) ≥30.0 mm: 0.19 (-0.08, 0.44)	26.0 - 28.0 mm: 0.06 (-0.17, 0.28) 28.0 - 30.0 mm: 0.13 (-0.35, 0.61) **≥30.0 mm: 0.39 (0.05, 0.73)**	26.0 - 28.0 mm: 0.19 (-0.05, 0.42) 28.0 - 30.0 mm: 0.1 (-0.39, 0.58) ≥30.0 mm: 0.14 (-0.17, 0.46)	26.0 - 28.0 mm: 0.01 (-0.22, 0.24) 28.0 - 30.0 mm: -0.08 (-0.59, 0.4) ≥30.0 mm: 0.15 (-0.18, 0.47)	26.0 - 28.0 mm: 0.02 (-0.21, 0.23) 28.0 - 30.0 mm: -0.03 (-0.51, 0.44) ≥30.0 mm: 0.01 (-0.29, 0.31)	26.0 - 28.0 mm: 0.21 (-0.03, 0.45) 28.0 - 30.0 mm: 0.13 (-0.36, 0.6) ≥30.0 mm: 0.11 (-0.2, 0.42)	26.0 - 28.0 mm: 0.08 (-0.12, 0.26) 28.0 - 30.0 mm: 0.01 (-0.4, 0.4) ≥30.0 mm: 0.22 (-0.02, 0.49)	26.0 - 28.0 mm: 0.07 (-0.15, 0.29) 28.0 - 30.0 mm: 0.01 (-0.47, 0.46) **≥30.0 mm: 0.37 (0.01, 0.74)**	26.0 - 28.0 mm: 0.1 (-0.13, 0.31) 28.0 - 30.0 mm: 0.08 (-0.41, 0.55) ≥30.0 mm: 0.23 (-0.09, 0.56)	XGBoost

FIGURE 4
Mixed comparison of subgroup analysis stratified by axial lengths.

been no study directly comparing Hill-RBF 3.0 and 2.0 in eyes with an AL >26.0 mm. Tsessler et al. (32) found that Hill-RBF 3.0 was more accurate than Hill-RBF 2.0, though their study did not necessarily focus on myopic eyes. Network meta-analysis offers an indirect approach to comparing two formulas without actually conducting the comparative trial. The SUCRA ranking of MAE in our study showed that the Hill-RBF 3.0 ranked higher than the 2.0 version. Moreover, the subgroup analysis further demonstrated the superiority of Hill-RBF 3.0 in eyes with an AL >30.0 mm. Although the MedAE ranking in Figure 2 shows Hill-RBF 2.0 with a higher ranking than the 3.0 version, it is worth mentioning that MedAE was not suitable for meta-analysis (33). Therefore, only descriptive analyses were presented in the MedAE ranking, and no statistical analysis could be undertaken. Hill-RBF formulas were suitable for highly myopic eyes, and Hill-RBF 3.0 was particularly accurate for extremely myopic eyes.

The Kane formula is an unpublished one, and the structure is largely unknown. It is based on theoretical optics, contains some elements of artificial intelligence, and uses AL, K, ACD, and gender to predict the IOL position, with LT and CCT being optional factors (34, 35). The formula considers factors such as ACD and LT, which are known to affect IOL power calculations in myopic eyes. Our findings suggest that the Kane formula was more accurate than conventional formulas such as Haigis and SRK/T in highly myopic eyes. However, as the subgroup analysis suggested, Kane was comparable to traditional formulas when dealing with extremely myopic eyes.

The Ladas super formula was created by Dr. John Ladas and further optimized in 2019 using the post-operative data of more than 4,000 eyes (35). It uses a three-dimensional model to choose the best IOL formula among existing ones for a particular AL or corneal power (12). Ang et al. (36) found that, in myopic eyes, Ladas was less accurate than AI-based or newer formulas such as Kane and Barrett. They also demonstrated a strong positive correlation between absolute prediction error and AL with the Ladas and SRK/T formulas, especially in extremely long ALs. Similarly, data in our study showed that the Ladas formula ranked 16th out of the 18 formulas in the MAE SUCRA ranking analysis and was not superior to traditional formulas in other evaluation analyses. The reason for its unexpected poor performance was partially due to the fact that Ladas was developed by combining conventional formulas such as Hoffer Q, Holladay 1, Holladay 2 (with Wang–Koch adjustment) (6), and SRK/T formulas (11) rather than creating new algorithms as most recent AI formulas did. When using Ladas formula in highly myopic eyes, other formulas should also be used to choose the most accurate one. However, physicians should be careful when using the Ladas formula in eyes with an AL >30.0 mm.

The newer generation of formulas showed superiority over traditional vergence formulas. Barrett formula incorporates AL, K, ACD, LT, WTW, age, corneal power, and estimated lens position (37–39). The Olsen formula is characterized by the ray-tracing technique and the C constant concept (40–43). In the present study, both Barrett and Olsen showed no significant difference from AI-based formulas. However, the Olsen formula had significantly lower MAE than SRK/T, Haigis, Hoffer Q, Holladay 1, and Holladay 2. Barrett formula had significantly lower MAE than most traditional formulas except for Hoffer Q. In terms of the percentage of eyes within ±0.50 and ±1.00 D PE, Barrett and Olsen formulas showed significant superiority over traditional formulas. The Wang–Koch adjustment was developed to be applied in eyes with longer ALs that have IOL power calculation with the Holladay 1 formula (6, 10). In the present study, there was no significant difference between formulas with Wang–Koch adjustment and AI-based formulas in terms of MAE and percentage of eyes within ±0.25 D, ±0.50 D, and ±1.00 D PE.

5. Limitations and recommendations

This network meta-analysis has several limitations inherent to the methodology applied in the study. First, 10 out of the 12 studies included in this study were conducted in China, and the other two were from Australia and Europe. Therefore, the conclusions of our study might not be generalized to other populations. Second, one study included both eyes of some patients, and the correlation between eyes is a potential limitation of the analysis. Third, two studies (27, 28) used the lens constants from the User Group for Laser Interference Biometry (ULIB), and more research is needed to analyze the effect of ULIB in the future.

6. Conclusion

In summary, the overall evidence indicated that in cataract patients with ALs >26.0 mm, AI-based formulas (especially XGBoost, Hill-RBF, and Kane) were promising in obtaining lower MAE and a higher percentage of eyes within ±0.25 D, ±0.50 D, and ±1.00 D of prediction error when compared to traditional vergence formulas. AI-based formulas tended to perform better than newer generations of formulas (such as Barrett and Olsen) and formulas with Wang–Koch adjustment, but the superiority was not significant. In future studies, sufficiently sized and geographically dispersed studies are warranted to validate the effect of AI-based formulas.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

YZ: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft. MD: Conceptualization, Data curation, Writing – original draft. LS: Writing – review & editing. XT: Software, Writing – review & editing. LZ: Software, Writing – review & editing. ZT: Software, Writing – review & editing. JJ: Conceptualization, Supervision, Writing – review & editing. XX: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was financially supported by the National Key Research and Development Program of China (No. 2020YFC2008205), the National Natural Science Foundation of China (Nos. 82171058 and 81974134), and the Key R&D Plan of Hunan Province of China (No. 2020SK2076).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1279718/full#supplementary-material

References

1. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology.* (2016) 123:1036–42. doi: 10.1016/j.ophtha.2016.01.006

2. Haigis W. Intraocular lens calculation in extreme myopia. *J Cataract Refract Surg.* (2009) 35:906–11. doi: 10.1016/j.jcrs.2008.12.035

3. Pan CW, Cheng CY, Saw SM, Wang JJ, Wong TY. Myopia and age-related cataract: a systematic review and meta-analysis. *Am J Ophthalmol.* (2013) 156:1021–33.e1. doi: 10.1016/j.ajo.2013.06.005

4. Voytsekhivskyy OV. Development and clinical accuracy of a new intraocular lens power formula (VRF) compared to other formulas. *Am J Ophthalmol.* (2018) 185:56–67. doi: 10.1016/j.ajo.2017.10.020

5. Hill DC, Sudhakar S, Hill CS, King TS, Scott IU, Ernst BB, et al. Intraoperative aberrometry versus preoperative biometry for intraocular lens power selection in axial myopia. *J Cataract Refract Surg.* (2017) 43:505–10. doi: 10.1016/j.jcrs.2017.01.014

6. Wang L, Shirayama M, Ma XJ, Kohnen T, Koch DD. Optimizing intraocular lens power calculations in eyes with axial lengths above 250 mm. *J Cataract Refract Surg.* (2011) 37:2018–27. doi: 10.1016/j.jcrs.2011.05.042

7. Abulafia A, Barrett GD, Rotenberg M, Kleinmann G, Levy A, Reitblat O, et al. Intraocular lens power calculation for eyes with an axial length greater than 260 mm: comparison of formulas and methods. *J Cataract Refract Surg.* (2015) 41:548–56. doi: 10.1016/j.jcrs.2014.06.033

8. Miyoshi T, Fujie S, Yoshida H, Iwamoto H, Tsukamoto H, Oshika T. Effects of capsular tension ring on surgical outcomes of premium intraocular

lens in patients with suspected zonular weakness. *PLoS ONE*. (2020) 15:e0228999. doi: 10.1371/journal.pone.0228999

9. Ning X, Yang Y, Yan H, Zhang J. Anterior chamber depth - a predictor of refractive outcomes after age-related cataract surgery. *BMC Ophthalmol*. (2019) 19:134. doi: 10.1186/s12886-019-1144-8

10. Wang L, Koch DD. Modified axial length adjustment formulas in long eyes. *J Cataract Refract Surg.* (2018) 44:1396–7. doi: 10.1016/j.jcrs.2018.07.049

11. Melles RB, Kane JX, Olsen T, Chang WJ. Update on intraocular lens calculation formulas. *Ophthalmology.* (2019) 126:1334–5. doi: 10.1016/j.ophtha.2019.04.011

12. Ladas JG, Siddiqui AA, Devgan U, Jun AS. A 3-D "super surface" combining modern intraocular lens formulas to generate a "super formula" and maximize accuracy. *JAMA Ophthalmol.* (2015) 133:1431–6. doi: 10.1001/jamaophthalmol.2015.3832

13. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol.* (2010) 63:854–61. doi: 10.1016/j.jclinepi.2009.10.002

14. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS ONE*. (2013) 8:e76654. doi: 10.1371/journal.pone.0076654

15. van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods.* (2016) 7:80–93. doi: 10.1002/jrsm.1167

16. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol.* (2011) 64:163–71. doi: 10.1016/j.jclinepi.2010.03.016

17. Bernardes J, Raimundo M, Lobo C, Murta JN. A comparison of intraocular lens power calculation formulas in high myopia. *J Refract Surg.* (2021) 37:207–11. doi: 10.3928/1081597X-20201123-01

18. Wan KH, Lam TCH, Yu MCY, Chan TCY. Accuracy and precision of intraocular lens calculations using the new hill-RBF version 20 in eyes with high axial myopia. *Am J Ophthalmol.* (2019) 205:66–73. doi: 10.1016/j.ajo.2019.04.019

19. Wei L, Song Y, He W, Chen X, Ma B, Lu Y, et al. Accuracy improvement of IOL power prediction for highly myopic eyes with an XGBoost machine learning-based calculator. *Front Med*. (2020) 7:592663. doi: 10.3389/fmed.2020.592663

20. Kane JX, Van Heerden A, Atik A, Petsoglou C. Accuracy of 3 new methods for intraocular lens power selection. *J Cataract Refract Surg.* (2017) 43:333–9. doi: 10.1016/j.jcrs.2016.12.021

21. Lin L, Xu M, Mo E, Huang S, Qi X, Gu S, et al. Accuracy of newer generation IOL power calculation formulas in eyes with high axial myopia. *J Refract Surg.* (2021) 37:754–8. doi: 10.3928/1081597X-2021 0712-08

22. Mo E, Lin L, Wang J, Huo Q, Yang Q, Liu E, et al. Clinical accuracy of 6 intraocular lens power calculation formulas in elongated eyes, according to anterior chamber depth. *Am J Ophthalmol.* (2022) 233:153–62. doi: 10.1016/j.ajo.2021.07.017

23. Guo C, Yin S, Qiu K. Zhang M. Comparison of accuracy of intraocular lens power calculation for eyes with an axial length greater than 290 mm. *Int Ophthalmol.* (2022) 42:2029–38. doi: 10.1007/s10792-021-02194-1

24. Chen Y, Wei L, He W, Lu Y, Zhu X. Comparison of Kane, Hill-RBF 2.0, Barrett Universal II, and emmetropia verifying optical formulas in eyes with extreme myopia [published correction appears in J Refract Surg. 2022;38:474]. *J Refract Surg*. (2021) 37:680–5. doi: 10.3928/1081597X-20210712-03

25. Ji J, Liu Y, Zhang J, Wu X, Shao W, Ma B, et al. Comparison of six methods for the intraocular lens power calculation in high myopic eyes. *Eur J Ophthalmol.* (2021) 31:96–102. doi: 10.1177/1120672119889016

26. Wei L, Cheng K, He W, Zhu X, Lu Y. Application of total keratometry in ten intraocular lens power calculation formulas in highly myopic eyes. *Eye Vis*. (2022) 9:21. doi: 10.1186/s40662-022-00293-3

27. Cheng H, Wang L, Kane JX Li J, Liu L, Wu M. Accuracy of artificial intelligence formulas and axial length adjustments for highly myopic eyes. *Am J Ophthalmol*. (2021) 223:100–7. doi: 10.1016/j.ajo.2020.09.019

28. Liu J, Wang L, Chai F, Han Y, Qian S, Koch DD, et al. Comparison of intraocular lens power calculation formulas in Chinese eyes with axial myopia. *J Cataract Refract Surg.* (2019) 45:725–31. doi: 10.1016/j.jcrs.2019.01.018

29. Guo D, He W, Wei L, Song Y, Qi J, Yao Y, et al. The Zhu-Lu formula: a machine learning-based intraocular lens power calculation formula for highly myopic eyes. *Eye Vis*. (2023) 10:26. doi: 10.1186/s40662-023-00342-5

30. Kane JX, Chang DF. Intraocular lens power formulas, biometry, and intraoperative aberrometry: a review. *Ophthalmology*. (2021) 128:e94–114. doi: 10.1016/j.ophtha.2020.08.010

31. Chung J, Bu JJ, Afshari NA. Advancements in intraocular lens power calculation formulas. *Curr Opin Ophthalmol.* (2022) 33:35–40. doi: 10.1097/ICU.0000000000000822

32. Tsessler M, Cohen S, Wang L, Koch DD, Zadok D, Abulafia A. Evaluating the prediction accuracy of the Hill-RBF 30 formula using a heteroscedastic statistical method. *J Cataract Refract Surg.* (2022) 48:37–43. doi: 10.1097/j.jcrs.0000000000000702

33. Wen D, Yu J, Zeng Z, McAlinden C, Hu L, Feng K, et al. Network meta-analysis of no-history methods to calculate intraocular lens power in eyes with previous myopic laser refractive surgery. *J Refract Surg.* (2020) 36:481–90. doi: 10.3928/1081597X-20200519-04

34. Connell BJ, Kane JX. Comparison of the Kane formula with existing formulas for intraocular lens power selection. *BMJ Open Ophthalmol*. (2019) 4:e000251. doi: 10.1136/bmjophth-2018-000251

35. Savini G, Taroni L, Hoffer KJ. Recent developments in intraocular lens power calculation methods-update 2020. *Ann Transl Med.* (2020) 8:1553. doi: 10.21037/atm-20-2290

36. Ang RET, Rapista AJB, Remo JTM, Tan-Daclan MAT, Cruz EM. Clinical outcomes and comparison of intraocular lens calculation formulas in eyes with long axial myopia. *Taiwan J Ophthalmol*. (2021) 12:305–11. doi: 10.4103/tjo.tjo_7_21

37. Barrett GD. Intraocular lens calculation formulas for new intraocular lens implants. *J Cataract Refract Surg.* (1987) 13:389–96. doi: 10.1016/S0886-3350(87)80037-8

38. Barrett GD. An improved universal theoretical formula for intraocular lens power prediction. *J Cataract Refract Surg.* (1993) 19:713–20. doi: 10.1016/S0886-3350(13)80339-2

39. Koch DD, Hill W, Abulafia A, Wang L. Pursuing perfection in intraocular lens calculations: I. Logical approach for classifying IOL calculation formulas. *J Cataract Refract Surg.* (2017) 43:717–8. doi: 10.1016/j.jcrs.2017.06.006

40. Olsen T. Theoretical approach to intraocular lens calculation using Gaussian optics. *J Cataract Refract Surg.* (1987) 13:141–5. doi: 10.1016/S0886-3350(87)80128-1

41. Olsen T, Corydon L, Gimbel H. Intraocular lens power calculation with an improved anterior chamber depth prediction algorithm. *J Cataract Refract Surg.* (1995) 21:313–9. doi: 10.1016/S0886-3350(13)80140-X

42. Olsen T. Prediction of the effective postoperative (intraocular lens) anterior chamber depth. *J Cataract Refract Surg.* (2006) 32:419–24. doi: 10.1016/j.jcrs.2005.12.139

43. Olsen T, Hoffmann P. C constant: new concept for ray tracing-assisted intraocular lens power calculation. *J Cataract Refract Surg.* (2014) 40:764–73. doi: 10.1016/j.jcrs.2013.10.037

frontiers | Frontiers in Public Health

TYPE Mini Review
PUBLISHED 27 November 2023
DOI 10.3389/fpubh.2023.1328918

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Daniele Spoladore,
Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, Italy
Ricardo Valentim,
Federal University of Rio Grande do Norte, Natal, Brazil

*CORRESPONDENCE
Amal Mousa Zaidan
✉ zaidanam@ksau-hs.edu.sa

RECEIVED 27 October 2023
ACCEPTED 13 November 2023
PUBLISHED 27 November 2023

CITATION
Zaidan AM (2023) The leading global health challenges in the artificial intelligence era. *Front. Public Health* 11:1328918.
doi: 10.3389/fpubh.2023.1328918

The leading global health challenges in the artificial intelligence era

Amal Mousa Zaidan[1,2,3]*

[1]King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia, [2]King Abdullah International Medical Research Center (KAIMRC), Riyadh, Saudi Arabia, [3]Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

Millions of people's health is at risk because of several factors and multiple overlapping crises, all of which hit the vulnerable the most. These challenges are dynamic and evolve in response to emerging health challenges and concerns, which need effective collaboration among countries working toward achieving Sustainable Development Goals (SDGs) and securing global health. Mental Health, the Impact of climate change, cardiovascular diseases (CVDs), diabetes, Infectious diseases, health system, and population aging are examples of challenges known to pose a vast burden worldwide. We are at a point known as the "digital revolution," characterized by the expansion of artificial intelligence (AI) and a fusion of technology types. AI has emerged as a powerful tool for addressing various health challenges, and the last ten years have been influential due to the rapid expansion in the production and accessibility of health-related data. The computational models and algorithms can understand complicated health and medical data to perform various functions and deep-learning strategies. This narrative mini-review summarizes the most current AI applications to address the leading global health challenges. Harnessing its capabilities can ultimately mitigate the Impact of these challenges and revolutionize the field. It has the ability to strengthen global health through personalized health care and improved preparedness and response to future challenges. However, ethical and legal concerns about individual or community privacy and autonomy must be addressed for effective implementation.

KEYWORDS
global health, artificial intelligence, public health, global health challenges, machine learning, health care, mortality

Introduction

Millions of people's health and well-being are at risk of several factors and multiple overlapping crises, including but not limited to infectious disease outbreaks, rising malnutrition rates, and lack of sufficient medical access; all hit the vulnerable the most (1). As we are heading toward the end of 2023, a record 339 million people globally need urgent aid. Several critical issues need to be addressed urgently to improve health globally and build resilience against future threats (1); the recent COVID-19 has shown that each country's security and prosperity depend on creating a healthier, safer, more resilient world.

Global public health priorities play a crucial role in addressing the most pressing health challenges faced by populations worldwide. They are dynamic and evolve in response to emerging health challenges and crises, which need effective collaboration among countries to secure global health. Mental Health, the Impact of climate change, cardiovascular diseases

(CVDs), diabetes, Infectious diseases, health system, and population aging are examples of challenges that are examples of these challenges that are known to pose a vast burden worldwide (Figure 1) (2, 3).

We are at a point known as the "digital revolution," characterized by the expansion of artificial intelligence (AI) and a fusion of technology types; the rapid and transformative changes brought about by these advances in digital technology led to increased connectivity and accessibility of vast amounts of information (4). AI has emerged as a powerful tool for addressing various challenges, and the last ten years have been influential in the digital health (5, 6). AI and its subfields or techniques, such as deep learning (DL), natural language processing (NLP), and machine learning (ML), have prospects to benefit healthcare, including public health, because of the rapid expansion in the production and accessibility of health-related data (7–9). The computational models and algorithms can understand complicated health and medical care data to perform various functions and deep-learning strategies. AI in health care improves disease surveillance, diagnosis, treatment selection, and clinical laboratory testing (10, 11). Harnessing its capabilities can ultimately mitigate the impact of global public health issues and revolutionize the field.

However, the burgeoning interest was accompanied by caution over using it, especially in health-related fields. Crucial ethics, privacy, and bias issues were raised to ensure AI's responsible and equitable integration in the global public health landscape. Additionally, the human inability to see how ML systems make their decisions "black box," brought uncertainty and threatened trust among the users regarding its application (12). This review aims to explore the current state of AI in those mentioned global public health challenges and provide insights about its current application in disease diagnosis, medicinal product development, and medical intervention. This step will provide more comprehension, and interpretability for most of AI's leading evidence-based applications and contribute to a better understanding of AI in dealing with the increasing scale and complexity of challenges to global health.

AI in non-communicable diseases

The global development over the last years was also associated with the change in disease burden, and NCDs have almost become the leading cause of mortality, resulting in 200 million premature deaths and expectation of another 150 million people deaths during the next ten years; most of them in low and middle-income countries (13).

NCDs are usually multifactorial diseases, and several factors are associated with their development, including genetic and environmental factors, which make them challenging to prevent and treat effectively. They are exacerbated by four key modifiable risk factors, namely tobacco use, harmful use of alcohol, physical inactivity, and unhealthy diets (14). It was estimated that behavioral factors and genetic factors are the main contributors to preterm death in the USA with (40%), and (30%) respectively (15), and the role of public health awareness and intervention have a vast impact (16). AI can offer the potential to analyze large and complex datasets obtained from lifestyle, clinical, and biological data in a way that exceeds the human ability to make sense of it. ML, NLP, Robotic, virtual agents (chatbots), and speech analysis are just a few examples of the available AI applications that are used to improve public health (17). For example, *Florence*, which is the WHO's first virtual health worker, is designed to help the world's 1.3 billion tobacco users quit smoking (18); *STop obesity Platform* that can offer personalized support to people with obesity (19), and *chatbots* to personalized fitness strategies.

If we discuss kidney disease progression as a case illustration, AI has its application in pre/post-diagnosis, which can ultimately lead to improved outcomes in a timely and accurate manner (20). During the analysis phase of patient data, AI can identify the early signs of the

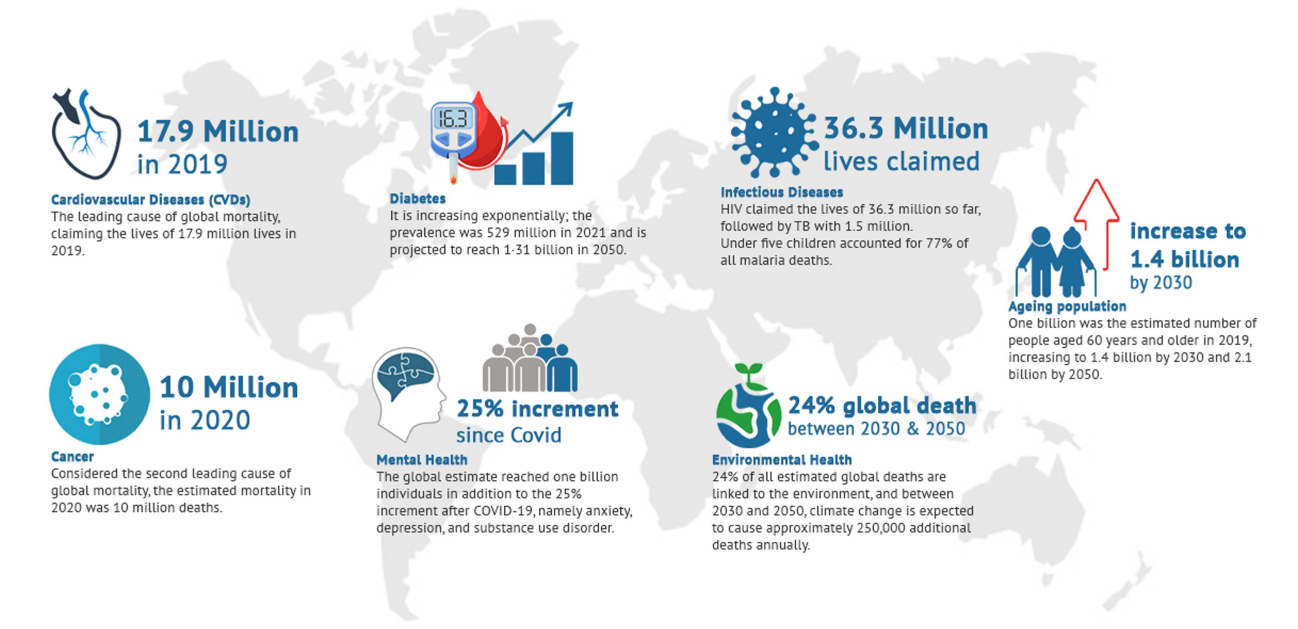


FIGURE 1
Examples of global health challenges.

diseases from the lab results, medical history, and images (21, 22), it can also help in the diagnosis of kidney disease from a kidney biopsy through deep learning-based approaches (23), and to improve outcome and early detection of other comorbidity in renal patients (24). The flowing part discusses AI's applications in the major of NCDs.

AI in cardiovascular diseases

Cardiovascular diseases (CVDs) are the leading cause of global mortality, taking an estimated 17.9 million lives in 2019. Strategies to tackle this considerable burden are directed toward reducing risk factors, enhancing the health system, and monitoring disease patterns and trends to inform national and global actions. The application of AI can identify, process, integrate, and analyze massive amounts of data, including but not limited to medical records, ultrasounds, medications, and experimental results. The gold standard in diagnosing most CVDs is an echocardiogram (ECG) and cardiac magnetic resonance (CMR). However, the clinician's interpretation of ECG depends on their experience. Hence, ECG information might be missed due to clinicians' difficulty analyzing them (25, 26).

Additionally, most ECGs are done on symptomatic patients, even though many CVDs, such as valvular heart diseases, involve long asymptomatic periods (27). However, the AI algorithm interpretation of ECG can diagnose heart failure, atrial fibrillation, hypertrophic cardiomyopathy, pulmonary hypertension, aortic valve stenosis, and anemia (28–36). An approach that reduced time and the physician's cognitive burden by offering pre-diagnosis, correcting clinician errors and preventing the occurrence of misdiagnosis. Moreover, AI enhances the prompt efficiency of medical tools such as computed tomography (CT), echocardiography, magnetic resonance imaging (MRI), and Coronary computed tomography angiography (CCTA) (37–39). In addition to its application in CVD prediction and predictive modeling, such as mortality prediction, vascular aging, and predicting major adverse cardiovascular events in asymptomatic subjects (40–43).

AI in cancer

Cancer is a large group of diseases and is considered the second leading cause of global mortality; the estimated mortality in 2020 was 10 million deaths. Despite the several accomplishments that have been made in the field of cancer diagnosis, prognosis, and treatment, individualized and data-driven care remains challenging. The challenge lies in the specific characteristics of distinct molecular, genetic, and tumor-based features (44). However, using machine learning and AI positively supports cancer prevention and management, as it is reshaping the existing picture, and is developing rapidly (45–47). AI has become pivotal as it can provide patients with forecasting and prediction and improved risk stratification according to specific criteria, such as in the cases of some breast, colon, ovarian, lung, and skin cancers (48–54). Further, it can detect hidden patterns from several sources such as molecular profiling, pathology, and medical imaging, and integration of-omics data to provide a more comprehensive understanding of cancer and improve the precision oncology (55). During a surgical procedure, AI can provide real-time detection and diagnosis of some cancers; through its ability to differentiate between cancerous and normal tissues (56, 57). Recently, mounting evidence indicated the potential role of microRNAs (miRNA) in cancer diagnosis and prognosis. MiRNAs, a small, single-stranded, non-coding ribonucleic acids (RNAs), are essential for all biological functions including cancer development. ML provides an opportunity to explore miRNA's ability to serve as a reliable biomarker targeting drugs and improve cancer clinical classification (58, 59).

A recent achievement is the genetics-based classification and treatment response of Cancer of unknown primary (CUP); this type of cancer usually leads to poor outcomes because primary cancer is unknown (60). The study used ML to classify the cancer based on its genetic profile. This model identifies the likely prior site and predicts the best treatment option (60). Another example is pancreatic cancer, one of the most challenging cancers to diagnose as it is often asymptomatic until it metastasises, causing poor and ineffective treatment. AI modeling enables the detection of individuals at high risk of developing pancreatic cancer; the detection was up to 3 years earlier than currently by using medical records (61).

AI in diabetes

A global crisis that is increasing exponentially and is considered a significant cause of blindness, kidney failure, heart attacks, stroke, and lower limb amputation. It was estimated that the prevalence was 529 million in 2021 and is projected to reach 1·31 billion in 2050 (62), a burden that has a global agreement to halt before. According to several recommendations, the starting point is prevention by screening, especially for obese or overweight adults (63, 64). However, a significant number of cases were missed with these approaches. The current clinical application of AI in diabetes diagnosis and management is categorized into four domains: (1) automatic retinal screening, (2) clinical diagnosis support, (3) patient self-management tools, and (4) risk stratification (65). AI for automatic retinal screening enables early diagnosis with high specificity and sensitivity (66). Several studies have evaluated the prediction of new-onset diabetes mellitus by AI and ML models, and it was recommended to include data as an omics database (e.g., genomics) (67). Recently, a new model based on AI was developed to detect diabetes warning signs, even in patients who did not meet the guidelines for diabetes elevated risk. This model can enhance type 2 diabetes (T2D) detection; it uses the patient's X-ray image collected during routine medical care and their medical records to detect T2D (68).

In diabetes management, AI devices can help patients monitor their glucose levels in real-time and predict spikes or drops in healthcare. AI-based medical devices, such as the Guardian Connect System by Medtronic and the DreaMed Diabetes system (DreaMed Diabetes Ltd), have been approved to help control diabetes (69, 70). Another mobile application (GoCARB) is used to estimate the carbohydrate content in meals, which can help enhance the patients' skills in managing diabetic disease through diet management (71, 72). The future application of AI will introduce a paradigm shift in Diabetes care from conventional management to more personalized and data-driven precision care.

AI in population aging

As a result of significant increases in life expectancy, the global population of people aged 60 years and older is increasing (73). One billion was the estimated number in 2019, rising to 1.4 billion by 2030, accounting for 16.7% of the global population, and projected to reach 2.1 billion by 2050; around 80% will live in low and middle-income countries (74, 75). The increase in age is associated with common health conditions, leading to several complex health states due to multiple underlying factors, such as disability. However, there is no linearity or consistency in developing these changes, and individual characteristics have a vast impact. They will continuously demand primary and long-term care, a more trained workforce, and physical and social environments for social support. Around 92% of global older adults have at least one chronic disease, and more than 81% of those aged ≥85 years suffer from two chronic diseases or more (76–78). Additionally, disability and its consequences also have a huge burden on the aging population. Globally, 1.3 billion people (16% of the world's population) suffer from a physical or cognitive disability; although these estimates cover all age groups, starting from 18 years old (79). However, AI can provide intelligent solutions for longer lives, satisfy the growing unmet healthcare needs, and overcome the limited number of insufficient healthcare resources. Currently, AI technologies for aging population are used in the robotic intervention (80, 81), applications on smartphones or computers (82, 83), social interaction and support, such as improved mental well-being and quality of life (84–86); rehabilitation therapy, such as its application in the recovery of upper and lower extremity functions, gait robotic rehabilitation, or improve sleep quality and daily living activities (87–90), and wearables, voice-activated (91–94). It can create more advanced algorithms to provide more precise holistic interventions tailored to address the elders' multiple needs in a safer and more friendly manner (78, 95, 96). Ambient-assisted working and ambient-assisted living are examples of smart systems that can adapt themselves to older adult needs by exploiting ambient intelligence solutions. These systems focus on using technology to support and enhance the quality of life of the older adult population, either in work, or indoor and outdoor environments (97, 98). Recently, a group of researchers have helped develop drugs that might potentially delay the effects of aging by eliminating senescent cells (96, 99).

AI in mental health

The global estimate of mental health disorders is one billion individuals (100). Since the beginning of the COVID-19 pandemic, the rates of anxiety, depression, and substance use disorder have increased (101, 102). This situation is worse in low-and middle-income countries, where the estimated number of people with limited access and no treatment is around 75% of people with mental, neurological and substance use disorders (100, 102).

Despite the significant advantages of using AI in healthcare, mental health has been slower to adopt AI since the primary factor contributing to successful psychiatric diagnosis and treatment is the interaction with the patients' (103–105). However, AI applications have great potential in diagnosing different kinds of mental illness. This is a great advantage given the available heterogeneity in the pathophysiology of mental illness. AI can access and analyze relevant information about a patient's unique bio-psycho-social characteristics and identify pertinent data patterns that might help provide more objective, improved definitions of these illnesses (106). Further, AI can be used in biomarkers identification, develop better diagnoses and formulate risk models to predict individual risk (105, 107). Moreover, it can be used for some cases, such as depression or autism, where face-to-face interaction might be challenging. In autism, for example, AI could be a more useful tool than a psychotherapy session with a human doctor; it can provide tailored, personalized interventions or bridge the communication gap they may experience (105). Nevertheless, the variation of AI applications is persistent in dealing with sensitive issues like mental health.

AI in infectious diseases

As the world becomes increasingly globalized, health and illnesses have no borders. Concerns about One Health have gained prominence recently, which is justified as the world emerges from the most significant global emergency and the increasing number of infectious pathogens that spread from humans, animals, or the environment. COVID-19 has highlighted the high spreading rate with which infections can devastate the world's health and economy. Results in more investment and investigation into the occurrence, prevalence, prevention, control, and treatment of infectious diseases to strengthen the epidemic response and mobilize quickly for public health priority.

Globally, the leading communicable diseases associated with high mortalities are HIV/AIDS, tuberculosis (TB), malaria, viral hepatitis, sexually transmitted infections, and neglected tropical diseases (NTDs). HIV continues to be a major global issue, claiming the lives of 36.3 million so far, and TB-associated mortalities reached 1.5 million annually, making it the world's second top infectious killer after COVID-19 (108). Tackling AI applications in HIV will yield several examples in HIV prevention, testing, and treatment to achieve sustained viral suppression (109–111). It was used also for rapid detection and response through monitoring clusters of vulnerable groups to reduce HIV transmission (111). Another example is Syphilis, which is a sexually transmitted disease (STD). To eliminate congenital syphilis (CS), the WHO launched an initiative in Latin America and the Caribbean (112). However, as the syphilis epidemic increased in Brazil, the government of Brazil developed a national project, the "Syphilis No!" Project (SNP), for implementing and integrating a syphilis response into healthcare networks, (113, 114). This project encompasses four dimensions: (a) management and governance, (b) surveillance, (c) comprehensive care, and (d) strengthening of the educommunication (113, 115). The application of AI such as data mining and NLP in these strategies augments the country's capabilities in combating syphilis (113).

It is essential to analyze global infectious disease cases regularly. However, some countries' investment in contagious disease identification was typically based on the identification of presenting symptoms and the likelihood of exposure due to the high cost and feasibility of the primary approach of detection (116, 117). However, using big data, AI and ML algorithms can contribute to global infection control and help with the spatial and temporal prediction of the evolution and spread of infectious diseases (118). Their advanced capabilities can analyze several factors: population demographics, environmental conditions, and individual behaviors, all of which can be used simultaneously (119). Such as case prediction according to historical data (120), predicting the likelihood of an individual contracting an infectious disease according to personal and behavioral characteristics, using pathogen genetic makeup to identify the most likely sources of an outbreak, identifying or anticipating an epidemic by analyzing massive data; it can be used for early warning systems, hot spot detection, forecasting, and improving the recourses allocation at a country and a global level (68–72). After the exposure or presence of a potential outbreak, AI can advance in diagnostic approaches and differentiate various pathogens by using the pathogen genetic makeup, such as its ability to distinguish between COVID-19 and other circulating respiratory viruses with COVID-like symptoms (121, 122).

Another example is the possible application to the rising incidence of antimicrobial resistance (AMR), which has become a significant challenge. For this purpose, a group of researchers were able to develop a mobile application to classify bacterial susceptibility to various antibiotics, especially in resource-limited settings (123). Further, reducing transmission is essential to control global widespread infections such as those that occur in pandemics. The application of AI for screening technologies targeting infections and integrating them into data visualization has been introduced broadly, especially during the COVID-19 pandemic (117). This improved the surveillance and generated meaningful insights from multidimensional data, which can be widely used for public health practice.

In addition to surveillance, early detection, and diagnosis, AI is used to develop anti-infective therapies, although it became challenging with the spread of drug resistance (124). ML models can help explore the pathway of pathogen's interaction with host cells and immune responses, facilitating antigen determination, vaccine design, and treatment strategies (124–127). Finally, the WHO global report on infection prevention and control estimated that implementing infection prevention and control (IPC) can reduce healthcare-associated infections (HAIs) by 70% (128). Using AI can improve current and past processes to speed infection prevention and control response, such as identifying the correlations associated with medically relevant conditions, identifying potential risk factors, and surveillance of emerging infectious diseases (129–131), improving hand hygiene compliance (132), and in-hospital analysis of transmission, and outbreak events identification and investigation (133).

AI in environmental health

The impact of environmental health on human lives and health are interconnected in various ways. The Global Health Observatory estimated that 24% of all estimated global deaths are linked to the environment. Between 2030 and 2050, climate change is expected to cause approximately 250,000 additional deaths per year, mainly from undernutrition, malaria, diarrhea and heat stress (134). Because of the adynamic of the environment, AI applications in this field are immense; its deployment will provide a better capacity to deal with the growing climate exigency and related challenges. In exposure assessment, AI can use satellite observations, meteorological variables, land use, and traffic data to predict the spatiotemporal patterns and concentrations of pollutants (135–138). AI were used in monitoring, such as its application during COVID-19, for airport security checks and patient tracking (139), or to improve the prediction of harmful algal blooms (140).

Additionally, it can predict diseases based on environmental factors, such as its application to predict the spread of Zika virus and Dengue fever (141, 142). In waste management, AI reduces fuel consumption and emissions, increases recycling rates, and reduces landfill waste (143). GeoAI is one of the emerging AI tools that can handle complex spatial and temporal data to adjust algorithms and workflows according to the specific characteristics of spatial processes (136, 144). It can develop various environmental exposure models across different geographical regions in prospective and retrospective approaches (136). In 2022, the United Nations launched The World Environment Situation Room, a new digital platform that can provide real-time analysis, track air quality, measure environmental footprint, and monitor (145).

AI in health systems

Through good stewardship, resource development, funding and services, health systems support initiatives to prevent, promote, and provide for more health and well-being (146). They are complex and are in a constant state of flux; according to the World Health Organization (WHO), "A well-functioning health system working in harmony is built on having trained and motivated health workers, a well-maintained infrastructure, and a reliable supply of medicines and technologies backed by adequate funding, strong health plans and evidence-based policies" (147). At the global level, it should be able to control and address global health challenges and severe events (147). However, several myriads of difficulties impede their ability to provide these services. This includes but is not limited to the sudden onset or the slowly growing crises, such as the COVID-19 pandemic, the natural disasters the world is encountering, or the slow time impact of climate change (148–150), the rising number of older people, and the associated complex chronic medical illness. To overcome these difficulties and achieve effective and lasting change, four factors were proposed: (a) the acorn-to-oak tree principle (small initiative), (b) the data-to-information-to-intelligence principle (information technology (IT) and data), (c) the many-hands principle (stakeholders); and (d) the patients-the-preeminent-player principle (individuals) (151). These factors were established across 60 health systems; the role of data and technology cannot be missed (5, 151). AI applications are steadily entering novel domains previously governed solely by human experts. They can improve health financing, make public health more effective, and reach underserved populations by making health care more efficient and effective through more personal health services (152).

Further, the Primary health care system (PHC) is vital to addressing health issues effectively; they are considered the front door of the health care system. Using AI will enhance the holistic approach of PHC in outcome prediction, data mining, and personalized treatment (153–155). The current tools in PHC have several applications, including the risk prediction (156–158), workforce assessment (159), record data extraction (160, 161), control of healthcare-associated infections (162), and performing medical tasks remotely that contribute to public health domain (154, 163, 164).

Conclusion

AI integration and application to global health challenges have immense potential to overcome them efficiently and effectively. Disease prevention, detection, and response can quickly mobilize and yield medicinal products. As mentioned earlier, around 40% of preterm death was associated with behavioral factors. With AI advancement, data analysis and segmentation can be done for several characteristics such as behavior, opinion, and attitude. Using these data, the ML can analyze the online health information and provide personalized massaging to influence individuals' health behaviors with high quality and clarity, amplifying their influence and effectiveness (165–167). This health communication can also inform AI technology in developing effective communication systems with patients and their healthcare providers. The health communication theories and models can highlight the available barriers to behavioral change and the available limitations of technology-driven health interventions. Which can help improve the efficacy of AI-supported systems or intervention designs (167–169).

AI in health care is expected to grow from nearly US $15 billion to $103 billion between 2023 and 2028 (170). However, incongruent with AI's benefits, exacerbation of inequities was accompanied, and ethical and legal concerns about individual or community privacy and autonomy were raised (118, 171). The EU AI Act is nearing implementation, and it will be the first comprehensive regulation that addresses the risks of artificial intelligence; European Parliament proposed it to ensure better conditions for developing and using this innovative technology (172). Further, to avoid the risk of hindering AI applications in healthcare due to lack of sufficient transparency "black box," researchers were urged to provide more research and explanation for AI; explainable AI (xAI), as an approach to more understandable and human-interpretable AI-based applications (173, 174).

However, AI stands as a cornerstone of the upcoming digital revolution. Despite the moral dilemmas in AI application in health care, it is likely to meager, co-exist or replace current systems and assets as a potent amplifier of human potential. It has the ability to strengthen global health through personalized health care and improved preparedness and response to future challenges.

Author contributions

AZ: Conceptualization, Writing – original draft.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. World Health Organization. *WHO's health emergency appeal 2023*. Geneva: World Health Organization (2023a).

2. Chan A Eleven global health issues to watch in 2023, according to IHME experts: the Institute for Health Metrics and Evaluation. (2022). Available at: https://www.healthdata.org/news-events/insights-blog/acting-data/11-global-health-issues-watch-2023-according-ihme-experts.

3. Lucero-Prisno DE, Shomuyiwa DO, Kouwenhoven MBN, Dorji T, Odey GO, Miranda AV, et al. Top 10 public health challenges to track in 2023: shifting focus beyond a global pandemic. *Public Health Challenges*. (2023) 2:e86. doi: 10.1002/puh2.86

4. Schwab K The fourth industrial revolution: what it means, how to respond New York. (2016). Available at: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.

5. World Bank. *Digital-in-health: Unlocking the value for everyone*. Washington DC: World Bank (2023).

6. U.S. Agency for International Development. *Artificial intelligence in Global Health: Defining a collective path forward*. Washington: U.S. Agency for International Development (2019).

7. Suleimenov IE, Vitulyova YS, Bakirov AS, Gabrielyan OA. Artificial intelligence: What is it? Proceedings of the 2020 6th International Conference on Computer and Technology Applications Antalya, Turkey: Association for Computing Machinery (2020).

8. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare J*. (2019) 6:94–8. doi: 10.7861/futurehosp.6-2-94

9. Kubat M. *A simple machine-learning task. An introduction to machine learning*. Cham: Springer International Publishing (2017).

10. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. (2023) 23:689. doi: 10.1186/s12909-023-04698-z

11. Rubens M, Ramamoorthy V, Saxena A, Shehadeh N. Public health in the twenty-first century: the role of advanced technologies. *Front Media*. (2014) 2:224. doi: 10.3389/fpubh.2014.00224

12. Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J. Implementing AI in healthcare—the relevance of trust: a scoping review. *Front Health Serv*. (2023) 3:150. doi: 10.3389/frhs.2023.1211150

13. World Health Organization. *Noncommunicable diseases*. Geneva: World Health Organization (2014).

14. World Health Organization. *Noncommunicable diseases: Key facts*. Geneva: World Health Organization (2023b).

15. Schroeder SA. We can do better — improving the health of the American people. *N Engl J Med*. (2007) 357:1221–8. doi: 10.1056/NEJMsa073350

16. Frieden TR. The future of public health. *N Engl J Med*. (2015) 373:1748–54. doi: 10.1056/NEJMsa1511248

17. Pan American Health Organization. *Artificial intelligence in public health*. Washington: Pan American Health Organization (2021).

18. World Health Organization. *Using AI to lead a healthier lifestyle: How Florence can help you*. Geneva: World Health Organization (2022a).

19. European Commission. STop obesity platform. (2019). Available at: https://cordis.europa.eu/project/id/823978.

20. Bellocchio F, Zhang H. Editorial: artificial intelligence in nephrology. *Front. Nephrol*. (2023) 3:769. doi: 10.3389/fneph.2023.1270769

21. Fayos De Arizón L, Viera ER, Pilco M, Perera A, De Maeztu G, Nicolau A, et al. Artificial intelligence: a new field of knowledge for nephrologists? *Clin Kidney J*. (2023). doi: 10.1093/ckj/sfad182

22. Nong Y, Wei X, Qiu H, Yang H, Yang J, Lu J, et al. Analysis of risk factors for severe acute kidney injury in patients with acute myocardial infarction: a retrospective study. *Front Nephrology*. (2023) 3:249. doi: 10.3389/fneph.2023.1047249

23. Basso MN, Barua M, Meyer J, John R, Khademi A. Machine learning in renal pathology. *Front Nephrology*. (2022) 2:7002. doi: 10.3389/fneph.2022.1007002

24. Canaud B, Kooman J, Davenport A, Campo D, Carreel E, Morena-Carrere M, et al. Digital health technology to support care and improve outcomes of chronic kidney

disease patients: as a case illustration, the Withings toolkit health sensing tools. *Front Nephrol*. (2023) 3:565. doi: 10.3389/fneph.2023.1148565

25. Elias P, Poterucha TJ, Rajaram V, Moller LM, Rodriguez V, Bhave S, et al. Deep learning electrocardiographic analysis for detection of left-sided Valvular heart disease. *J Am Coll Cardiol*. (2022) 80:613–26. doi: 10.1016/j.jacc.2022.05.029

26. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. (2021) 18:465–78. doi: 10.1038/s41569-020-00503-2

27. Lancellotti P, Magne J, Dulgheru R, Clavel MA, Donal E, Vannan MA, et al. Outcomes of patients with asymptomatic aortic stenosis followed up in heart valve clinics. *JAMA Cardiol*. (2018) 3:1060–8. doi: 10.1001/jamacardio.2018.3152

28. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. (2019) 25:70–4. doi: 10.1038/s41591-018-0240-2

29. Attia ZI, Kapa S, Noseworthy PA, Lopez-Jimenez F, Friedman PA. Artificial intelligence ECG to detect left ventricular dysfunction in COVID-19: a case series. *Mayo Clin Proc*. (2020) 95:2464–6. doi: 10.1016/j.mayocp.2020.09.020

30. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. (2019) 394:861–7. doi: 10.1016/S0140-6736(19)31721-0

31. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*. (2021) 27:815–9. doi: 10.1038/s41591-021-01335-4

32. Kwon JM, Cho Y, Jeon KH, Cho S, Kim KH, Baek SD, et al. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health*. (2020) 2:e358–67. doi: 10.1016/S2589-7500(20)30108-4

33. Kwon JM, Kim KH, Medina-Inojosa J, Jeon KH, Park J, Oh BH. Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography. *J Heart Lung Transplant*. (2020) 39:805–14. doi: 10.1016/j.healun.2020.04.009

34. Ko WY, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. (2020) 75:722–33. doi: 10.1016/j.jacc.2019.12.030

35. Kwon JM, Lee SY, Jeon KH, Lee Y, Kim KH, Park J, et al. Deep learning-based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc*. (2020) 9:e014717. doi: 10.1161/JAHA.119.014717

36. Cohen-Shelly M, Attia ZI, Friedman PA, Ito S, Essayagh BA, Ko WY, et al. Electrocardiogram screening for aortic valve stenosis using artificial intelligence. *Eur Heart J*. (2021) 42:2885–96. doi: 10.1093/eurheartj/ehab153

37. Choi AD, Marques H, Kumar V, Griffin WF, Rahban H, Karlsberg RP, et al. CT evaluation by artificial intelligence for atherosclerosis, stenosis and vascular morphology (CLARIFY): a multi-center, international study. *J Cardiovasc Comput Tomogr*. (2021) 15:470–6. doi: 10.1016/j.jcct.2021.05.004

38. Lin A, Manral N, McElhinney P, Killekar A, Matsumoto H, Kwiecinski J, et al. Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *Lancet Digit Health*. (2022) 4:e256–65. doi: 10.1016/S2589-7500(22)00022-X

39. Knott KD, Seraphim A, Augusto JB, Xue H, Chacko L, Aung N, et al. The prognostic significance of quantitative myocardial perfusion: an artificial intelligence-based approach using perfusion mapping. *Circulation*. (2020) 141:1282–91. doi: 10.1161/CIRCULATIONAHA.119.044666

40. CGdS E Silva, Buginga GC, Eads E Silva, Arena R, Rouleau CR, Aggarwal S, et al., Prediction of mortality in coronary artery disease: role of machine learning and maximal exercise capacity. Mayo Clinic proceedings; (2022). Elsevier.

41. Raghunath S, Ulloa Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med*. (2020) 26:886–91. doi: 10.1038/s41591-020-0870-z

42. Toya T, Ahmad A, Attia Z, Cohen-Shelly M, Ozcan I, Noseworthy PA, et al. Vascular aging detected by peripheral endothelial dysfunction is associated with ECG-derived physiological aging. *J Am Heart Assoc*. (2021) 10:e018656. doi: 10.1161/JAHA.120.018656

43. Eisenberg E, McElhinney PA, Commandeur F, Chen X, Cadet S, Goeller M, et al. Deep learning-based quantification of Epicardial adipose tissue volume and attenuation predicts major adverse cardiovascular events in asymptomatic subjects. *Circ Cardiovasc Imaging*. (2020) 13:e009829. doi: 10.1161/CIRCIMAGING.119.009829

44. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in Cancer research and precision medicine. *Cancer Discov*. (2021) 11:900–15. doi: 10.1158/2159-8290.CD-21-0090

45. Gaur K, Jagtap MM. Role of artificial intelligence and machine learning in prediction, diagnosis, and prognosis of Cancer. *Cureus*. (2022) 14:e31008. doi: 10.7759/cureus.31008

46. Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett*. (2020) 471:61–71. doi: 10.1016/j.canlet.2019.12.007

47. Wang S-H, Chen G, Zhong X, Lin T, Shen Y, Fan X, et al. Global development of artificial intelligence in cancer field: a bibliometric analysis range from 1983 to 2022. *Front Oncol*. (2023) 13:1215729. doi: 10.3389/fonc.2023.1215729

48. Enshaei A, Robson CN, Edmondson RJ. Artificial intelligence systems as prognostic and predictive tools in ovarian Cancer. *Ann Surg Oncol*. (2015) 22:3970–5. doi: 10.1245/s10434-015-4475-6

49. Nartowt BJ, Hart GR, Muhammad W, Liang Y, Stark GF, Deng J. Robust machine learning for colorectal Cancer risk prediction and stratification. *Front Big Data*. (2020) 3:6. doi: 10.3389/fdata.2020.00006

50. Nartowt BJ, Hart GR, Roffman DA, Llor X, Ali I, Muhammad W, et al. Scoring colorectal cancer risk with an artificial neural network based on self-reportable personal health data. *PLoS One*. (2019) 14:e0221421. doi: 10.1371/journal.pone.0221421

51. Hart GR, Roffman DA, Decker R, Deng J. A multi-parameterized artificial neural network for lung cancer risk prediction. *PLoS One*. (2018) 13:e0205264. doi: 10.1371/journal.pone.0205264

52. Stark GF, Hart GR, Nartowt BJ, Deng J. Predicting breast cancer risk using personal health data and machine learning models. *PLoS One*. (2019) 14:e0226765. doi: 10.1371/journal.pone.0226765

53. Roffman D, Hart G, Girardi M, Ko CJ, Deng J. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Sci Rep*. (2018) 8:1701. doi: 10.1038/s41598-018-19907-9

54. Hunter B, Chen M, Ratnakumar P, Alemu E, Logan A, Linton-Reid K, et al. A radiomics-based decision support tool improves lung cancer diagnosis in combination with the herder score in large lung nodules. *EBioMedicine*. (2022) 86:104344. doi: 10.1016/j.ebiom.2022.104344

55. Perez-Lopez R, Reis-Filho JS, Kather JN. A framework for artificial intelligence in cancer research and precision oncology. *NPJ Prec Oncology*. (2023) 7:43. doi: 10.1038/s41698-023-00383-y

56. Hollon TC, Pandian B, Adapa AR, Urias E, Save AV, Khalsa SSS, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med*. (2020) 26:52–8. doi: 10.1038/s41591-019-0715-9

57. Zhang B, Shi H, Wang H. Machine learning and AI in Cancer prognosis, prediction, and treatment selection: a critical approach. *J Multidiscip Healthc*. (2023) 16:1779–91. doi: 10.2147/JMDH.S410301

58. Luo Y, Peng L, Shan W, Sun M, Luo L, Liang W. Machine learning in the development of targeting microRNAs in human disease. *Front Genet*. (2023) 13:189. doi: 10.3389/fgene.2022.1088189

59. Azari H, Nazari E, Mohit R, Asadnia A, Maftooh M, Nassiri M, et al. Machine learning algorithms reveal potential miRNAs biomarkers in gastric cancer. *Sci Rep*. (2023) 13:6147. doi: 10.1038/s41598-023-32332-x

60. Moon I, LoPiccolo J, Baca SC, Sholl LM, Kehl KL, Hassett MJ, et al. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. *Nat Med*. (2023) 29:2057–67. doi: 10.1038/s41591-023-02482-6

61. Placido D, Yuan B, Hjaltelin JX, Zheng C, Haue AD, Chmura PJ, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nat Med*. (2023) 29:1113–22. doi: 10.1038/s41591-023-02332-5

62. The Lancet. Diabetes: a defining disease of the 21st century. *Lancet*. (2023) 401:2087. doi: 10.1016/S0140-6736(23)01296-5

63. Duan D, Kengne AP, Echouffo-Tcheugui JB. Screening for diabetes and prediabetes. *Endocrinol Metab Clin*. (2021) 50:369–85. doi: 10.1016/j.ecl.2021.05.002

64. Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, Davis EM, et al. Screening for prediabetes and type 2 diabetes: US preventive services task force recommendation statement. *JAMA*. (2021) 326:736–43. doi: 10.1001/jama.2021.12531

65. Ellahham S. Artificial intelligence: The future for diabetes care. *Am J Med*. (2020) 133:895–00. doi: 10.1016/j.amjmed.2020.03.033

66. Huang X, Wang H, She C, Feng J, Liu X, Hu X, et al. Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy. *Front Endocrinol*. (2022) 13:946915. doi: 10.3389/fendo.2022.946915

67. Nomura A, Noguchi M, Kometani M, Furukawa K, Yoneda T. Artificial intelligence in current diabetes management and prediction. *Curr Diab Rep*. (2021) 21:61. doi: 10.1007/s11892-021-01423-2

68. Pyrros A, Borstelmann SM, Mantravadi R, Zaiman Z, Thomas K, Price B, et al. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. *Nat Commun*. (2023) 14:4039. doi: 10.1038/s41467-023-39631-x

69. Abraham SB, Arunachalam S, Zhong A, Agrawal P, Cohen O, McMahon CM. Improved real-world glycemic control with continuous glucose monitoring system predictive alerts. (2021) 1932–2968 Contract No: 1

70. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. (2020) 3:118. doi: 10.1038/s41746-020-00324-0

71. Guan Z, Li H, Liu R, Cai C, Liu Y, Li J, et al. Artificial intelligence in diabetes management: advancements, opportunities, and challenges. *Cell Rep Med*. (2023) 4:101213. doi: 10.1016/j.xcrm.2023.101213

72. Vasiloglou MF, Mougiakakou S, Aubry E, Bokelmann A, Fricker R, Gomes F, et al. A comparative study on carbohydrate estimation: GoCARB vs. *Diet Nutr*. (2018) 10:741. doi: 10.3390/nu10060741

73. Beard JR, Officer A, De Carvalho IA, Sadana R, Pot AM, Michel J-P, et al. The world report on ageing and health: a policy framework for healthy ageing. *Lancet*. (2016) 387:2145–54. doi: 10.1016/S0140-6736(15)00516-4

74. World Health Organization. Ageing and health. (2022b). Available at: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health.

75. World health Organization. (2023c). *Ageing Geneva* Available at: https://www.who.int/health-topics/ageing#tab=tab_1.

76. Tkatch R, Musich S, MacLeod S, Alsgaard K, Hawkins K, Yeh CS. Population health management for older adults: review of interventions for promoting successful aging across the health continuum. *Gerontol Geriatr Med*. (2016) 2:787. doi: 10.1177/2333721416667877

77. Salive ME. Multimorbidity in older adults. *Epidemiol Rev*. (2013) 35:75–83. doi: 10.1093/epirev/mxs009

78. Ma B, Yang J, Wong FKY, Wong AKC, Ma T, Meng J, et al. Artificial intelligence in elderly healthcare: a scoping review. *Ageing Res Rev*. (2023) 83:101808. doi: 10.1016/j.arr.2022.101808

79. World Health Organization. *Disability*. Geneva: WHO (2023d).

80. Hu X, Zeng X, Xu Y, Luo C, Jia L, Zhao Z, et al. A soft robotic intervention for gait enhancement in older adults. *IEEE Trans Neural Syst Rehabil Eng*. (2021) 29:1838–47. doi: 10.1109/TNSRE.2021.3109729

81. Pu L, Moyle W, Jones C, Todorovic M. The effect of a social robot intervention on sleep and motor activity of people living with dementia and chronic pain: a pilot randomized controlled trial. *Maturitas*. (2021) 144:16–22. doi: 10.1016/j.maturitas.2020.09.003

82. Valero MÁ, Bravo J, Chamizo JMG, López-de-Ipiña D. Integration of multisensor hybrid reasoners to support personal autonomy in the smart home. *Sensors*. (2014) 14:17313–30. doi: 10.3390/s140917313

83. VandeWeerd C, Yalcin A, Aden-Buie G, Wang Y, Roberts M, Mahser N, et al. HomeSense: design of an ambient home health and wellness monitoring platform for older adults. *Heal Technol*. (2020) 10:1291–09. doi: 10.1007/s12553-019-00404-6

84. Follmann A, Schollemann F, Arnolds A, Weismann P, Laurentius T, Rossaint R, et al. Reducing loneliness in stationary geriatric care with robots and virtual encounters—a contribution to the COVID-19 pandemic. *Int J Environ Res Public Health*. (2021) 18:4846. doi: 10.3390/ijerph18094846

85. Pei-Ti H, Chen-Shie H, Ya-Fang H, Jeu-Jung C. The effects of a social participation app on seniors. *J Nurs Res*. (2021) 29:e168. doi: 10.1097/JNR.0000000000000451

86. Aggar C, Sorwar G, Seton C, Penman O, Ward A. Smart home technology to support older people's quality of life: a longitudinal pilot study. *Int J Older People Nursing*. (2023) 18:e12489. doi: 10.1111/opn.12489

87. Park J-H. The effects of robot-assisted left-hand training on hemispatial neglect in older patients with chronic stroke: a pilot and randomized controlled trial. *Medicine*. (2021) 100:e24781. doi: 10.1097/MD.0000000000024781

88. Calabrò RS, De Luca R, Leo A, Balletta T, Marra A, Bramanti P. Lokomat training in vascular dementia: motor improvement and beyond! *Aging Clin Exp Res*. (2015) 27:935–7. doi: 10.1007/s40520-015-0343-2

89. Koumpouros Y, Toulias TL, Tzafestas CS, Moustris G. Assessment of an intelligent robotic rollator implementing navigation assistance in frail seniors. *Technol Disabil*. (2020) 32:159–77. doi: 10.3233/TAD-200271

90. Mizuno J, Saito D, Sadohara K, Nihei M, Ohnaka S, Suzurikawa J, et al. Effect of the information support robot on the daily activity of older people living alone in actual living environment. *Int J Environ Res Public Health*. (2021) 18:2498. doi: 10.3390/ijerph18052498

91. Ahmed MU. A personalized health-monitoring system for elderly by combining rules and case-based reasoning. *pHealth*. (2015) 1:249–54.

92. Tsamis KI, Rigas G, Nikolaos K, Fotiadis DI, Konitsiotis S. Accurate monitoring of Parkinson's disease symptoms with a wearable device during COVID-19 pandemic. *In Vivo*. (2021) 35:2327–30. doi: 10.21873/invivo.12507

93. Wang L-H, Hsiao Y-M, Xie X-Q, Lee S-Y. An outdoor intelligent healthcare monitoring device for the elderly. *IEEE Trans Consum Electron*. (2016) 62:128–35. doi: 10.1109/TCE.2016.7514671

94. Lazarou I, Stavropoulos TG, Meditskos G, Andreadis S, Kompatsiaris IY, Tsolaki M. Long-term impact of intelligent monitoring technology on people with cognitive impairment: an observational study. *J Alzheimers Dis*. (2019) 70:757–92. doi: 10.3233/JAD-190423

95. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. *The Gerontologist*. (2019) 59:e37–51. doi: 10.1093/geront/gny046

96. Marino N, Putignano G, Cappilli S, Chersoni E, Santuccione A, Calabrese G, et al. Towards AI-driven longevity research: an overview. *Front Aging*. (2023) 4:1057204. doi: 10.3389/fragi.2023.1057204

97. Guerra BMV, Torti E, Marenzi E, Schmid M, Ramat S, Leporati F, et al. Ambient assisted living for frail people through human activity recognition: state-of-the-art, challenges and future directions. *Front Neurosci*. (2023) 17:1256682. doi: 10.3389/fnins.2023.1256682

98. Spoladore D, Trombetta A. Ambient assisted working solutions for the ageing workforce: a literature review. *Electronics*. (2023) 12:101. doi: 10.3390/electronics12010101

99. Smer-Barreto V, Quintanilla A, Elliott RJR, Dawson JC, Sun J, Campa VM, et al. Discovery of senolytics using machine learning. *Nat Commun*. (2023) 14:3445. doi: 10.1038/s41467-023-39120-1

100. World Health Organization (2020). World mental health day: an opportunity to kick-start a massive scale-up in investment in mental health. Geneva: World Health Organization.

101. National Institutes of Health (2023). Mental Health During the COVID-19 Pandemic: an urgent issue. Available at: https://covid19.nih.gov/covid-19-topics/mental-health#:~:text=Rates%20of%20anxiety%2C%20depression%2C%20and,have%20mental%20illnesses%20or%20disorders.

102. Gao Y, Bagheri N, Furuya-Kanamori L. Has the COVID-19 pandemic lockdown worsened eating disorders symptoms among patients with eating disorders? systematic review. *A Z Gesundh Wiss*. (2022) 30:2743–52. doi: 10.1007/s10389-022-01704-4

103. Fei J, Yong J, Hui Z, Yi D, Hao L, Sufeng M, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. (2017) 2:230–43. doi: 10.1136/svn-2017-000101

104. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med*. (2018) 131:129–033. doi: 10.1016/j.amjmed.2017.10.035

105. Minerva F, Giubilini A. Is AI the future of mental healthcare? *Topoi (Dordr)*. (2023) 42:1–9. doi: 10.1007/s11245-023-09932-3

106. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep*. (2019) 21:116. doi: 10.1007/s11920-019-1094-0

107. D'Alfonso S. AI in mental health. *Curr Opin Psychol*. (2020) 36:112–7. doi: 10.1016/j.copsyc.2020.04.005

108. World Health Organization (n.d.). Communicable and noncommunicable diseases, and mental health. Available at: https://www.who.int/our-work/communicable-and-noncommunicable-diseases-and-mental-health

109. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using clinical notes and natural language processing for automated HIV risk assessment. *J Acquir Immune Defic Syndr*. (2018) 77:160–6. doi: 10.1097/QAI.0000000000001580

110. Kakalou C, Lazarus JV, Koutkias V. Mining social Media for Perceptions and Trends on HIV pre-exposure prophylaxis. *Stud Health Technol Inform*. (2019) 264:959–63. doi: 10.3233/SHTI190366

111. Xiang Y, Du J, Fujimoto K, Li F, Schneider J, Tao C. Application of artificial intelligence and machine learning for HIV prevention interventions. *Lancet HIV*. (2022) 9:e54–62. doi: 10.1016/S2352-3018(21)00247-2

112. Pan American Health Organization. PAHO launches new initiative to eliminate mother-to-child transmission of four diseases. (2017) Available at: https://www3.paho.org/hq/index.php?option=com_content&view=article&id=13567:paho-launches-new-initiative-to-eliminate-mother-to-child-transmission-of-four-diseases&Itemid=0&lang=en#gsc.tab=0.

113. da Rocha MA, dos Santos MM, Fontes RS, de Melo ASP, Cunha-Oliveira A, Miranda AE, et al. The text mining technique applied to the analysis of health interventions to combat congenital syphilis in Brazil: The case of the "syphilis no!" Project. *Front Public Health*. (2022) 10:680. doi: 10.3389/fpubh.2022.855680

114. Pinto R, Valentim R, Fernandes da Silva L, Fontoura de Souza G, Farias G, De Moura Santos Lima T, et al. Use of interrupted time series analysis in understanding the course of the congenital syphilis epidemic in Brazil. The Lancet regional health. *Americas*. (2022) 7:100163. doi: 10.1016/j.lana.2021.100163

115. Caitano AR, Gusmão CMG, Dias-Trindade S, Barbalho IMP, Morais PSG, Caldeira-Silva GJP, et al. Massive health education through technological mediation: analyses and impacts on the syphilis epidemic in Brazil. *Front Public Health*. (2022) 10:944213. doi: 10.3389/fpubh.2022.944213

116. Sisimayi C, Harley C, Nyabadza F, Visaya MV. AI-enabled case detection model for infectious disease outbreaks in resource-limited settings. *Front Appl Math Statis*. (2023) 9:9. doi: 10.3389/fams.2023.1133349

117. Brownstein JS, Rader B, Astley CM, Tian H. Advances in artificial intelligence for infectious-disease surveillance. *N Engl J Med*. (2023) 388:1597–607. doi: 10.1056/NEJMra2119215

118. World Health Organization. *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization (2021).

119. Santangelo OE, Gentile V, Pizzo S, Giordano D, Cedrone F. Machine learning and prediction of infectious diseases: a systematic review. *Mach Learn Knowl Extract*. (2023) 5:175–98. doi: 10.3390/make5010013

120. Palaniappan S, David B. Prediction of epidemic disease dynamics on the infection risk using machine learning algorithms. *SN Comp Sci*. (2021) 3:97. doi: 10.1007/s42979-021-00902-3

121. Maharaj AS, Parker J, Hopkins JP, Gournis E, Bogoch II, Rader B, et al. The effect of seasonal respiratory virus transmission on syndromic surveillance for COVID-19 in Ontario Canada. *Lancet Infect Dis*. (2021) 21:593–4. doi: 10.1016/S1473-3099(21)00151-1

122. Roy S, Biswas P, Ghosh P. Spatiotemporal tracing of pandemic spread from infection data. *Sci Rep*. (2021) 11:17689. doi: 10.1038/s41598-021-97207-5

123. Pascucci M, Royer G, Adamek J, Asmar MA, Aristizabal D, Blanche L, et al. AI-based mobile application to fight antibiotic resistance. *Nat Commun*. (2021) 12:1173. doi: 10.1038/s41467-021-21187-3

124. Wong F, de la Fuente-Nunez C, Collins JJ. Leveraging artificial intelligence in the fight against infectious diseases. *Science*. (2023) 381:164–70. doi: 10.1126/science.adh1114

125. Bojar D, Powers RK, Camacho DM, Collins JJ. Deep-learning resources for studying glycan-mediated host-microbe interactions. *Cell Host Microbe*. (2021) 29:132–44.e3. doi: 10.1016/j.chom.2020.10.004

126. Hie B, Zhong ED, Berger B, Bryson B. Learning the language of viral evolution and escape. *Science*. (2021) 371:284–8. doi: 10.1126/science.abd7331

127. Wheeler NE, Gardner PP, Barquist L. Machine learning identifies signatures of host adaptation in the bacterial pathogen *Salmonella enterica*. *PLoS Genet*. (2018) 14:e1007333. doi: 10.1371/journal.pgen.1007333

128. World Health Organization. *Global report on infection prevention and control*. Geneva, World Health Organization: (2022c).

129. Fitzpatrick F, Doherty A, Lacey G. Using artificial intelligence in infection prevention. *Curr Treat Options Infect Dis*. (2020) 12:135–44. doi: 10.1007/s40506-020-00216-7

130. Bechmann L, Geginat G. Semi-automated contact tracing and management of contact precautions during the COVID-19 pandemic within a tertiary hospital. *Infect Prev Pract*. (2023) 5:100266. doi: 10.1016/j.infpip.2022.100266

131. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis*. (2014) 14:160–8. doi: 10.1016/S1473-3099(13)70244-5

132. Worlikar H, Vadhiraj VV, Murray A, O'Connell J, Connolly C, Walsh JC, et al. Is it feasible to use a humanoid robot to promote hand hygiene adherence in a hospital setting? *Infect Prev Pract*. (2022) 4:100188. doi: 10.1016/j.infpip.2021.100188

133. Branch-Elliman W, Sundermann AJ, Wiens J, Shenoy ES. The future of automated infection detection: innovation to transform practice (part III/III). *Antimicrob Steward Healthc Epidemiol*. (2023) 3:e26. doi: 10.1017/ash.2022.333

134. World Health Organization. *Public health and environment*. Geneva: World Health Organization (2023e).

135. Nti EK, Cobbina SJ, Attafuah EA, Senanu LD, Amenyeku G, Gyan MA, et al. Water pollution control and revitalization using advanced technologies: uncovering artificial intelligence options towards environmental health protection, sustainability and water security. *Heliyon*. (2023) 9:e18170. doi: 10.1016/j.heliyon.2023.e18170

136. VoPham T, Hart JE, Laden F, Chiang Y-Y. Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environ Health*. (2018) 17:40. doi: 10.1186/s12940-018-0386-x

137. Chinnaswamy A, Walton N. AI and environmental health: a platform ecosystem perspective In: HL Ong, R-a Doong, R Naguib, CP Lim and AK Nagar, editors. *Artificial intelligence and environmental sustainability: Challenges and solutions in the era of industry 40*. Singapore: Springer Nature Singapore (2022)

138. Weichenthal S, Hatzopoulou M, Brauer M. A picture tells a thousand… exposures: opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environ Int*. (2019) 122:3–10. doi: 10.1016/j.envint.2018.11.042

139. Alafif T, Tehame AM, Bajaba S, Barnawi A, Zia S. Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *Int J Environ Res Public Health*. (2021) 18:1117. doi: 10.3390/ijerph18031117

140. Lee S, Lee D. Improved prediction of harmful algal blooms in four Major South Korea's Rivers using deep learning models. *Int J Environ Res Public Health*. (2018) 15:1322. doi: 10.3390/ijerph15071322

141. Gangula R, Thirupathi L, Parupati R, Sreeveda K, Gattoju S. Ensemble machine learning based prediction of dengue disease with performance and accuracy elevation patterns. *Mater Today Proceed*. (2023) 80:3458–63.

142. Dadheech P, Mehbodniya A, Tiwari S, Kumar S, Singh P, Gupta S, et al. Zika virus prediction using AI-driven technology and hybrid optimization algorithm in healthcare. *J Healthcare Eng*. (2022) 2022:1–13. doi: 10.1155/2022/2793850

143. Fayyazi M, Sardar P, Thomas SI, Daghigh R, Jamali A, Esch T, et al. Artificial intelligence/machine learning in energy management systems, control, and optimization of hydrogen fuel cell vehicles. *Sustainability*. (2023) 15:5249. doi: 10.3390/su15065249

144. Fan Z, Yan Z, Wen S. Deep learning and artificial intelligence in sustainability: a review of SDGs, renewable energy, and environmental health. *Sustainability*. (2023) 15:13493. doi: 10.3390/su151813493

145. United Nations. World environment situation room: data, information and knowledge on the environment. (n.d.). United Nation Environment Programme Available at: https://wesr.unep.org

146. Darrudi A, Ketabchi Khoonsari MH, Tajvar M. Challenges to achieving universal health coverage throughout the world: a systematic review. *J Prev Med Public Health*. (2022) 55:125–33. doi: 10.3961/jpmph.21.542

147. World Health Organization. *Health systems strengthening*. Geneva: World Health Organization (2023f).

148. Mora C, McKenzie T, Gaw IM, Dean JM, von Hammerstein H, Knudson TA, et al. Over half of known human pathogenic diseases can be aggravated by climate change. *Nat Clim Chang*. (2022) 12:869–75. doi: 10.1038/s41558-022-01426-1

149. World Health Organization. *Climate action: fast facts [press release]*. Geneva: World Health Organization (2022d).

150. Wilson D, Sheikh A, Görgens M, Ward K. Technology and universal health coverage: examining the role of digital health. *J Glob Health*. (2021) 11:16006. doi: 10.7189/jogh.11.16006

151. Braithwaite J, Mannion R, Matsuyama Y, Shekelle P, Whittaker S, Al-Adawi S, et al. Accomplishing reform: successful case studies drawn from the health systems of 60 countries. *Int J Qual Health Care*. (2017) 29:880–6. doi: 10.1093/intqhc/mzx122

152. McKinsey & Company. *How digital tools could boost efficiency in African health systems*. Chicago: McKinsey & Company (2023).

153. Kueper JK. Primer for artificial intelligence in primary care. *Can Fam Physician*. (2021) 67:889–93. doi: 10.46747/cfp.6712889

154. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. *J Gen Intern Med*. (2019) 34:1626–30. doi: 10.1007/s11606-019-05035-1

155. Terry AL, Kueper JK, Beleno R, Brown JB, Cejic S, Dang J, et al. Is primary health care ready for artificial intelligence? What do primary health care stakeholders say? *BMC Med Inform Decis Mak*. (2022) 22:237. doi: 10.1186/s12911-022-01984-6

156. Amit G, Girshovitz I, Marcus K, Zhang Y, Pathak J, Bar V, et al. Estimation of postpartum depression risk from electronic health records using machine learning. *BMC Pregnancy Childbirth*. (2021) 21:630. doi: 10.1186/s12884-021-04087-8

157. Black JE, Kueper JK, Terry AL, Lizotte DJ. Development of a prognostic prediction model to estimate the risk of multiple chronic diseases: constructing a copula-based model using Canadian primary care electronic medical record data. *Int J Popul Data Sci*. (2021) 6:1395. doi: 10.23889/ijpds.v5i1.1395

158. Pearce C, McLeod A, Rinehart N, Whyte R, Deveny E, Shearer M. Artificial intelligence and the clinical world: a view from the front line. *Med J Aust*. (2019) 210:S38–s40. doi: 10.5694/mja2.50025

159. Wingrove P, Liaw W, Weiss J, Petterson S, Maier J, Bazemore A. Using machine learning to predict primary care and advance workforce research. *Ann Fam Med*. (2020) 18:334–40. doi: 10.1370/afm.2550

160. MacRae J, Darlow B, McBain L, Jones O, Stubbe M, Turner N, et al. Accessing primary care big data: the development of a software algorithm to explore the rich content of consultation records. *BMJ Open*. (2015) 5:e008160. doi: 10.1136/bmjopen-2015-008160

161. Turner NM, MacRae J, Nowlan ML, McBain L, Stubbe MH, Dowell A. Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study using a natural language processing software inference algorithm. *BMJ Open*. (2018) 8:e021241. doi: 10.1136/bmjopen-2017-021241

162. Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A. Artificial intelligence-based tools to control healthcare associated infections: a systematic review of the literature. *J Infect Public Health*. (2020) 13:1061–77. doi: 10.1016/j.jiph.2020.06.006

163. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. *Artif Intell Med*. (2009) 46:5–17. doi: 10.1016/j.artmed.2008.07.017

164. Morgenstern JD, Rosella LC, Daley MJ, Goel V, Schünemann HJ, Piggott T. AI's gonna have an impact on everything in society, so it has to have an impact on public health: a fundamental qualitative descriptive study of the implications of artificial intelligence for public health. *BMC Public Health*. (2021) 21:40. doi: 10.1186/s12889-020-10030-x

165. Stalgaitis CA, Jordan JW, Djakaria M, Saggese DJ, Bruce HR. Psychographic segmentation to identify higher-risk teen peer crowds for health communications: validation of Virginia's Mindset Lens Survey. *Front Public Health*. (2022) 10:864. doi: 10.3389/fpubh.2022.871864

166. Hasikin K, Lai KW, Satapathy SC, Sabanci K, Aslan MF. Editorial: emerging applications of text analytics and natural language processing in healthcare. *Front Digital Health*. (2023) 5:948. doi: 10.3389/fdgth.2023.1227948

167. Lim S, Schmälzle R. Artificial intelligence for health message generation: an empirical study using a large language model (LLM) and prompt engineering. *Front Commun*. (2023) 8:82. doi: 10.3389/fcomm.2023.1129082

168. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. *Patient Educ Couns*. (2013) 92:160–6. doi: 10.1016/j.pec.2013.05.007

169. Green N, Rubinelli S, Scott D, Visser A. Health communication meets artificial intelligence. *Patient Educ Couns*. (2013) 92:139–41. doi: 10.1016/j.pec.2013.06.013

170. The Lancet. Can artificial intelligence improve cancer care? *Lancet Oncol*. (2023) 24:577. doi: 10.1016/S1470-2045(23)00240-1

171. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front Surgery*. (2022) 9:322. doi: 10.3389/fsurg.2022.862322

172. Europarleuropaeu. Report on the proposal for a regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation). (2023) Available at: https://www.europarl.europa.eu/doceo/document/A-7-2013-0402_EN.html.

173. Spoladore D, Sacco M, Trombetta A. A review of domain ontologies for disability representation. *Expert Syst Appl*. (2023) 228:120467. doi: 10.1016/j.eswa.2023.120467

174. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inform Fusion*. (2020) 58:82–15. doi: 10.1016/j.inffus.2019.12.012

frontiers | Frontiers in Public Health

TYPE Brief Research Report
PUBLISHED 29 November 2023
DOI 10.3389/fpubh.2023.1276211

Check for updates

OPEN ACCESS

EDITED BY
Natalia Kryvinska,
Comenius University, Slovakia

REVIEWED BY
Ana Afonso,
NOVA University of Lisbon, Portugal
Dávid Líška,
Matej Bel University, Slovakia

*CORRESPONDENCE
Oleksii Honchar
ov.honchar@knmu.edu.ua

RECEIVED 11 August 2023
ACCEPTED 25 October 2023
PUBLISHED 29 November 2023

CITATION
Honchar O, Ashcheulova T, Chumachenko T, Chumachenko D, Bobeiko A, Blazhko V, Khodosh E, Matiash N, Ambrosova T, Herasymchuk N, Kochubiei O and Smyrnova V (2023) A prognostic model and pre-discharge predictors of post-COVID-19 syndrome after hospitalization for SARS-CoV-2 infection.
Front. Public Health 11:1276211.
doi: 10.3389/fpubh.2023.1276211

A prognostic model and pre-discharge predictors of post-COVID-19 syndrome after hospitalization for SARS-CoV-2 infection

Oleksii Honchar[1]*, Tetiana Ashcheulova[1], Tetyana Chumachenko[2], Dmytro Chumachenko[3], Alla Bobeiko[4], Viktor Blazhko[4], Eduard Khodosh[4], Nataliia Matiash[4], Tetiana Ambrosova[1], Nina Herasymchuk[1], Oksana Kochubiei[1] and Viktoriia Smyrnova[1]

[1]Department of Propedeutics of Internal Medicine No.1, Fundamentals of Bioethics and Biosafety, Kharkiv National Medical University, Kharkiv, Ukraine, [2]Department of Epidemiology, Kharkiv National Medical University, Kharkiv, Ukraine, [3]Department of Mathematical Modelling and Artificial Intelligence, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine, [4]Department of Pulmonology, MNE "Clinical City Hospital No.13" of Kharkiv City Council, Kharkiv, Ukraine

Background: Post-COVID-19 syndrome (PCS) has been increasingly recognized as an emerging problem: 50% of patients report ongoing symptoms 1 year after acute infection, with most typical manifestations (fatigue, dyspnea, psychiatric and neurological symptoms) having potentially debilitating effect. Early identification of high-risk candidates for PCS development would facilitate the optimal use of resources directed to rehabilitation of COVID-19 convalescents.

Objective: To study the in-hospital clinical characteristics of COVID-19 survivors presenting with self-reported PCS at 3 months and to identify the early predictors of its development.

Methods: 221 hospitalized COVID-19 patients underwent symptoms assessment, 6-min walk test, and echocardiography pre-discharge and at 1 month; presence of PCS was assessed 3 months after discharge. Unsupervised machine learning was used to build a SANN-based binary classification model of PCS development.

Results: PCS at 3 months has been detected in 75% patients. Higher symptoms level in the PCS group was not associated with worse physical functional recovery or significant echocardiographic changes. Despite identification of a set of pre-discharge predictors, inclusion of parameters obtained at 1 month proved necessary to obtain a high accuracy model of PCS development, with inputs list including age, sex, in-hospital levels of CRP, eGFR and need for oxygen supplementation, and level of post-exertional symptoms at 1 month after discharge (fatigue and dyspnea in 6MWT and MRC Dyspnea score).

Conclusion: Hospitalized COVID-19 survivors at 3 months were characterized by 75% prevalence of PCS, the development of which could be predicted with an 89% accuracy using the derived neural network-based classification model.

KEYWORDS
COVID-19, hospitalization, convalescence, post-acute COVID-19 syndrome, walk test, echocardiography, clinical decision rules, machine learning

1 Background

Since 2019–2020, the novel SARS-CoV-2 infection has emerged as a global health challenge. During the first year of COVID-19 pandemic, the cumulative effect of overwhelming of healthcare systems, lack of effective etiological treatment, and more severe course of disease that was typical to early SARS-CoV-2 variants had led to high morbidity and significant excess mortality (1), and the cascade of prolonged lockdowns that were introduced in an attempt to slow down the epidemic process imposed an additional strain on economies worldwide (2). Subsequent success of the global vaccination campaign has dramatically reduced the prevalence of severe and life-threatening cases whilst imposing less significant initial effect on the morbidity (3). Moreover, gradual lifting of lockdowns contributed to multifocal re-acceleration of the COVID-19 spread, with the daily new cases count in January 2022 exceeding the parameters during previous peaks more than three-fold (4). As a result, the current phase of lower morbidity that we observe since the beginning of 2023 is being continuously accompanied by an increasing number of patients suffering from long lasting sequelae of SARS-CoV-2 infection.

Unlike in most other respiratory viral infections, persistence of symptoms after the end of the acute phase is highly prevalent in COVID-19 convalescents (5, 6) and is shown to be associated with a decreased quality of life (7). Potential pathogenetic mechanisms are heterogenous and include consequences of direct organs injury during the acute phase, persistent SARS-CoV-2 replication, dysautonomia, altered immune reactivity, coagulopathy, endotheliopathy, and gastrointestinal microbiome disturbances (8, 9). "Long COVID," the most popular term reserved for this scenario, accounts for all manifestations that persist beyond 4 weeks after disease onset and, therefore, includes the early post-acute period of 4 to 12 weeks (also referred to as "ongoing symptomatic COVID-19"), when one expects to observe good natural dynamics of residual symptoms resolving in a significant proportion of patients (10). At the same time, development of the "post-COVID-19 syndrome," being defined as persistence of symptoms beyond 12 weeks, presents a more significant socio-economic problem, given the debilitating effect of some of the most prevalent symptoms, such as fatigue, dyspnea, apathy, and cognitive dysfunction (6, 10, 11).

The prevalence of post-COVID-19 syndrome (PCS) ranges from 5 to 37% in the general population of convalescents but may reach as high as 76 to 81% in those who required hospitalization (6, 12). To date, cardiopulmonary rehabilitation remains the mainstay of management of these patients, with locally available resources being at times insufficient to deal with the total flow of new convalescents (13–15). In this setting, creation of tools to assess the risk of subsequent development of PCS basing on pre-discharge data might be used to optimize the selection of candidates for supervised rehabilitation programs following hospitalization.

1.1 Objective

To study the in-hospital clinical characteristics of COVID-19 survivors presenting with self-reported post-COVID-19 syndrome at 3 months and to identify the early predictors of its development.

2 Materials and methods

2.1 Study design and population

By design, this is a cross-sectional prospective observational single-center study that was performed at Kharkiv City Hospital No.13 that at the time served as a specialized tertiary COVID-19 care center. Between January and November 2021, patients that were hospitalized with the diagnosis of pneumonia were evaluated for eligibility criteria that included the age of ≥18 years and positive polymerase chain reaction test for SARS-CoV-2. In total, 265 consecutive eligible patients were identified and invited to participate in the study; 44 of those have declined participation, and 221 were enrolled. The final cohort included 166 patients who have completed the repeated visit at 3 months post-discharge – see Supplementary Figure S1 for the study flowchart.

The study was conducted in compliance with the standards of Helsinki Declaration and was approved by the ethical committee of Kharkiv National Medical University (No. 3/2021). All participants provided written informed consent.

2.2 Clinical data collection

The first visit was performed pre-discharge, in clinically stable patients who were meeting clinical criteria of epidemic safety [resting capillary blood oxygen saturation (SpO2) >93% on room air, absence of acute respiratory symptoms and normal body temperature for ≥3 days starting from the 10th day after the onset of disease (16)] and included the analysis of medical records to collect data on patients' age, gender, laboratory and instrumental findings, and treatment used; an interview to obtain data on symptoms and medical history; and anthropometry.

Transthoracic echocardiography (TTE) was performed by an expert physician who was blinded to patients' data prior to procedure, using Radmir ULTIMA ultrasound system (Radmir Co., Ukraine). The measurements were performed in strict accordance with the respective guidelines by the American Society of Echocardiography (ASE) and European Association of Cardiovascular Imaging (EACVI) (17–19) and included cardiac chambers morphometry (left atrial and left ventricular [LV] volumes, right atrial area, right ventricular size, relative walls thickness and myocardial mass index of the LV), assessment of the indices of systolic function (LV ejection fraction using Simpson biplane method, mitral and tricuspid annular planes systolic excursion, global longitudinal strain [GLS] of ventricles using the linear method (20–22), and mitral annular s' velocity in the tissue Doppler mode) and LV diastolic function (mitral e' velocity and E/e' ratio).

6-min walk test (6MWT) was performed by a physician using the standard methodology as recommended by the American Thoracic Society guidelines (23), in a self-paced mode, with no use of practice tests, warm-up or non-standardized encouragement. A 20-m hallway was used, determining the selection of models to calculate the individual predicted distance (24). Pulse rate and SpO2 were registered at start and every 30 s thereafter using a Bluetooth-connected pulse oximeter; levels of fatigue and dyspnea were assessed at the baseline and at finish using modified Borg scale. Along with 6-min walk distance (6MWD), reached percent of predicted distance (6MWD%), baseline and minimal capillary oxygen saturation (SpO2base and SpO2min), baseline and maximal heart rate (HRbase and HRmax) were analyzed. Peak oxygen desaturation was calculated as SpO2drop = SpO2min - SpO2base, and reached percent of the predicted maximal heart rate as HRmax% = 100% × HRmax / (208–0.8 × Age) (25).

A follow-up visit at 1 month after discharge followed the same protocol, and the final visit at 3 months was performed distantly (by means of telephone, email, and text messengers) and included re-assessment of symptoms and detection of the PCS which was defined as a self-reported perceived worsening of health status compared to the pre-COVID-19 state or persistence of new symptoms during the last month before the visit.

2.3 Statistical analysis

The data analysis was performed using StatSoft STATISTICA Version 12 software package. Shapiro–Wilk test was used to assess the distribution of data. Continuous variables are reported as mean ± standard deviation (SD) in case of normal distribution and as median [interquartile range] in case of skewed distribution. Categorical variables are reported as counts (percentages). Independent samples t-test was used to compare normally distributed continuous variables, and paired samples t-test was used for longitudinal comparisons. For skewed variables, the comparisons were made using Mann–Whitney U-test or Wilcoxon signed-rank test. Binary and categorical variables were compared using Chi-Square test. The differences were considered significant if $p < 0.05$. Marginal effects of potential PCS predictors in logistic regression analysis were used as a measure to select inputs for unsupervised machine learning based training of simple artificial neural networks (SANN). Random sampling was used to select training, test and validation subsets of the study cohort in the 70:15:15 proportion. For each set of input variables, 500 binary classification models were trained. This number represented an empirical balance between computational resource allocation and model reproducibility – fluctuating prediction accuracy was observed with fewer models, whereas the consistency in results was achieved at n = 500; further increasing the model count did not yield significant improvements but incurred greater computational time. Automated neural architecture search strategy and Broyden-Fletcher-Goldfarb-Shanno optimization algorithm were used; missing data was imputed by mean values. Predictive performance of the obtained models was assessed as percentage of correctly classified cases from the test and validation subsets. For the final predictive model, 10-fold cross-validation was used to ensure its reproducibility, and ROC analysis performed. A post-hoc approach incorporating assessment of the model accuracy and the dataset effect size using Cohen's d statistic was used to evaluate the sample size adequacy (26).

3 Results

3.1 Baseline characteristics

The final study cohort included 76 male and 90 female patients at the mean age of 53.7 ± 13.3 years. Visit 1 was performed at the median of 22 days, visit 2 at 53 days, and visit 3 at 116 days after the symptoms onset. Among the observed cohort, 124 (75%) were reporting ongoing new symptoms and/or self-estimated worsening of health status that was classified as post-COVID-19 syndrome. The comparative clinical characteristic of study participants with and without ongoing symptoms at 3 months is presented in Table 1. The patients with PCS were older, more frequently female, had higher BMI and comorbidities burden. In their in-hospital laboratory profile, patients with ongoing symptoms at 3 months had higher values of C-reactive protein, erythrocyte sedimentation rate, and higher proportion of patients with very high interleukin-6 values, which attested to higher inflammatory activity. The observed lower values in estimated glomerular filtration rate were explained by age and gender differences. There were no differences in received treatment, but PCS patients more frequently required oxygen support during hospitalization.

3.2 Physical performance assessment

Analysis of the 6MWT parameters (see Table 2) has revealed that the apparent decrease of 6-min walk distance (6MWD) that was observed pre-discharge in PCS patients was explained by age and sex differences, evidenced by close values of the reached percent of individually predicted distance (6MWD%). Moreover, the PCS group has paradoxically demonstrated larger between-visits increment of both parameters, resulting in higher 6MWD% values at 1 month. At the same time, the level of subjective symptoms in these patients was significantly higher at visit 2, being explained by much worse improvement from the pre-discharge baseline compared to non-PCS participants. This difference was not explained by the observed values of capillary oxygen saturation throughout the test (there was no difference between groups), nor could it be attributed to worse dysautonomia – despite initially lower heart rate (HR) increment during the test at visit 1, the PCS patients have demonstrated better dynamics of utilization of HR reserve, which resulted in higher reached percent of the individual HR maximum at 1 month. MRC dyspnea scale score at both visits was also higher in PCS participants with the mean values of 2.5 ± 1.1 vs. 2.0 ± 1.0 pre-discharge ($p = 0.005$) and 1.8 ± 0.8 vs. 1.3 ± 0.6 at visit 2 ($p = 0.002$).

3.3 Echocardiographic assessment

Retrospective assessment of echocardiographic features of observed patients has only revealed minor differences between PCS and non-PCS study participants (see Table 3). Both groups were showing a strong trend to concentric LV remodeling [refer to (27) for the detailed comparison to matched control], and patients who were subsequently reporting long-lasting symptoms had somewhat smaller ventricular cavities. Despite this fact, systolic atrioventricular annuli excursion was comparable to PCS-free patients, translating into somewhat higher longitudinal strain values (reaching statistical significance in case of RV). The borderline changes in diastolic LV parameters became more apparent at visit 2, with significantly lower e' velocity (9.4 ± 2.6 vs. 10.7 ± 3.1 cm/s, $p = 0{,}025$) and higher E/e' ratio (7.6 ± 2.4 vs. 6.4 ± 1,6, p = 0,020) being observed in the PCS group; the only other difference that persisted at 1 month was the slightly higher RV free wall strain values in the PCS cohort.

TABLE 1 Demographics and pre-discharge clinical characteristics of the study participants with and without post-COVID-19 syndrome.

Parameters	No post-COVID-19 syndrome	Post-COVID-19 syndrome	Difference (95% CI)	2-sided *p*
Subjects	42	124		
Female sex	16 (38)	74 (60)		0,015
Age, years	48,7 ± 17,0	55,4 ± 12,1	6,7 (1,9; 11,5)	0,006
BMI, kg/m^2	27,3 ± 4,9	29,6 ± 5,2	2,3 (0,5; 4,1)	0,012
Comorbidities				
Hypertension	15 (36)	50 (40)		0,597
Obesity	10 (24)	51 (41)		0,044
Diabetes mellitus, type 2	2 (5)	15 (12)		0,175
History of peptic ulcer	0 (0)	12 (10)		0,036
History of cancer	0 (0)	7 (6)		0,116
History of stroke / TIA	0 (0)	6 (5)		0,147
Chronic kidney disease	0 (0)	4 (3)		0,239
Bronchial asthma	0 (0)	4 (3)		0,239
COPD	1 (2)	2 (2)		0,747
Angina pectoris	0 (0)	3 (2)		0,309
Chronic liver disease	0 (0)	2 (2)		0,408
Charlson comorbidity index	0,19 ± 0,40	0,57 ± 0,82		0,005
Active smoking status	7 (17)	16 (13)		0,542
Pulmonary involvement by CT, %	37,8 ± 25,5	30,6 ± 18,7	−7,2 (−17,0; 2,6)	0,147
Minimal in-hospital SpO2, %	90,0 ± 6,4	87,6 ± 7,9	−2,4 (−5,1; 0,2)	0,072
Oxygen supplementation				
Via nasal cannula	16 (38)	77 (62)		0,007
Noninvasive/invasive ventilation	4 (10)	5 (4)		0,174
Laboratory parameters				
Peak IL-6, pg./mL	8,6 [3,3; 11,7]	11,7 [3,0; 47,0]	13,4 (0,2; 26,7)	0,458
Peak CRP, mg/L	11 [6; 27]	25 [12; 74]	27,7 (0,0; 55,3)	0,007
Peak ESR, mm/h	26,8 ± 10,6	32,2 ± 13,0	5,4 (0,2; 10,5)	0,044
Peak procalcitonin, ng/mL	0,06 [0,05; 0,11]	0,06 [0,04; 0,12]	0,3 (−0,7; 1,2)	0,750
Peak D-dimer, ng/mL	323 [199; 432]	279 [156; 508]	−18 (−253; 217)	0,524
Peak creatinine, μmol/L	99,0 ± 21,0	104,6 ± 23,0	5,5 (−4,7; 15,8)	0,285
Lowest eGFR, ml/min/1,73m^2	76,9 ± 25,1	60,7 ± 13,7	−16,2 (−24,2; 8,2)	< 0,001
Hemoglobin, g/dL	14,1 ± 1,7	13,8 ± 1,5	−3,1 (−9,4; 3,3)	0,342
Treatment				
Dexamethasone	40 (95)	107 (86)		0,115
Methylprednisolone	28 (67)	83 (67)		0,975
Remdesivir	21 (50)	53 (43)		0,413

BMI, body mass index; TIA, transient ischemic attack; COPD, chronic obstructive pulmonary disease; CT, computed tomography; SpO2, peripheral capillary oxygen saturation; IL-6, interleukin 6; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; eGFR, estimated glomerular filtration rate by CKD-EPI equation.

3.4 Prediction of post-COVID-19 syndrome development

To create a tool that would predict the PCS development, unsupervised machine learning was used to train SANN-based binary classification models. In order to avoid overfitting effect in the setting of a relatively small sample size, we applied a two-step approach, where the list of potential inputs was first narrowed down using the assessment of their marginal effects in the logistic regression analysis. In the order of decreasing prognostic value, significant pre-discharge predictors included pre-discharge 6MWD, eGFR, heart rate increment during 6MWT, SBP, RV free wall strain, ESR, LV end-diastolic index, BMI, in-hospital oxygen supplementation, Charlson's comorbidity index, height, sex, age, and obesity (see Supplementary tables 1–2 and Supplementary figure 2 for exact Somers' D values and the resulting regression model parameters). As a next step, identified parameters were used as inputs for SANN-based classification model.

The use of the complete set of significant pre-discharge predictors, however, did not result in generation of high-performance models: the best predictive accuracy in the test/validation subset was 83%, which (given the 75% prevalence of PCS) meant that only 1/3 of otherwise false-positive cases could be correctly reclassified compared to the blunt assumption that all participants will develop the PCS. Addition

TABLE 2 One-month post-discharge dynamics of 6-min walk test parameters in patients with and without post-COVID-19 syndrome.

Parameters	No post-COVID-19 syndrome	Post-COVID-19 syndrome	Difference (95% CI)	2-sided *p*
Distance:				
6MWD at Visit 1, m	444 ± 57	380 ± 64	−64 (−90; −38)	< 0,001
6MWD% at Visit 1, %	62,5 ± 9,8	62,5 ± 9,6	0,0 (−4,1; 4,0)	0,986
6MWD at Visit 2, m	490 ± 64	454 ± 63	−36 (−65; −8)	0,012
6MWD% at Visit 2, %	68,2 ± 11,6	75,1 ± 9,5	6,9 (2,4; 11,3)	0,003
Delta 6MWD between visits, m	49 ± 33	77 ± 44	28 (9; 46)	0,004
Delta 6MWD% between visits, %	7,3 ± 5,0	13,4 ± 7,1	6,1 (3,0; 9,1)	< 0,001
Heart rate:				
HRbase at Visit 1, bpm	81,1 ± 11,0	83,3 ± 13,2	2,2 (−2,8; 7,2)	0,388
HRmax at Visit 1, bpm	109,0 ± 9,2	105,0 ± 15,6	−4,0 (−10,4; 2,4)	0,216
HRmax% at Visit 1, %	61,2 ± 6,0	60,4 ± 12,5	−0,8 (−5,7; 4,0)	0,734
HRrise at Visit 1, bpm	26,7 ± 9,7	21,5 ± 11,5	−5,2 (−10,1; −0,3)	0,039
HRbase at Visit 2, min^{-1}	78,6 ± 15,7	82,2 ± 13,5	3,6 (−2,5; 9,8)	0,239
HRmax at Visit 2, min^{-1}	108,2 ± 11,3	110,5 ± 14,6	2,4 (−3,8; 8,6)	0,451
HRmax% at Visit 2, %	60,7 ± 5,8	65,3 ± 8,3	4,6 (1,1; 8,1)	0,010
HRrise at Visit 2, bpm	29,1 ± 9,9	27,7 ± 12,1	−1,4 (−6,6; 3,9)	0,608
Oxygen saturation:				
SpO2base at Visit 1, %	95,2 ± 5,6	95,1 ± 5,2	−0,1 (−2,0; 1,8)	0,887
SpO2min at Visit 1, %	94,7 ± 3,9	93,8 ± 4,2	−0,9 (−2,6; 0,8)	0,293
SpO2drop at Visit 1, %	2,5 ± 2,5	3,1 ± 2,7	0,6 (−0,5; 1,7)	0,317
SpO2base at Visit 2, %	98,2 ± 0,6	98,0 ± 0,8	−0,2 (−0,6; 0,1)	0,155
SpO2min at Visit 2, %	95,7 ± 2,8	95,8 ± 2,1	−0,1 (−0,9; 1,1)	0,837
SpO2drop at Visit 2, %	2,5 ± 2,8	2,2 ± 1,9	−0,3 (−1,3; 0,6)	0,520
Symptoms*:				
Dyspnea at start, Visit 1, pts.	1,1 ± 1,3	1,4 ± 1,6	0,3 (−0,3; 1,0)	0,346
Fatigue at start, Visit 1, pts.	2,1 ± 1,6	2,5 ± 2,3	0,4 (−0,5; 1,3)	0,379
Dyspnea at finish, Visit 1, pts.	2,6 ± 1,8	3,8 ± 1,9	1,2 (0,1; 2,3)	0,027
Fatigue at finish, Visit 1, pts.	3,1 ± 2,4	3,5 ± 1,8	0,4 (−0,7; 1,4)	0,482
Dyspnea at start, Visit 2, pts.	0,4 ± 0,8	0,9 ± 1,1	0,5 (0,1; 1,0)	0,022
Fatigue at start, Visit 2, pts.	0,8 ± 1,2	1,5 ± 1,3	0,7 (0,1; 1,3)	0,014
Dyspnea at finish, Visit 2, pts.	1,3 ± 1,0	3,2 ± 1,7	2,0 (1,1; 2,8)	< 0,001
Fatigue at finish, Visit 2, pts	0,9 ± 1,0	2,9 ± 1,9	2,0 (1,1; 2,9)	< 0,001

*Assessed using modified Borg scale; CI, confidence interval; 6MWD, 6-min walk distance; 6MWD%, reached % of predicted 6-min walk distance; HRbase, baseline heart rate; HRmax, maximal reached heart rate; HRmax%, reached percent of the individual maximum heart rate; SpO2base, baseline oxygen saturation; SpO2min, minimal oxygen saturation; SpO2drop, peak oxygen desaturation.

of the parameters obtained during the first follow-up visit that was performed 1 month after discharge (see Supplementary Table 3 for marginal effects in logistic regression analysis) with subsequent stepwise deletion of excessive variables based on the results of global sensitivity analysis of the current version of the model have resulted in creation of the model that was characterized by a 93% predictive performance in the training and 89% in the randomly selected test/validation subsets of the study group (91% for sensitivity, 82% for specificity). The model utilized 13–7-2 SANN architecture and used age, sex, in-hospital levels of CRP, eGFR and need for oxygen supplementation, and level of post-exertional symptoms at 1 month after discharge (fatigue and dyspnea in 6MWT and MRC Dyspnea score) as inputs (Figure 1; see Supplementary Table S4 for network weights and connections; source file available at https://doi.org/10.5281/zenodo.8395451). 10-fold cross-validation that was performed by re-training the neural networks using alternative sampling has allowed to consistently obtain non-inferior accuracy in the test/validation subsets.

4 Discussion

Despite the end of the COVID-19-related public health emergency that was declared by the World Health organization on May 5, 2023, it remains an ongoing global health issue (28). Moreover, asynchronous trends of its spread and non-uniform levels of vaccination worldwide contribute to the possibility of persistence of epidemic process in separate areas, being associated with a potential for selection of new SARS-CoV-2 variants. Another issue of growing concern is the frequent transition of symptomatic COVID-19 to the chronic phase. Moreover, the recent large meta-analysis shows that the unusually high incidence of post-acute sequelae as assessed at 1 to 3 month after onset of disease (6, 12) is translating into a long-lasting trail of impaired health status: 50% of COVID-19 survivors continue to report at least one new symptom 1 year after acute infection, with the most typical manifestations (such as fatigue, dyspnea, psychiatric symptoms, cognitive deficit, memory impairment) having a potential to impose a long-term debilitating effect (29). Thus, despite having

10.3389/fpubh.2023.1276211

TABLE 3 Pre-discharge echocardiographic characteristic of hospitalized COVID-19 patients.

Parameters	No post-COVID-19 syndrome	Post-COVID-19 syndrome	Difference (95% CI)	2-sided *p*
Left chambers morphometry				
LA volume index, ml/m²	28,7 ± 7,7	28,7 ± 6,3	0,1 (−2,3; 2,4)	0,961
LV end-diastolic volume index, ml/m²	51,2 ± 8,2	47,5 ± 9,2	−3,8 (−6,9; −0,6)	0,020
LV end-systolic volume index, ml/m²	17,8 ± 5,0	16,4 ± 4,7	−1,4 (−3,1; 0,2)	0,093
LV relative wall thickness	0,44 ± 0,07	0,46 ± 0,08	0,01 (−0,01; 0,04)	0,298
LV mass index (height^{2,7}), g/m^{2,7}	38,5 ± 10,7	37,8 ± 8,6	−0,7 (−3,9; 2,6)	0,674
Left ventricular function				
LV ejection fraction, %	65,3 ± 7,5	65,5 ± 6,6	0,2 (−2,3; 2,6)	0,900
Lateral MAPSE, mm	15,2 ± 2,8	15,2 ± 2,2	0,0 (−0,1; 0,1)	0,961
LV global longitudinal strain, %	16,9 ± 2,3	17,6 ± 2,4	0,7 (−0,2; 1,6)	0,137
LV s', cm/s	9,8 ± 1,6	9,7 ± 1,8	−0,1 (−0,7; 0,5)	0,787
LV e', cm/s	9,7 ± 2,2	9,0 ± 2,2	−0,7 (−1,5; 0,1)	0,070*
LV E/e' ratio	7,4 ± 2,0	7,5 ± 1,7	0,1 (−0,5; 0,7)	0,763
Right chambers evaluation				
RA area index, cm²/m²	8,1 ± 1,1	8,2 ± 2,4	0,0 (−0,8; 0,9)	0,908
RV size (proximal outflow tract), mm	31,2 ± 3,7	32,0 ± 3,2	0,8 (−0,5; 2,1)	0,241
RV longitudinal size, mm	70,7 ± 5,9	67,6 ± 7,4	−3,1 (−5,8; −0,5)	0,021
TAPSE, mm	24,5 ± 4,9	25,0 ± 4,1	0,4 (−1,2; 2,1)	0,604
RV free wall longitudinal strain, %	33,5 ± 5,0	37,2 ± 7,1	3,7 (0,9; 6,4)	0,009

*1-sided $p = 0{,}034$; CI, confidence interval; LA left atrium; LV, left ventricle; MAPSE, mitral annular plane systolic excursion; RA, right atrium; RV, right ventricle; TAPSE, tricuspid annular plane systolic excursion.

hopefully stabilized the acute COVID-19 problem, we are currently facing a new protracted global health challenge presented by the post-COVID-19 syndrome, and the timely identification of patients at risk of its development might be instrumental for subsequent targeted attempts to improve their expected poor trajectory.

The current study was focused on the assessment of pre- and early post-discharge characteristics of hospitalized COVID-19 survivors who presented with ongoing symptoms or a self-reported persistent decline in the general health status at 3 months after discharge. Compared to the patients who have completely recovered by this term, subjects with PCS were characterized by older age, higher BMI, were more frequently female, had higher burden of comorbidities and more intensive inflammatory response during the acute phase, as evidenced by higher values of CRP, ESR, and larger share of patients with very high IL-6 levels. Despite non-inferior physical functional status as assessed by the 6MWT, PCS patients were reporting higher levels of fatigue and dyspnea, which became more pronounced at 1 month post-discharge due to worse dynamics of improvement compared to symptom-free individuals. Considering an insignificant difference in the minimal SpO2 levels between groups, it was most likely the higher level of subjective symptoms that was driving more frequent in-hospital oxygen supplementation in the PCS cohort. Comprehensive TTE has revealed similar changes in both groups that persisted at 1 month and foremost included a high incidence of concentric LV remodeling and grade I diastolic dysfunction; PCS patients were characterized by smaller ventricular cavities pre-discharge and worse diastolic filling at 1 month but higher RV longitudinal strain throughout the period of follow-up.

Compared to the set of predictors of poor physical functional recovery at 1 month post-discharge in the same cohort of patients (30), age, ESR, eGFR, need for in-hospital oxygen supplementation, and pre-discharge 6MWD have retained their prognostic value for predicting the outcome at 3 months, whereas the extent of radiological pulmonary involvement, pre-discharge SpO2, and history of hypertension became insignificant. At the same time, we revealed a subset of additional predictors of PCS that were either irrelevant (sex, height, BMI, SBP) or had not been analyzed at 1 month (RV free wall strain, LV end-diastolic index, and Charlson's comorbidity index).

Despite the large number of studies that assessed the epidemiology of PCS, the data on its predictors differ according to different sources, and the strength of observed associations is frequently weak (29). Among the variety of potential predictors, the few that have been identified regularly in different studies include older age (31, 32), history of smoking and/or lung disease (31, 33, 34), disease severity (31, 33, 35), longer hospital stay (35, 36), higher levels of CRP (32, 37), severity of symptoms (36), grade of pulmonary CT changes (35, 38), and requirement of in-hospital and post-discharge oxygen support (34, 35, 39), which puts the results of our study in line with the available body of knowledge.

Despite the poor pre-discharge results of 6MWT and evidence of incomplete early physical functional recovery, we did not find these parameters to be independently predictive of preserved self-reported COVID-19 sequelae at 3 months when accounted for sex, age, and anthropometrical parameters, thus confirming the conclusion of Ladlow et al. (40) about the lack of association between the objective functional limitations and the level of residual symptoms in the setting

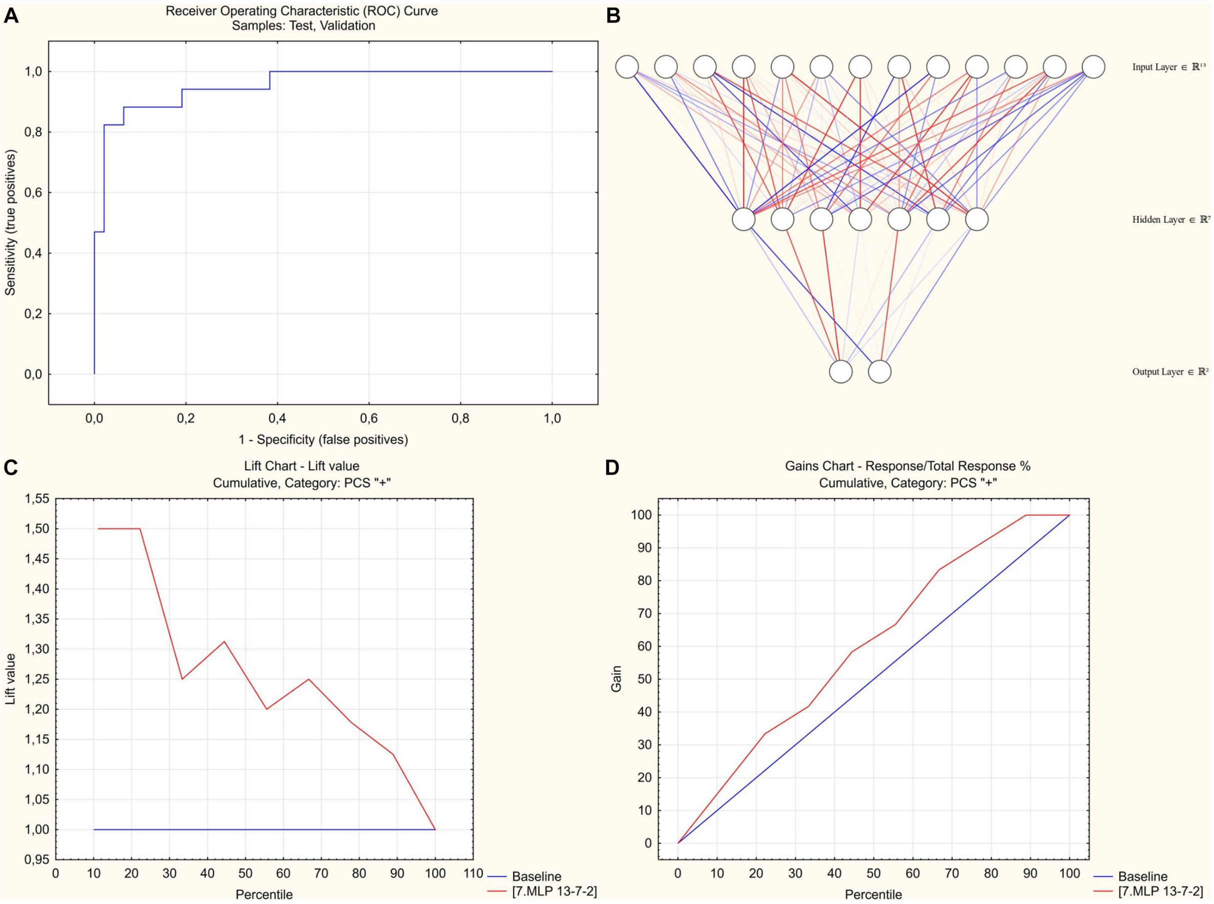


FIGURE 1
SANN-based classification model to predict post-COVID-19 syndrome at 3 months post-discharge. **(A)** Receiver operator characteristic analysis. Area under curve = 0,955. **(B)** 13–7-2 SANN architecture. **(C)** Lift chart for predicting PCS-positive cases. **(D)** Gains chart for predicting PCS-positive cases.

of PCS. We also have not found significant differences in cardiac structure and function both pre-discharge and after a one-month follow-up: patients with and without the PCS seemed to express similar changes that were in line with those reported before (41, 42).

The difficulties in predicting the development of post-COVID-19 syndrome might be related to the heterogenous underlying mechanisms, differences in vaccination status at the population level, and the constantly changing landscape of temporarily and locally prevailing SARS-CoV-2 variants. The literature search has allowed us to identify several finished studies proposing tools for predicting the outcomes of the acute phase of disease (43), effects of post COVID-19 rehabilitation (44), development of post-acute cardiovascular complications (45), and dynamics of the radiological recovery (46). Additionally, we have identified a published protocol of the study directed at solving the problem of predicting the development of the long COVID syndrome (47). To our knowledge, the current study is the first to date presenting a classification model that allows to predict the development of self-reported post-COVID-19 syndrome based on the acute phase parameters in the cohort of hospitalized patients that were shown to be at higher risk of PCS.

Logistic regression analysis is a traditional first line tool to solve the binary classification tasks; at the same time, it is vulnerable to overfitting when using smaller datasets ($n < 500$) (48). Machine learning approach in our study has allowed to overcome this obstacle; minimization of the input variables number led the training subset to meet the usual sample size requirements for ML projects with a cases-to-predictors ratio of 15.4:1 (49). Post-hoc analysis involving the input dataset effect size (Cohen's d statistic = 0,684) and the model's predictive accuracy of 89% additionally confirmed the appropriateness of the sample size (26).

4.1 Limitations of the study

The limitations of our study include possible center-related effects, potential selection bias related to inclusion of less severe cases as a result of inability of persistently O2-dependent patients to participate, negligible level of vaccination in the study population and different prevailing SARS-CoV-2 variants at the time of enrollment compared to today. Hence, the results of the study should be cautiously generalized to the current post-COVID-19 care practice. Given the ever-changing landscape of the acute SARS-CoV-2 infection setting, any types of newly developed predictive tools would require an external validation on the current local cohorts of COVID-19 convalescents as a part of their proper implementation. The proposed model should, therefore, be perceived both as a potential ready-to-use predictive tool and as a proof of concept for the development of similar models based on the more recent data from local populations.

5 Conclusion

Hospitalized patients with SARS-CoV-2 infection were characterized by a 75% prevalence of post-COVID-19 syndrome at 3 months after discharge, with PCS subjects being older, more frequently female, having higher BMI, more intensive acute inflammatory response, and lower eGFR. Higher level of symptoms in the PCS group was not associated with worse physical functional recovery or significant changes on TTE compared to symptoms-free participants. Despite identification of a set of pre-discharge predictors, inclusion of parameters obtained at 1 month proved necessary to obtain a high accuracy ML-based classification model of PCS development; the final list of inputs included age, sex, in-hospital levels of CRP, eGFR and need for oxygen supplementation, and level of post-exertional symptoms at 1 month after discharge (fatigue and dyspnea in 6MWT and MRC dyspnea score).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethical committee of Kharkiv National Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

OH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. TA: Conceptualization, Supervision, Writing – review & editing. TC: Writing – review & editing. DC: Writing – review & editing. AB: Writing – review & editing. VB: Writing – review & editing. EK: Writing – review & editing. NM: Writing – review & editing. TA: Writing – review & editing. NH: Writing – review & editing. OK: Writing – review & editing. VS: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank M. Yerzina for the technical support in data collection and to acknowledge the work of the hospital staff: O. Morozova, L. Kolisnyk, I. Talalay, V. Kozlov, L. Avdeyeva, who have set conditions for the proper functioning of the clinic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1276211/full#supplementary-material

References

1. Covid-19 Excess Mortality Collaborators. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21. *Lancet*. (2022) 399:1513–36. doi: 10.1016/S0140-6736(21)02796-3

2. Delardas O, Kechagias KS, Pontikos PN, Giannos P. Socio-economic impacts and challenges of the coronavirus pandemic (COVID-19): an updated review. *Sustainability*. (2022) 14:13. doi: 10.3390/su14159699

3. Doroftei B, Ciobica A, Ilie OD, Maftei R, Ilea C. Mini-review discussing the reliability and efficiency of COVID-19 vaccines. *Diagnostics (Basel)*. (2021) 11. doi: 10.3390/diagnostics11040579

4. Worldometers.info. (2023). Delaware, U.S.A.: Dover. Available at: https://www.worldometers.info/coronavirus/worldwide-graphs/ (Accessed March 1, 2023).

5. Astin R, Banerjee A, Baker MR, Dani M, Ford E, Hull JH, et al. Long COVID: mechanisms, risk factors and recovery. *Exp Physiol*. (2023) 108:12–27. doi: 10.1113/EP090802

6. Gyongyosi M, Alcaide P, Asselbergs FW, Brundel B, Camici GG, da Costa MP, et al. Long COVID and the cardiovascular system - elucidating causes and cellular mechanisms in order to develop targeted diagnostic and therapeutic strategies: a joint scientific statement of the ESC working groups on cellular biology of the heart and Myocardial & Pericardial Diseases. *Cardiovasc Res*. (2022) 119:336–56. doi: 10.1093/cvr/cvac115

7. Liska D, Liptakova E, Babicova A, Batalik L, Banarova PS, Dobrodenkova S. What is the quality of life in patients with long COVID compared to a healthy control group? *Front Public Health*. (2022) 10:975992. doi: 10.3389/fpubh.2022.975992

8. Proal AD, VanElzakker MB. Long COVID or post-acute sequelae of COVID-19 (PASC): an overview of biological factors that May contribute to persistent symptoms. *Front Microbiol*. (2021) 12:698169. doi: 10.3389/fmicb.2021.698169

9. Kenny G, Townsend L, Savinelli S, Mallon PWG. Long COVID: clinical characteristics, proposed pathogenesis and potential therapeutic targets. *Front Mol Biosci*. (2023) 10:1157651. doi: 10.3389/fmolb.2023.1157651

10. NICE. *COVID-19 rapid guideline: managing the long-term effects of COVID-19. COVID-19 rapid guideline: managing the long-term effects of COVID-19*. London: National Institute for health and care excellence: Clinical guidelines (2020).

11. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. (2022) 22:e102–7. doi: 10.1016/S1473-3099(21)00703-9

12. Carfi A, Bernabei R, Landi F, Gemelli Against C-P-ACSG. Persistent symptoms in patients after acute COVID-19. *JAMA*. (2020) 324:603–5. doi: 10.1001/jama.2020.12603

13. Fugazzaro S, Contri A, Esseroukh O, Kaleci S, Croci S, Massari M, et al. Rehabilitation interventions for post-acute COVID-19 syndrome: a systematic review. *Int J Environ Res Public Health*. (2022) 19:24. doi: 10.3390/ijerph19095185

14. Reinert G, Muller D, Wagner P, Martinez-Pozas O, Cuenca-Zaldivar JN, Fernandez-Carnero J, et al. Pulmonary rehabilitation in SARS-CoV-2: a systematic review and Meta-analysis of post-acute patients. *Diagnostics*. (2022) 12:20. doi: 10.3390/diagnostics12123032

15. Chen H, Shi H, Liu X, Sun T, Wu J, Liu Z. Effect of pulmonary rehabilitation for patients with post-COVID-19: a systematic review and meta-analysis. *Front Med*. (2022) 9:837420. doi: 10.3389/fmed.2022.837420

16. World Health Organization (WHO). *COVID-19 clinical management: living guidance*, (2021). WHO. Available at: https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-1 (Acessed January 25, 2021).

17. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American society of echocardiography and the European association of cardiovascular imaging. *Eur Heart J Cardiovasc Imaging*. (2015) 16:233–71. doi: 10.1093/ehjci/jev014

18. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, Dokainish H, Edvardsen T, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European association of cardiovascular imaging. *J Am Soc Echocardiogr*. (2016) 29:277–314. doi: 10.1016/j.echo.2016.01.011

19. Rudski LG, Lai WW, Afilalo J, Hua L, Handschumacher MD, Chandrasekaran K, et al. Guidelines for the echocardiographic assessment of the right heart in adults: a report from the American society of echocardiography endorsed by the european association of echocardiography, a registered branch of the European society of cardiology, and the Canadian society of echocardiography. *J Am Soc Echocardiogr*. (2010) 23:685–713. doi: 10.1016/j.echo.2010.05.010

20. Stoylen A, Molmen HE, Dalen H. Relation between mitral annular plane systolic excursion and global longitudinal strain in normal subjects: the HUNT study. *Echocardiography*. (2018) 35:603–10. doi: 10.1111/echo.13825

21. Stoylen A, Molmen HE, Dalen H. Left ventricular global strains by linear measurements in three dimensions: interrelations and relations to age, gender and body size in the HUNT study. *Open Heart*. (2019) 6:e001050. doi: 10.1136/openhrt-2019-001050

22. Aurich M, Fuchs P, Muller-Hennessen M, Uhlmann L, Niemers M, Greiner S, et al. Unidimensional longitudinal strain: a simple approach for the assessment of longitudinal myocardial deformation by echocardiography. *J Am Soc Echocardiogr*. (2018) 31:733–42. doi: 10.1016/j.echo.2017.12.010

23. Laboratories ATSCoPSfCPF. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. (2002) 166:111–7. doi: 10.1164/ajrccm.166.1.at1102

24. Duncan MJ, Mota J, Carvalho J, Nevill AM. An evaluation of prediction equations for the 6 minute walk test in healthy European adults aged 50-85 years. *PLoS One*. (2015) 10:e0139629. doi: 10.1371/journal.pone.0139629

25. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol*. (2001) 37:153–6. doi: 10.1016/S0735-1097(00)01054-8

26. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. *BMC Bioinform*. (2023) 24:48. doi: 10.1186/s12859-023-05156-9

27. Honchar, O, and Ashcheulova, T. Short-Term Echocardiographic Follow-Up After Hospitalization for COVID-19: A Focus on Early Post-Acute Changes. *Front Cardiovasc Med*. (2023) 10:1250656. doi: 10.3389/fcvm.2023.1250656

28. World Health Organization. Statement on the fifteenth meeting of the IHR (2005) emergency committee on the COVID-19 pandemic. (2023). Available at: https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic (Accessed May 20, 2023).

29. Zeng N, Zhao YM, Yan W, Li C, Lu QD, Liu L, et al. A systematic review and meta-analysis of long term physical and mental sequelae of COVID-19 pandemic: call for research priority and action. *Mol Psychiatry*. (2023) 28:423–33. doi: 10.1038/s41380-022-01614-7

30. Honchar O, Ashcheulova T. Spontaneous physical functional recovery after hospitalization for COVID-19: insights from a 1 month follow-up and a model to predict poor trajectory. *Front Med*. (2023) 10:1212678. doi: 10.3389/fmed.2023.1212678

31. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. (2021) 397:220–32. doi: 10.1016/S0140-6736(20)32656-8

32. Zhao YM, Shang YM, Song WB, Li QQ, Xie H, Xu QF, et al. Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. *EClinicalMedicine*. (2020) 25:100463. doi: 10.1016/j.eclinm.2020.100463

33. Liu D, Baumeister RF, Veilleux JC, Chen C, Liu W, Yue Y, et al. Risk factors associated with mental illness in hospital discharged patients infected with COVID-19 in Wuhan. *China Psychiatry Res*. (2020) 292:113297. doi: 10.1016/j.psychres.2020.113297

34. Bellan M, Soddu D, Balbo PE, Baricich A, Zeppegno P, Avanzi GC, et al. Respiratory and psychophysical sequelae among patients with COVID-19 four months after hospital discharge. *JAMA Netw Open*. (2021) 4:e2036142. doi: 10.1001/jamanetworkopen.2020.36142

35. Wu X, Liu X, Zhou Y, Yu H, Li R, Zhan Q, et al. 3-month, 6-month, 9-month, and 12-month respiratory outcomes in patients following COVID-19-related hospitalisation: a prospective study. *Lancet Respir Med*. (2021) 9:747–54. doi: 10.1016/S2213-2600(21)00174-0

36. Xiong Q, Xu M, Li J, Liu Y, Zhang J, Xu Y, et al. Clinical sequelae of COVID-19 survivors in Wuhan, China: a single-Centre longitudinal study. *Clin Microbiol Infect*. (2021) 27:89–95. doi: 10.1016/j.cmi.2020.09.023

37. Weng J, Li Y, Li J, Shen L, Zhu L, Liang Y, et al. Gastrointestinal sequelae 90 days after discharge for COVID-19. *Lancet Gastroenterol Hepatol*. (2021) 6:344–6. doi: 10.1016/S2468-1253(21)00076-5

38. Frija-Masson J, Debray MP, Gilbert M, Lescure FX, Travert F, Borie R, et al. Functional characteristics of patients with SARS-CoV-2 pneumonia at 30 days post-infection. *Eur Respir J*. (2020) 56:2001754. doi: 10.1183/13993003.01754-2020

39. Mandal S, Barnett J, Brill SE, Brown JS, Denneny EK, Hare SS, et al. 'Long-COVID': a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. *Thorax*. (2021) 76:396–8. doi: 10.1136/thoraxjnl-2020-215818

40. Ladlow P, O'Sullivan O, Houston A, Barker-Davies R, May S, Mills D, et al. Dysautonomia following COVID-19 is not associated with subjective limitations or symptoms but is associated with objective functional limitations. *Heart Rhythm*. (2022) 19:613–20. doi: 10.1016/j.hrthm.2021.12.005

41. Ingul CB, Grimsmo J, Mecinaj A, Trebinjac D, Berger Nossen M, Andrup S, et al. Cardiac dysfunction and arrhythmias 3 months after hospitalization for COVID-19. *J Am Heart Assoc*. (2022) 11:e023473. doi: 10.1161/JAHA.121.023473

42. Ovrebotten T, Myhre P, Grimsmo J, Mecinaj A, Trebinjac D, Nossen MB, et al. Changes in cardiac structure and function from 3 to 12 months after hospitalization for COVID-19. *Clin Cardiol*. (2022) 45:1044–52. doi: 10.1002/clc.23891

43. Rasmy L, Nigo M, Kannadath BS, Xie Z, Mao B, Patel K, et al. Recurrent neural network models (CovRNN) for predicting outcomes of patients with COVID-19 on admission to hospital: model development and validation using electronic health record data. *Lancet Digit Health*. (2022) 4:e415–25. doi: 10.1016/S2589-7500(22)00049-8

44. Adamo S, Ambrosino P, Ricciardi C, Accardo M, Mosella M, Cesarelli M, et al. A machine learning approach to predict the rehabilitation outcome in convalescent COVID-19 patients. *J Pers Med*. (2022) 12:12. doi: 10.3390/jpm12030328

45. Gupta A, Jain V, Singh A. Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications. *N Gener Comput*. (2022) 40:987–1007. doi: 10.1007/s00354-021-00144-0

46. Sonnweber T, Tymoszuk P, Sahanic S, Boehm A, Pizzini A, Luger A, et al. Investigating phenotypes of pulmonary COVID-19 recovery: a longitudinal observational prospective multicenter trial. *elife*. (2022) 11:11. doi: 10.7554/eLife.72500

47. Daines L, Mulholland RH, Vasileiou E, Hammersley V, Weatherill D, Katikireddi SV, et al. Deriving and validating a risk prediction model for long COVID-19: protocol for an observational cohort study using linked Scottish data. *BMJ Open*. (2022) 12:e059385. doi: 10.1136/bmjopen-2021-059385

48. Bujang MA, Sa'at N, Sidik T, Joo LC. Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malays J Med Sci*. (2018) 25:122–30. doi: 10.21315/mjms2018.25.4.12

49. Baeza-Delgado C, Cerda Alberich L, Carot-Sierra JM, Veiga-Canuto D, de Las M, Heras B, et al. A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data. *Eur Radiol Exp*. (2022) 6:10. doi: 10.1186/s41747-022-00276-y

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 11 December 2023
DOI 10.3389/fpubh.2023.1310016

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Sonu Rajak,
National Institute of Technology Patna, India
Anbesh Jamwal,
Malaviya National Institute of Technology Jaipur, India

*CORRESPONDENCE
Lin Lu
✉ lulin355@163.com
Xiaochun Luo
✉ 57747278@qq.com
Qianlan Chen
✉ 393565992@qq.com

RECEIVED 10 October 2023
ACCEPTED 23 November 2023
PUBLISHED 11 December 2023

CITATION
Long P, Lu L, Chen Q, Chen Y, Li C and Luo X (2023) Intelligent selection of healthcare supply chain mode – an applied research based on artificial intelligence.
Front. Public Health 11:1310016.
doi: 10.3389/fpubh.2023.1310016

Intelligent selection of healthcare supply chain mode – an applied research based on artificial intelligence

Ping Long[1], Lin Lu[1]*, Qianlan Chen[1]*, Yifan Chen[2], Chaoling Li[1] and Xiaochun Luo[1,3]*

[1]School of Economics and Management, Guangxi Normal University, Guilin, China, [2]Adam Smith Business School, University of Glasgow, Scotland, United Kingdom, [3]School of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Introduction: Due to the inefficiency and high cost of the current healthcare supply chain mode, in order to adapt to the great changes in the global economy and public health, it is urgent to choose an effective mode for sustainable development of healthcare supply chain. The aim of this paper is to use artificial intelligence systems to make intelligent decisions for healthcare supply chain mode selection.

Methods: Firstly, according to the economic benefits, social benefits and environmental benefits of healthcare supply chain, this paper identifies different healthcare supply chain modes in combination with artificial intelligence technology. Secondly, this paper presents the intelligent choice optimization method of healthcare supply chain mode based on deep reinforcement learning algorithm. Finally, the effect of artificial intelligence in healthcare supply chain mode selection is verified by simulation experiment.

Results and Discussion: The experimental results show that healthcare supply chain mode selected by artificial intelligence is basically consistent with the target mode, while healthcare supply chain mode selected by the basic selection method, BP neural network method and big data method is different from the target mode, which indicates that AI has more advantages in the selection of medical supply chain mode. Therefore, we recommend the application of artificial intelligence to healthcare supply chain management. This study not only makes up for the ineffective problems of existing methods, but also makes up for the gaps in the application of AI technology in the field of healthcare supply chain. The scientific value of this paper is that the proposed framework and the artificial intelligence algorithm enrich the relevant theories of healthcare supply chain research and provide methodological guidance for intelligent decision-making of healthcare supply chain. At the same time, for medical enterprises, this research provides a new practical guideline for the application of artificial intelligence in the sustainable development and modern management of healthcare supply chain.

KEYWORDS

healthcare supply chain, mode selection, artificial intelligence, intelligent selection, deep reinforcement learning algorithms

1 Introduction

In the post epidemic era, the industry environment and consumer demand have changed significantly, which has accelerated the digitalization of the healthcare supply chain (1, 2), the development of the modern healthcare supply chain has entered a new era, and the study of the healthcare supply chain has become a hot topic in both academia and industry. The healthcare supply chain plays a crucial role in health economics, as it directly impacts the cost, accessibility, and quality of healthcare services. Specifically, the healthcare supply chain occupies the core position in ensuring the availability and availability of pharmaceutical products and services (3). Studying the healthcare supply chain helps to ensure high-quality healthcare products and services (4) and minimize the risk of harm to patients. Second, studying the healthcare supply chain can help to improve its operational efficiency by minimizing waste, reducing costs, and optimizing resource utilization (5), this can promote the improvement of medical economic level of medical enterprises. Third, during emergencies or epidemics, the healthcare supply chain plays a key role in ensuring the availability of critical supplies (e.g., medicines, personal protective equipment, and medical devices) (6), and a resilient healthcare supply chain can respond quickly to emergencies and minimize losses due to risks. Finally, the healthcare supply chain involves information, logistics, and financial flows from suppliers to healthcare providers and ultimately to patients, which involves multiple stakeholders, including manufacturers, distributors, healthcare providers, and regulatory agencies (7). The research of the healthcare supply chain is beneficial for coordinating the interests of the supply chain actors and enhancing the competitiveness of the supply chain. Therefore, considering the importance of healthcare supply chain in ensuring timely supply of medical products and services, minimizing the risk of patient harm and minimizing waste, as well as the dominant position of healthcare supply chain mode in the supply chain operation process, this paper focuses on the intelligent selection of healthcare supply chain mode.

With the emergence of enterprise clusters, the competition mode among enterprises has begun to transition from single-enterprise competition to supply chain competition, and the competitiveness of supply chains increasingly depends on whether they can meet the requirements of sustainable development (8). The development and pattern evolution of the manufacturing industry have an important impact on the world economy (9), but at the same time, due to the production characteristics of the industry, it also becomes the biggest hidden danger to the ecological environment. The special supply chain such as healthcare supply chain involves a large number of chemical reactions in the process of research and development of pharmaceuticals, which will produce a large number of harmful pollutants such as waste water, waste gas, experimental waste liquid and other harmful pollutants that have a negative impact on the environment. The ingredients contained in these pollutants are also more complex, and if they are not disposed of in a compliant way, it will have a great negative impact on the environment. In addition, the business system of healthcare supply chain covers pharmaceutical research and development, distribution and other fields, and involves multiple links such as procurement, production, delivery, marketing and recycling in daily operation (10). Whether each link meets the "green standard" is closely related to human health. Therefore, some pharmaceutical companies have been implementing environmental protection actions through green practices such as recycling and medical waste treatment, and adopting new clean technologies such as low-pollution gas heating/cooling, electric vehicles and energy-saving policies to increase social benefits (11). Therefore, in order to meet the new industry needs, there is a need to study the optimal mode of healthcare supply chain and use modern decision-making technology tools.

At the same time, the application of AI appears in a number of different fields, including the supply chain (12). AI systems are able to make informed decisions and perform tasks automatically without human intervention. The implantation of AI systems in the supply chain is still an innovation, and AI is expected to optimize and upgrade various industries, including medical enterprises (13). As far as the healthcare supply chain is concerned, AI offers plenty of opportunities to improve operational efficiency and sustainability. Specifically, firstly, artificial intelligence algorithms help healthcare supply chain to forecast demand and manage inventory (14). Artificial intelligence can minimize waste, improve cost efficiency, and ensure timely supply of key materials. Secondly, artificial intelligence algorithms can optimize the logistics of healthcare supply chain, improve transportation efficiency, reduce fuel consumption, and minimize carbon emissions, thus promoting the sustainable development of healthcare supply chain. Finally, artificial intelligence has significant advantages in continuous improvement and learning. It can clearly and quickly identify the optimal operation mode of healthcare supply chain through the analysis of big data (15), which makes healthcare supply chain more flexible. To sum up, artificial intelligence plays a vital role in promoting the development of health economy, and the integration of artificial intelligence technology into all aspects of healthcare supply chain operation can bring significant benefits (16). This provides technical support for the healthcare supply chain to gain competitive advantages, improve operational efficiency, and become a more low-carbon and environmentally friendly medical system with sustainable development potential (17).

However, few scholars have conducted in-depth studies on the issue of healthcare supply chain mode selection. At the same time, given the great advantages of AI technology in optimizing healthcare supply chain, there are few researches on the role of AI technology in healthcare supply chain mode selection. Therefore, to fill this gap and given that deep reinforcement learning algorithms are particularly suitable for highly variable environments and applicable to the research scenarios of the article, we propose a deep reinforcement learning algorithm based on AI with high scalability to optimize healthcare supply chain mode selection. Since AI technology has a crucial role in the development of healthcare supply chains and is valuable for optimizing healthcare supply chain modes and improving competitiveness, this is the main motivation for this study. In addition, the competition of current business modes is becoming more and more intense, and enterprises need to find suitable modes to realize their sustainable development, which further illustrates the necessity of applying AI technology to the selection of healthcare supply chain modes. In summary, the article identifies the following three research objectives:

(1) Determine the indicators for evaluating healthcare supply chain modes and clarify the relevant calculation methods to support the effective selection of the optimal healthcare supply chain mode;
(2) Combined with the evaluation indexes, different healthcare supply chain modes are identified using artificial intelligence technology, and an intelligent selection optimization method

for healthcare supply chain modes is proposed based on deep reinforcement learning algorithm;

(3) Constructing simulation experiments to verify the effectiveness of the application of artificial intelligence technology.

This study has three major contributions to the field of healthcare supply chain research. First, we discuss a neglected and understudied problem in previous studies: how to choose the optimal healthcare supply chain mode? Based on the characteristics of healthcare supply chain, this paper uses artificial intelligence technology to choose healthcare supply chain mode, and puts forward a new method of healthcare supply chain mode selection, which enhances the objectivity and effectiveness of the selection process. Secondly, based on the deep reinforcement learning algorithm, this paper puts forward a specific algorithm to optimize the mode selection of healthcare supply chain, enriching the relevant theories and methods of healthcare supply chain research, and ensuring the feasibility of the results. Finally, the simulation experiment verifies the effectiveness of artificial intelligence in the application of healthcare supply chain mode selection, which provides new practical guidance and valuable insights for artificial intelligence to improve the level of health economy, sustainable development of medical supply chain and modern management.

The paper is organized as follows. Section 2 will review the literature on research on healthcare supply chain and AI applications. Section 3 describes the related research methodology and the process of intelligent selection of healthcare supply chain modes. Section 4 conducts simulation experiments for validation. Section 5 is a discussion and analysis of the results of the simulation experiments. Section 6 is the conclusion and further indicates the future research direction.

2 Literature review

2.1 Healthcare supply chain mode selection

Existing research on healthcare supply chain is very rich and many scholars have explored different aspects of healthcare supply chain in depth. For the logistics process of healthcare supply chain, Smith et al. argued that the logistics process of healthcare supply chain has become a key factor affecting the cost and quality of healthcare services, and they conducted two surveys of healthcare supply chain professional practitioners to understand the operational efficiency of the healthcare supply chain and the potential opportunities for improvement, and the article pointed out the benefits and difficulties of standardizing data in the healthcare supply chain, as well as how managers can provide support in achieving data standardization (18). Addressing the aspect of healthcare supply chain digitization and performance, it was noted that the global manufacturing industry is working hard to improve productivity by leveraging digital manufacturing technologies such as artificial intelligence and the Internet of Things (IoT), which are expected to improve the connectivity between supply chains, and this trend extends to the healthcare supply chain sector. Meanwhile, social capital, a potential resource for healthcare supply chains, can affect supply chain performance in the healthcare industry. Therefore, the article conducted an empirical study on the relationship between healthcare supply chain digitization, social capital and supply chain performance (19). For risk management in healthcare supply chain, some studies have made guidance for improving healthcare supply chain practices by revealing the relationship between supply chain integration, supply chain risk management, and supply chain 4.0 and healthcare supply chain performance, so that healthcare supply chain can better cope with risks and enhance performance (20).

It is not difficult to see that many scholars have made certain achievements in the study of healthcare supply chain, and most of the research content is closely related to the operation of healthcare supply chain, based on which, it is necessary to mention the healthcare supply chain mode, which dominates the various stages of the supply chain, and largely determines the structure of the healthcare supply chain and its operation mode. The choice of healthcare supply chain mode plays a fundamental role in the development of the supply chain, specifically, from the viewpoint of patient safety, an effective healthcare supply chain mode can help control the quality and safety of medicines (21), and filter the potential risks and loopholes that may affect the safety of patients to ensure the safety of patients. From the perspective of cost control, an appropriate healthcare supply chain mode is conducive to improving inventory management, streamlining processes, and optimizing logistics as a means to achieve cost savings. From the perspective of healthcare supply chain risk management, an efficient healthcare supply chain operation mode is conducive to making quick and effective decisions to reduce risks and ensure the continuity of healthcare supply chain operations. From the perspective of sustainable development, an excellent good healthcare supply chain mode is conducive to the reduction of carbon emissions, which improves economic benefits and ensures the growth of environmental and social benefits at the same time.

However, there is little literature studying the selection of healthcare supply chain modes. In terms of the existing researched healthcare supply chain modes, one study proposed the healthcare supply chain e-commerce mode (22), which is the combination of e-commerce and healthcare supply chain. This mode solves the drawbacks existing in the traditional healthcare supply chain, optimizes the structure of the healthcare supply chain, and improves the efficiency of the healthcare supply chain. There is also a study on the SPD mode of the healthcare supply chain (23), which refers to a medical supplies management mode integrating supply/processing/distribution links, and its core carrier is a medical consumables supply chain platform, aiming to help medical institutions realize refined operation and cost control. The SPD mode comprehensively takes into consideration the operation rules and characteristics of each management link of medical consumables in hospitals, as well as the interconnection between links. With the support of supply chain management theory and information technology, it optimizes and improves the traditional way of medical consumables management, and is a consumables management mode applicable to the current social and medical background.

In terms of the methodology of the healthcare supply chain mode selection, some scholars created a vaccine supply chain model in response to the global crisis and used a multi-objective mixed integer planning (MIP) model to develop the supply chain (24), and this study also integrated the TOPSIS (Order Preference Technique by Similarity to Ideal Solution) model into the optimization model to identify the best solution from a set of non-dominated solutions that prioritize environmental sustainability, although the algorithm used in this study can be used to design a sustainable vaccine supply chain mode, it has a narrow scope of application and the study can be expanded to utilize other hybrid meta-heuristic algorithms for the analysis of the experimental results for the purpose of comparing the findings. Other scholars have finally constructed and proposed a new model of a sustainable healthcare supply chain that prioritizes patient safety by identifying key attributes of a sustainable healthcare supply chain and then applying the Fuzzy Delphi Method (FDM) to determine the appropriateness of these key attributes and categorize them in order to finally construct and propose a new mode of a sustainable healthcare supply chain that prioritizes patient safety, which improves the safety of healthcare professionals and patients in a sustainable manner (25). Furthermore, to deal with the risks faced in the operation of the supply chain. Rajak and other scholars used the hybrid BWM-QFD methodology to identify the stakeholders' requirements and critical success factors (CSFs) for the sustainability initiative in SC during this pandemic situation (26). Because the choice of healthcare supply chain mode needs to focus on the impact on the sustainable development of medical supply chain. The framework proposed by the study and the 16 key success factors identified provide good solutions to supply chain sustainability problems. Thus, about the measurement of sustainability dimensions in healthcare supply chain mode selection, this study provides important methodological references and content guidance. Finally, an integrated DEA enhanced Russell measure (ERM) model was studied to select the best transportation system provider. This model is used for the performance evaluation of the transportation system to minimize capital cost, cost of energy consumption, cost of work safety and labor's health, and CO_2 emission (27). In terms of healthcare supply chain mode selection, the assessment of medical goods transportation, capital costs, energy consumption, exhaust emission and other indicators are particularly important. The value obtained by these indicators will directly affect the choice of healthcare supply chain mode. Therefore, this model provides strong method support for the selection of healthcare supply chain transportation system, which is important for the selection of effective healthcare supply chain mode.

2.2 Application of artificial intelligence in the healthcare supply chain

With the rapid development of a series of emerging technologies such as big data and cloud computing, artificial intelligence technology is also increasingly used in the medical field (28). As an advanced technology, artificial intelligence can use the high-speed processing and distributed computing ability of computers to simulate the information processing ability and learning ability of the human brain, so that computers can analyze and solve certain problems that cannot be handled by medical technology (29), and the feasibility of the application of AI technology in the field of healthcare has been confirmed, and the future prospects are broad (30).

Currently, the application of artificial intelligence technology in the healthcare supply chain is mainly reflected in the following five aspects. First, the application of artificial intelligence technology in pharmaceutical research and development (31), the significant changes in the field of pharmaceutical research and development are closely related to the breakthrough progress in the field of artificial intelligence (32). In the face of the extremely large pharmaceutical industry data, AI can be used to process the data through appropriate algorithmic models, and then used in the process of pharmaceutical R&D, which can effectively accelerate the speed of pharmaceutical product development, reduce the cost of R&D, and improve the success rate of research and development. Second, AI can improve the sustainability of the healthcare supply chain (33). Obviously, AI technology has gradually become a strategic tool to improve competitiveness, and the application of AI in all aspects of the supply chain guarantees the normal operation of the business, and its fast and effective decision-making and execution ability prevents the disruption of the healthcare supply chain after the impact. Third, AI technology can be used in healthcare supply chain management (34). The healthcare supply chain is a very complex network, and the use of innovative tools for healthcare supply chain management can create economic surpluses and take into account the interests of the supply chain actors. Fourth, AI techniques can be used for demand forecasting and patient health management (35). Artificial intelligence analyzes and compares the patient's indicators with the standard data by using the data analysis system and transmits them to the server's terminal, and finally gives reasonable suggestions. This technology can combine relevant data to assess personal health, establish personal health service files, and provide health control for key patients. Fifth, artificial intelligence technology can be used to assist in the diagnosis of medical imaging (36) and improve the efficiency of the healthcare supply chain. Traditional medical diagnosis is mainly obtained by relying on the doctor's analysis of the image data, mostly by virtue of personal experience as well as subjective judgment, and therefore the phenomenon of misdiagnosis will occur. Medical imaging can be divided into two parts: imaging and image analysis, and the use of artificial intelligence technology can make imaging faster, effectively shorten the imaging time and improve the clarity of the image. In the image analysis part, artificial intelligence technology can quickly extract data from the image database and analyze it to help doctors carry out auxiliary diagnosis, avoiding diagnostic errors due to doctors' subjective factors, which effectively improves the accuracy of diagnosis.

However, the application of AI in the healthcare supply chain field still has some challenges and limitations. First of all, the application scope of artificial intelligence in the field of healthcare supply chain is relatively narrow (37). Compared with other fields (such as retail industry and banking industry), the healthcare supply chain is subject to more strict legal constraints, and its data security and privacy protection are more important, which limits the application of artificial intelligence in the medical field to a certain extent. The research of artificial intelligence in helping the healthcare supply chain make intelligent decisions needs to be further explored. Second, the public health system is gradually transforming, its main focus is to realize the value chain through the use of digital technology (38), while past research has mainly discussed the scalability of artificial intelligence and some ethical issues, as well as how to realize the value chain in the medical field in the new business environment has not been clarified. More theoretical and empirical research is needed to address the challenges facing AI. Finally, the current research cognition of artificial intelligence is still limited and scattered, which may be caused by the disjoint of previous research. Some scholars only focus on the application of AI in disease monitoring and drug development (39, 40), and lack a comprehensive understanding of artificial intelligence systems, which limits the practical value of artificial intelligence. In fact, the application of artificial intelligence can benefit different subjects in the healthcare supply chain (41).

In short, with the development of the times, artificial intelligence technology has been serving more and more aspects of the healthcare supply chain. The popularity of artificial intelligence technology has greatly improved the degree of intelligent development of the medical system. However, in view of the limitations and challenges existing in the application of artificial intelligence in the field of public health, the application of artificial intelligence in the choice of healthcare supply chain mode is still a blank. Further discussion is needed.

2.3 Research gaps

To sum up, it is not difficult to see that the existing research literature on healthcare supply chain and artificial intelligence is relatively rich and the research system is relatively mature. Most scholars focus on discussing the application of artificial intelligence in healthcare supply chain from the perspectives of sustainable development, supply chain management and logistics, so they ignore the research on healthcare supply chain mode. In particular, there are few researches on the application of artificial intelligence in the selection of healthcare supply chain modes. In addition, in view of the existing healthcare supply chain mode selection methods, the significant problem is the lack of systematicness and objectivity, and it is difficult to apply such methods in practical analysis.

Healthcare supply chain mode determines the structure and operation mode of the supply chain to a large extent. Only by choosing an effective mode can healthcare supply chain have an advantageous competitive position in the highly competitive market. And it is conducive to the long-term development of the supply chain. The selection of healthcare supply chain mode needs to be interpreted from a deeper perspective, and the existing research methods and tools need to be improved and optimized. Therefore, this paper not only makes up for the gaps in the existing healthcare supply chain research content, but also makes up for the defects of research methods.

3 Research methodology

3.1 Deep reinforcement learning algorithm

Reinforcement Learning (RL), also known as evaluation learning, is an important branch of machine learning. Its feature is that there is no prior knowledge, it does not need to give the correct action, but only needs to give the return, and adjust the strategy to maximize the total cumulative return. To put it simply, reinforcement learning is that the actions generated by an Agent interact with the environment and generate sample sequences. Through continuous learning, the agent searches for actions with high returns and avoids actions with low returns. Obviously, reinforcement learning is a trial-and-error process that simulates the learning process of animals. It obtains reward and punishment information by constantly trying actions and iteratively updates its reward value, so as to find the optimal solution. Since reinforcement learning is the reward information obtained from environmental feedback in the learning process, reinforcement learning is applicable to unknown environments and has decision-making ability (42). The paradigm of reinforcement learning is very similar to the process of human learning knowledge, so reinforcement learning is regarded as an important way to achieve general AI (43), but it is helpless to solve the perception problem.

Deep learning is one of the most popular branches of machine learning in recent years, which has made great achievements in image recognition and natural language processing. The simple explanation of deep learning is that when we know an input and mapping, we can get its output, but the reality is that the mapping is very complex, cannot be expressed in an expression or simply unknown state. At this time, deep learning constructs a deep neural network as the connection bridge between sample input and feature output. We constantly input samples for training, so that the neural network can fit and approximate a mapping corresponding to these samples as much as possible, and the finally trained network can be approximately equivalent to the mapping we need. Although deep learning has strong perception ability, it lacks certain decision-making ability.

Therefore, in recent years, with the continuous improvement of deep learning theory, the combination of reinforcement learning and deep learning has received more and more attention and research, and the deep reinforcement learning algorithm has gradually become a research hotspot in the field of artificial intelligence (44). Introducing deep learning into the reinforcement learning framework and replacing the original Q-table with neural network can greatly improve the computing power of the reinforcement learning algorithm and deal with higher dimensional and more complex situations. In other words, the neural network provides the perception ability for reinforcement learning that it lacks. By combining the fitting approximation ability of deep learning with the strategy generation ability of reinforcement learning, the transformation from table form to network form is achieved, and deep reinforcement learning is formed. Choosing deep reinforcement learning algorithm for healthcare supply chain mode selection mainly takes into account the unique functions and characteristics of deep reinforcement learning. First of all, the healthcare supply chain mode selection involves a complex decision-making process, which requires the AI model to learn in a dynamic and uncertain environment. Among many AI algorithms, DRL is the most suitable for unknown environment and has strong decision-making ability. Second, in healthcare supply chain mode selection, decisions made at one stage may affect subsequent stages, DRL's ability to learn from sequential interactions enables it to effectively address such impacts. Finally, healthcare supply chain mode selection needs to balance multiple objectives, such as economic cost, patient safety and environmental sustainability. DRL can learn more complex decision spaces through interaction with the environment and identify solutions that are consistent with the organization's goals to achieve a balance between conflicting goals. In short, the key to choosing the deep reinforcement learning model over other artificial intelligence models for healthcare supply chain mode selection is that, compared with other artificial intelligence algorithms, DRL is more suitable for uncertain environments and can handle complex and continuous decision-making tasks. These characteristics make DRL very suitable for solving the multi-faceted challenges of healthcare supply chain management.

Current deep reinforcement learning algorithms mainly include two types, namely deep reinforcement learning algorithms based on value function approximation and deep reinforcement learning algorithms based on policy gradient. Among them, the most representative ones in the field of value function approximation algorithms are Deep Q network algorithm (DQN) (45) and dual Deep Q network algorithm (DDQN) (46). The most representative in the field of policy gradient algorithms is the depth deterministic policy gradient algorithm (DDPG) (47). In deep reinforcement learning, we use neural networks to represent value functions and strategies and solve their modeling problems. We continuously input samples to continuously improve the values of weight W and bias b in the network, and use Error Back Propagation (BP) methods to continuously iterate and optimize our objective functions. The network can approximate the real value function and strategy as much as possible. The DRL framework is shown in Figure 1, and its learning process can be described as follows:

(1) At each moment agent interacts with the environment to get a high-dimensional observation, and utilizes the DL method to perceive the observation to get a specific state feature representation.
(2) Evaluate the value function of each action based on the expected return and map the current state to the corresponding action by some strategy.
(3) The environment reacts to the action and obtains the next observation. By continuously cycling the above process, the optimal strategy to realize the goal can be obtained eventually.

3.2 Design of intelligent selection method of healthcare supply chain mode

In view of the current healthcare supply chain mode selection problem, there is no specific method to help healthcare institutions choose the optimal mode, and the current research ignores the potential of technology and innovation in optimizing the healthcare supply chain. Therefore, this study applied artificial intelligence technology to the intelligent choice of healthcare supply chain mode, designed a new intelligent choice method and adopted a more scientific calculation method to select the optimal healthcare supply chain mode to achieve sustainable development. In this study, the intelligent selection process of healthcare supply chain mode is set up as shown in Figure 2. This study mainly provides the basis for the selection of healthcare supply chain mode through indicators such as economic benefit, social benefit (48) and environmental benefit (49). By considering the economic, social, and environmental dimensions of each supply chain mode, healthcare companies can make informed decisions to optimize their supply chain strategies while promoting

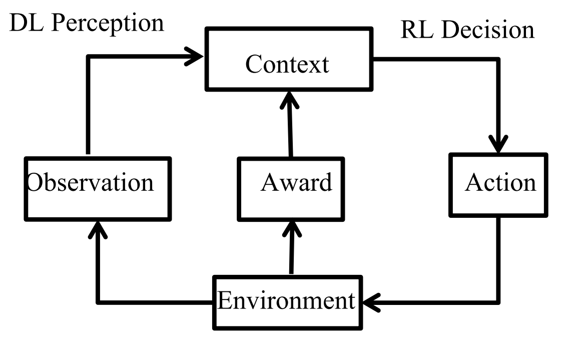


FIGURE 1
DRL schematic framework.

sustainability, ethical practices, and cost-effective operations. The selection process involves a lot of calculation parts, in order to ensure the scientific selection results, it is necessary to ensure the feasibility of the calculation algorithm.

3.2.1 Economic benefits of healthcare supply chain

The economic efficiency reflects the overall economic operation effect of the healthcare supply chain, which is the performance of the ability of capital circulation. The economic benefits of healthcare supply chain include financial value, internal operation and management level of the supply chain, and pharmaceutical R&D innovation capability.

For the financial value of the healthcare supply chain, this article selects the total asset turnover rate, the profit rate of pharmaceutical sales and the current ratio of the healthcare supply chain as the indicators for the assessment of the financial value, which fully reflects the financial operation status and profitability of the entire healthcare supply chain.

(1) Total Asset Turnover Ratio. This index mainly measures the ratio between the net income from pharmaceutical sales and the asset investment scale of healthcare supply chain in a certain period of time, which to a certain extent can reflect the operational efficiency of total assets of healthcare supply chain. The larger the value of this index, the higher the level of sales of medicines in the healthcare supply chain, the greater the return on asset investment, and the stronger the operating ability of the total assets of the supply chain.

$$\text{Total asset turnover} = \frac{\begin{array}{c}\text{Net proceeds from the}\\ \text{sale of pharmaceuticals}\end{array}}{\begin{array}{c}\text{Total healthcare}\\ \text{supply chain assets}\end{array}} \times 100\% \tag{1}$$

(2) Profitability of pharmaceutical sales. This indicator is the ratio of total profit to product sales revenue. It reflects the profitability of the product and indirectly reflects the size of the competitiveness of the enterprise's products in the market.

$$\text{Profitability of pharmaceutical sales} = \frac{\text{Sales profit}}{\text{Total sales}} \times 100\% \tag{2}$$

(3) Current ratio. The current ratio is the ratio of current assets to current liabilities of the healthcare supply chain for a certain period of time. The current ratio can reflect whether the healthcare supply chain has enough working capital for production to supply products to the demand side.

$$\text{Current ratio} = \frac{\text{Current asset}}{\text{Current liabilities}} \times 100\% \tag{3}$$

The healthcare supply chain involves many aspects such as procurement, production, logistics, sales, and reverse logistics. It is necessary to ensure the overall circulation of the healthcare supply chain as a whole in all links. Therefore, each link is crucial. Based on

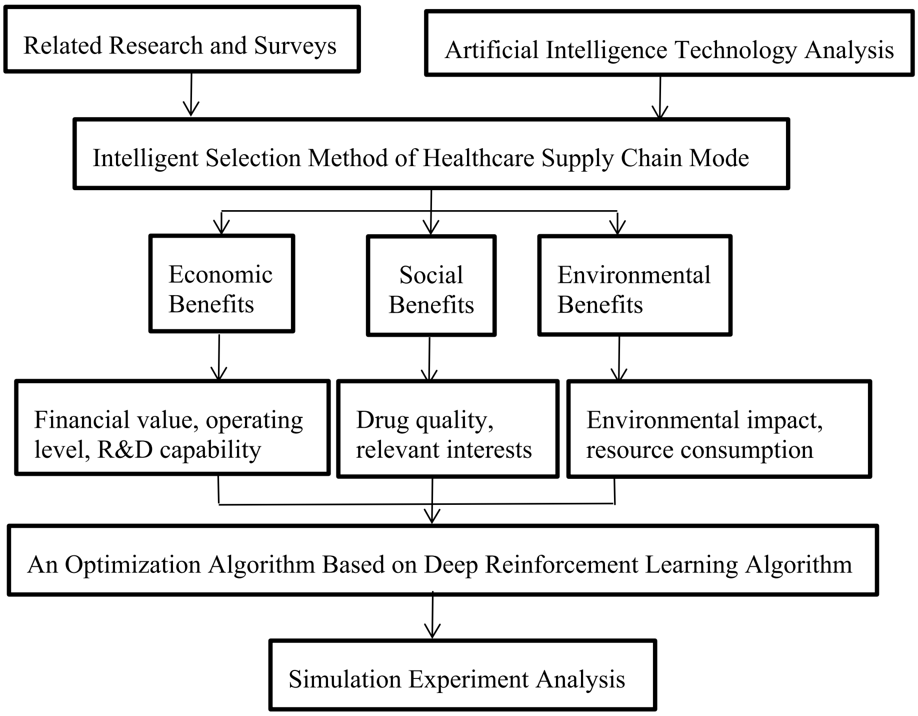


FIGURE 2
Intelligent selection process for healthcare supply chain mode.

the characteristics of the healthcare supply chain and the acceptability of the indicator data, the article selects the two indicators of pharmaceutical inventory turnover rate and the healthcare supply chain management cost rates to evaluate the internal operation and management level of the healthcare supply chain.

(1) Pharmaceutical inventory turnover rate. Inventory turnover rate is an important indicator for measuring the management capacity of healthcare supply chain inventory. The higher the value of this indicator, the faster the represented inventory transition to the capital.

$$\text{Pharmaceutical inventory turnover rate} = \frac{\text{Sales cost}}{\text{Average inventory balance}} \times 100\% \quad (4)$$

(2) The healthcare supply chain management cost rate. Management cost rate can reflect the current cost of the current healthcare supply chain operation cost.

$$\text{Management cost rate} = \frac{\text{Total operating cost}}{\text{Total operating income}} \times 100\% \quad (5)$$

Pharmaceutical research and development and innovation capabilities are the key factors affecting the sustainable development of the healthcare supply chain. Only when healthcare supply chain continues to launch new products and services can the supply chain be maintained and effectively expand its market share. The healthcare supply chain pharmaceutical R&D and innovation capabilities are measured by two indicators of new technologies and research and development investment rates.

(1) New technology adoption rate. The adoption rate of new technology refers to the ratio of the output value of the new technology product to the total output value of the enterprise. This indicator reflects the contribution of new technologies to the total value of corporate products.

$$\text{New technology adoption rate} = \frac{\text{Gross new technology product value}}{\text{Gross product value}} \times 100\% \quad (6)$$

(2) R&D investment rate. This index refers to the proportion of the cost of healthcare supply chain for research and development over a certain period of time in the total operating income of the supply chain, reflecting the importance of healthcare supply chain to innovation development.

$$\text{R \& D investment rate} = \frac{\text{R \& D total investment}}{\text{Total operating income}} \times 100\% \quad (7)$$

3.2.2 Social benefits of healthcare supply chain

The supply of high-quality drugs is one of the main targets of healthcare supply chain operations, and the quality of drugs must have strict implementation standards. The quality of drugs can be measured from the three indicators of pharmaceutical qualification rate, pharmaceutical return rate, and market share.

(1) Pharmaceutical qualification rate. The qualified rate of drugs refers to the percentage of the number of qualified drugs in a certain period of time. For a certain period of time, the medical supply chain production drug A, a total of N batches. The output of the $i(1 \le i \le N)$ is T, of which the number of qualified products is Qi,the qualification rate of the drug during this period is:

$$RQA = \frac{\sum_{i=1}^{N} Qi}{\sum_{i=1}^{N} Ti} \times 100\% \tag{8}$$

(2) Pharmaceutical return rate. Product return rate refers to the percentage of the number of drugs in the total supply of drugs in a certain period of time. For a certain period of time, the healthcare supply chain issues a total of N batches for drug A. The supply volume of the $i(1 \le i \le N)$ is Si. The number of returns is Ri. The return rate of the drug product during this period is:

$$RRA = \frac{\sum_{i=1}^{N} Ri}{\sum_{i=1}^{N} Si} \times 100\% \tag{9}$$

(3) Market share. The market share of drugs is a key indicator of measurement of the competitiveness of the drug market. It can be expressed as the ratio of drugs in a regional medical supply chain to account for the proportion of similar drugs in the region. For the sales of drugs supplied by a certain period of healthcare supply chain is S', the total number of pharmaceutical sales in the region is S,the market share of pharmaceutical is:

$$RMS = \frac{S'}{S} \times 100\% \tag{10}$$

During the management of healthcare supply chain, enterprises need to regularly conduct education and training of employees to help employees improve their learning ability. This is not only the manifestation of the responsibility of enterprises in terms of employee management, but also the obligation of the enterprise. Therefore, the growth rate of the total number of employee training in this paper reflects the protection of employees' rights and interests of pharmaceutical manufacturing enterprises. In a certain period of time, the total number of employee training changes from NT to the end of NT', and the total growth rate of employee training is:

$$RT = \frac{NT' - NT}{NT} \times 100\% \tag{11}$$

The embodiment of shareholders' equity is mainly due to the legitimate rights of shareholders. All business behaviors must be carried out on the premise that the capital security of shareholders can ensure the security of shareholders and whether it can be divorced and creating profits on time. The article uses two indicators of net assets per share and net assets to measure shareholders' rights in the medical supply chain.

(1) Net assets per share. This indicator measures the number of assets that shareholders can obtain in each shareholder issued by an enterprise.

$$\text{Net assets per share} = \frac{\text{Total shareholder equity}}{\text{Total shareholders}} \times 100\% \tag{12}$$

(2) Net asset yield. This indicator refers to the percentage share of net profit obtained by medical supply chain sales during a period of time.

$$\text{Net asset yield} = \frac{\text{Total profit of drug sales}}{\text{Owners' equity}} \times 100\% \tag{13}$$

In the process of medical supply chain management, it is also important to protect the rights and interests of creditors. For creditors, the most concerned question is whether the enterprise has the ability to pay debt. Therefore, the article selects the asset -liability ratio to measure the ability of the medical supply chain.

(1) Asset-liability ratio. Asset-liability ratio refers to the ratio of the total liabilities of the healthcare supply chain in a certain period of time. This indicator is a comprehensive indicator of evaluating the level of healthcare supply chain liabilities, and an important criterion for measuring its degree of risk.

$$\text{Asset - liability ratio} = \frac{\text{Total liability}}{\text{Total asset}} \times 100\% \tag{14}$$

3.2.3 Environmental benefits of healthcare supply chains

The environmental benefits of healthcare supply chain is an important index to highlight whether the supply chain is "green," measuring the environmental benefits of healthcare supply chain meets the requirements of the present era. This paper evaluates the environmental impact, resource consumption intensity and ISO certification rate of healthcare supply chain.

(1) Environmental impact of healthcare supply chain. Total waste emissions from healthcare supply chain. This index refers to the sum of three types of waste emissions from a series of supply chain activities, such as production by pharmaceutical manufacturing enterprises.

(2) Resource consumption intensity. The type, quantity and scarcity of medical resources used in the operation of the pharmaceutical supply chain.
(3) ISO certification rate. The enterprise through the implementation of a series of standards, the implementation of quality system certification, proof of its technical ability to meet customer requirements, provide the right products. The number of ISO certifications achieved in a medical supply chain is $NISO$, and the number of ISO certifications required in the industry is $NNISO$, the company's ISO certification rate is:

$$RISO = \frac{NISO}{NNISO} \times 100\% \quad (15)$$

3.3 Optimization algorithm based on deep reinforcement learning

Based on the above indicators and related computing formulas, with the help of the artificial intelligence system, the selection of the healthcare supply chain mode in this section is completed. This paper has selected three healthcare supply chain modes through artificial intelligence technology.

3.3.1 Healthcare supply chain cooperative mode

In this mode, medical institutions cooperate with pharmaceutical logistics enterprises or supply chain service providers, and pharmaceutical logistics enterprises participate in pharmaceutical inventory and management. Medical institutions have the choice of suppliers, centralized control of consumables, drugs acceptance, warehousing, distribution and other activities to ensure the quality of medical materials. Under the cooperative healthcare supply chain mode, the economic, social and environmental aspects have been greatly improved. In terms of economy, the cooperative mode of healthcare supply chain can achieve economies of scale, reduce procurement costs and achieve total cost savings through collective bargaining, centralized procurement and resource sharing. Secondly, the participants in healthcare supply chain can reduce personal financial risks by sharing the financial burden of procurement, inventory management and logistics. Socially, the collaborative mode of the healthcare supply chain promotes community engagement by supporting local suppliers and businesses, and promotes local economic development and spreads social well-being by prioritizing local procurement. Environmentally, the collaborative mode of the healthcare supply chain supports the development of a green environment by encouraging the use of environmentally friendly products, materials and logistics. Healthcare supply chains under the cooperative mode reduce carbon emissions by consolidating orders, optimizing logistics and sharing distribution networks, and minimize the environmental impact of their supply chain activities. Overall, healthcare supply chain cooperative mode can address economic, social and environmental issues through cost reduction, community engagement and sustainable practices. The main feature of this mode is to open up the capital flow, logistics and information flow, and greatly improve the management level of healthcare supply chain.

3.3.2 SPD mode of healthcare supply chain

SPD mode is also known as the third-party oriented mode. In the SPD mode, the third-party supply chain service provider uses modern information technology to manage the logistics, information flow and capital flow of the supply chain of medical institutions in a unified manner, aiming to build a value-adding, quality and efficiency improvement, and innovative service-oriented healthcare supply chain management system. Economically, effective SPD management can extend the service life of medical devices and equipment and reduce the frequency of replacement. This method minimizes the expenditure of fixed costs. Secondly, SPD supply chain service provider, as a logistics center, provides medical institutions with centralized procurement, warehousing and distribution services of medical consumables, helps medical institutions realize "one-to-one" procurement mode and zero inventory management of materials, realizes effective utilization of resources and reduces warehousing and logistics costs in the operation process. Socially, the SPD mode maintains high standards of sterilization handling and the rational use of safe medical devices, which are essential to ensure patient safety. Second, the SPD mode supports regular training and self-development of employees, ensuring the well-being of employees. Finally, management under the SPD mode protects the interests of creditors. In terms of the environment, the SPD mode reduces waste emissions through the efficient treatment of medical waste. Secondly, through effective sterilization treatment measures, the disposal of disposable items is optimized, contributing to environmental sustainability. The main feature of this mode is to organically integrate the logistics, information flow and capital flow of the supply chain of medical institutions and third-party supply chain service providers into one system, so as to reduce the storage cost of medical institutions and improve the efficiency of material distribution.

3.3.3 Healthcare supply chain e-commerce mode

It consists of healthcare supply chain + e-commerce platforms, and its development goal is to promote the transformation of traditional healthcare supply chain to smart medical supply chain. On the economic side, e-commerce platforms can achieve cost-effectiveness through simplified procurement processes, competitive pricing, and reduced overhead costs. E-commerce platforms facilitate efficient procurement and price negotiations by directly connecting medical institutions with suppliers, contributing to overall cost savings. Second, the healthcare supply chain e-commerce mode improves supply chain transparency, enabling healthcare organizations to make informed purchasing decisions, and this transparency can lead to better cost management and budget allocation. In terms of society, the e-commerce mode improves the accessibility of medical supplies, especially for remote areas or areas with insufficient medical supplies. By providing a digital market for purchasing medical products, people can quickly obtain medical products, so as to realize the social responsibility of enterprises. In terms of the environment, the healthcare supply chain e-commerce mode can maintain environmental protection by reducing the demand for physical transportation and the associated carbon emissions. Second, e-commerce platforms influence consumer preferences and promote the development of green practices by offering environmentally friendly and reusable products. The main feature of this mode is to reduce the hierarchical structure of the healthcare supply chain through Internet technology, enhance the information sharing and communication between various subjects in the medical supply chain, and integrate the Internet into the medical logistics to create an Internet + medicine distribution mode, and improve the operational efficiency of the healthcare supply chain.

This selection is optimized based on the Deep Deterministic Strategy Gradient Algorithm(DDPG) in the reinforcement learning algorithm (50, 51). DDPG algorithm combined with neural network for value function model approximation and deterministic strategy selection. This algorithm requires fewer samples and has high training efficiency. In addition, this algorithm is a reinforcement learning method based on deep learning related theories and technical methods, and takes deterministic strategy gradient as its optimization goal.

DDPG algorithm has four networks:

(1) Actor online network Q: Responsible for the iterative update of parameter θ, while selecting the current action a according to the current state s, interacting with the environment to get the state s' and rewarding r.
(2) Actor target network Q': Select the best action a' from the next state s', θ' is the parameter.
(3) Critic online network $\propto$: Responsible for the iterative update of parameter ω. Calculate the current value function at the same time, and the expression of the target Q value is as follows:

$$y_i = r_i + \gamma Q^{'}\left(s',a',\omega'\right) \tag{16}$$

λ represents the discount function.

(4) Critic target network $\propto'$: Calculate Q' in the target Q value, ω' is the parameter.

The network parameters in DDPG are updated as shown in equation (17):

$$\begin{cases} \theta' \leftarrow \tau\theta + (1-\tau)\theta' \\ \omega' \leftarrow \tau\omega + (1-\tau)\omega' \end{cases} \tag{17}$$

τ is the update coefficient, generally selected a small number, usually 0.01 or 0.001.

In the learning process, in order to reduce the regularity of learning, a noise N is added to the selected action a, and the obtained action becomes:

$$\mathrm{a} = \pi\theta(\mathrm{s}) + N \tag{18}$$

Update the Critic network according to Critical loss:

$$L = \frac{1}{N}\sum_{\mathrm{i}}^{N}\left(yi - Q\left(si,ai\middle|\omega\right)\right)^2 \tag{19}$$

Update the Actor online network:

$$\nabla\theta\mu\,\mu|\mathrm{si} \cong \frac{1}{N}\sum_{\mathrm{i}}\nabla\mathrm{ai}Q\left(\mathrm{si,a\,i}\middle|\omega\right)\nabla\theta\mu\left(\mathrm{si}\middle|\theta\right) \tag{20}$$

Artificial intelligence DDPG algorithm is used to select the medical supply chain model, and its process is mainly as follows:

1 Build a simulation environment.
2 Design a container BUFFER with storage data and random sampling.
3 Build model, define DDPG algorithm, specify loss function and optimizer. The Agent will interact with its environment and learn from it. After the interaction, it will obtain a state's provided by the environment. After receiving the state, the agent will use the actor network to calculate the action a that should be taken in the current state. Meanwhile, the agent will predict the Q value corresponding to action A in this state according to the critic network. When the action is fed back to the environment, the environment will give the corresponding reward r, the new state's, and whether to trigger the termination condition done. After each interaction is completed, the DDPG algorithm will store it as an experience in the experience pool and extract a certain amount of experience from the experience pool as input data to train the neural network. Through the execution of multiple rounds of training, constantly adjust the parameters to achieve better results.
4 Evaluate and test the trained model and observe the reward.
5 Save the model to a specified location for subsequent reasoning or continued training.
6 Test the performance of the model in the test set data.

Thus, the healthcare supply chain mode can be determined according to the above algorithm. The overall framework flow chart of DDPG algorithm is shown in the Figure 3. At this point, the intelligent selection method of healthcare supply chain mode based on artificial intelligence technology has been designed and completed.

4 Simulation experiment

4.1 Simulation experiment plan

In this study, in order to make up for the shortcomings of the current healthcare supply chain mode selection method, the intelligent selection method of healthcare supply chain mode based on artificial intelligence technology is proposed. To verify the application effect of this method, constructing simulation experiment to verify the effect of artificial intelligence method. Experimental environment settings: in the DDPG algorithm, Relu functions are used as activation functions for the Actor primary network, and the activation function for the target network is the sigmoid function. The update rate for each network parameter is 0.01 and the learning rate for each network is set to 0.00001. The path formed by healthcare supply chain to achieve the maximum economic, social and environmental benefits under ideal conditions is set as the target mode. In this study, the setting of the target mode of the healthcare supply chain is shown in Figure 4.

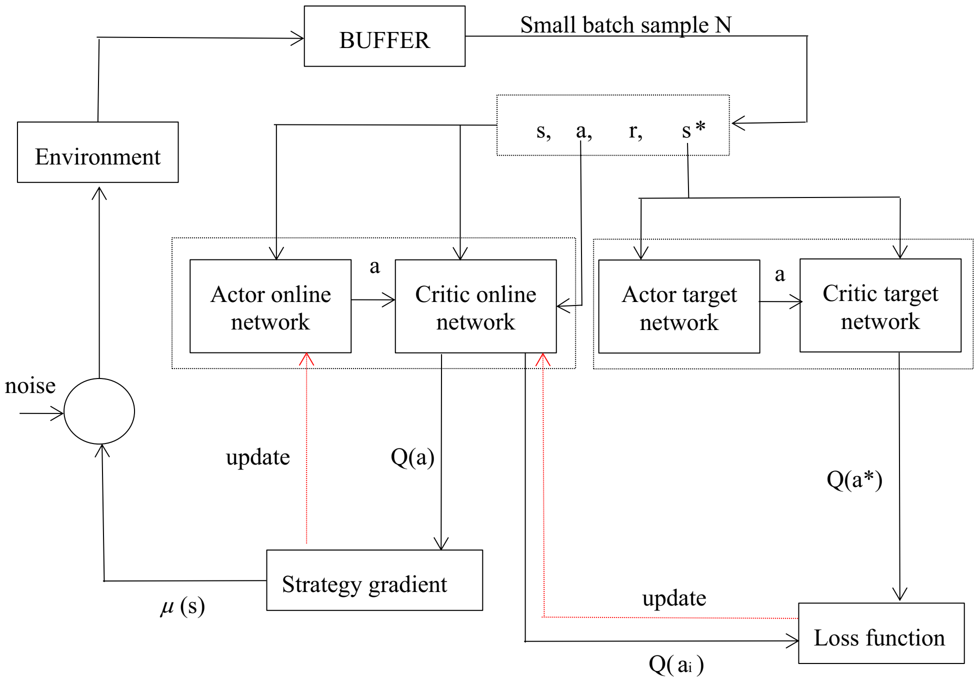


FIGURE 3
DDPG algorithm framework.

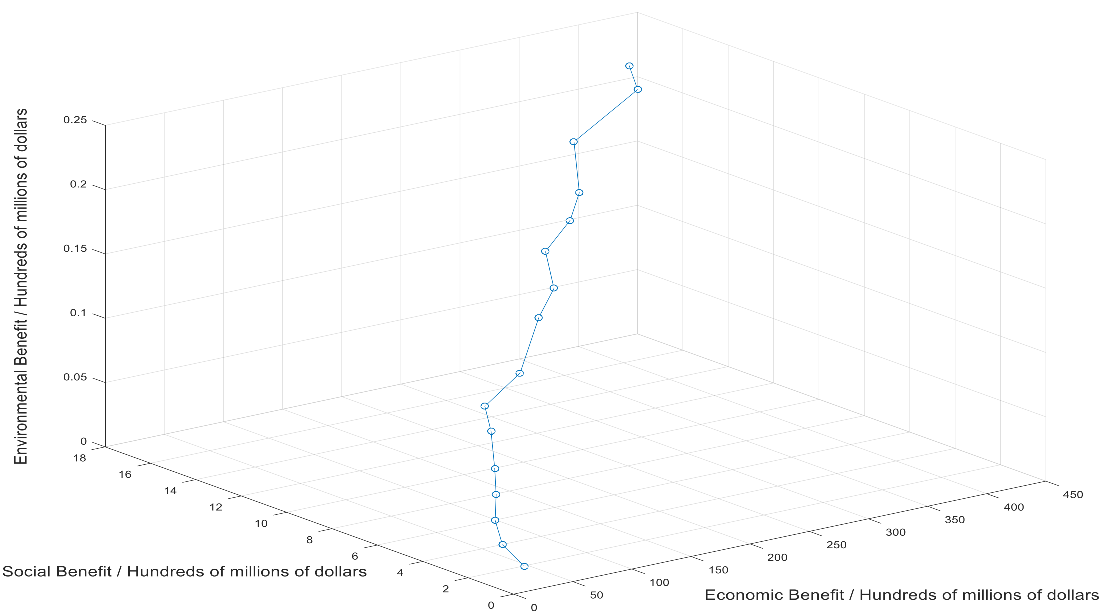


FIGURE 4
Optimal mode setting for healthcare supply chain.

In order to verify the effect of AI selection method, after analyzing a large number of literature and cases (52–54), the paper selects big data selection method, basic selection method and neural network selection method to compare with AI selection method, and presents the experimental comparison results through different curves.

4.2 Experimental results

DDPG algorithm generates actions through Actor network, and the collected data (state-action) is trained on the algorithm model. The results of healthcare supply chain mode selection based on DDPG algorithm can be obtained from the simulation data in

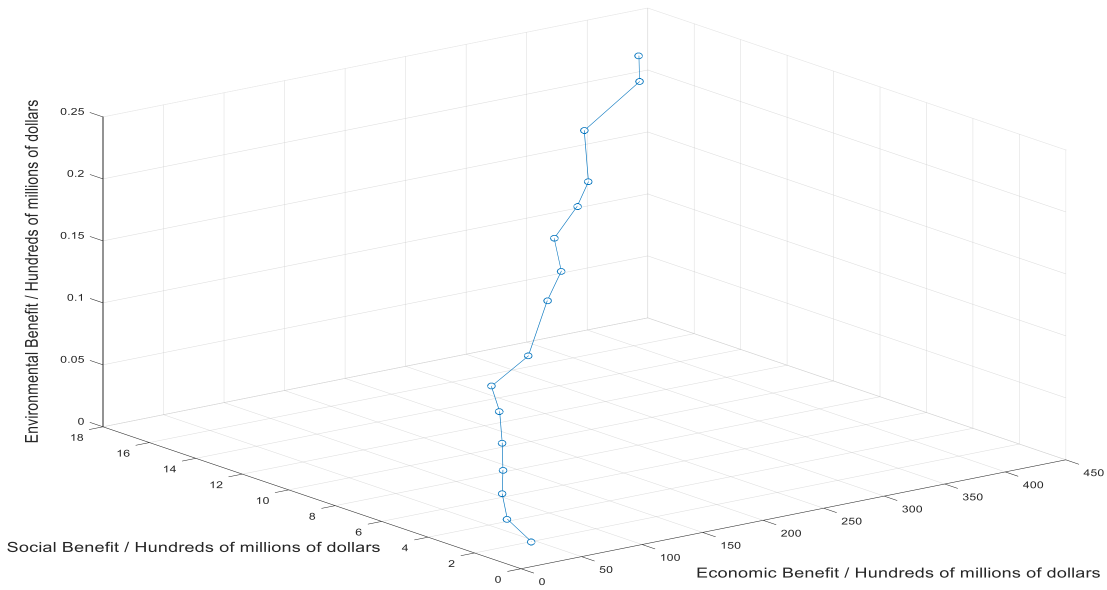


FIGURE 5
Healthcare supply chain mode selection results based on DDPG algorithm.

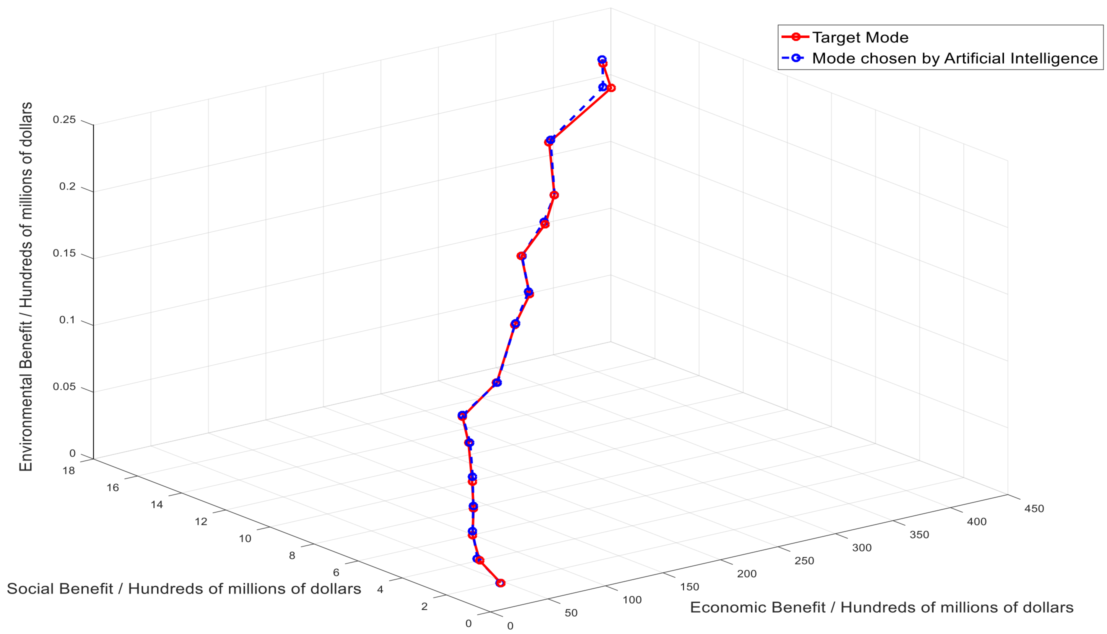


FIGURE 6
Comparison between target mode and AI-based mode.

Figure 5. From the experimental results, we can intuitively see the path of the healthcare supply chain mode selected by DDPG algorithm in the three dimensions of economy, society and environment. Whether it is in economic benefit dimension, social benefit dimension or environmental dimension, it is on the rise, and the highest economic benefit is $44.27 billion, the highest social benefit is $1.8 billion, the highest environmental benefit is $21.3 million. In order to verify the effect of application of artificial intelligence algorithm in healthcare supply chain mode selection, the results of healthcare supply chain mode selected based on DDPG algorithm were compared and analyzed with the results of target mode, as shown in Figure 6. Through comparative analysis, it can be seen that the results of healthcare supply chain mode selection obtained by artificial intelligence algorithm are consistent with the results of target mode. There are many overlapping points and the highest points of economic, social and environmental benefits almost coincide. Therefore, it can be seen that the application effect of artificial intelligence in healthcare supply chain mode selection is good, and artificial intelligence can help healthcare supply chain choose the optimal target mode.

In order to conduct a more detailed analysis of the application effect of artificial intelligence, this study introduces three other commonly used methods for comparison. Under the same experimental environment, we conducted simulation experiments on healthcare supply chain mode selected by different methods, as shown in Figure 7. Different healthcare supply chain mode selection results can be clearly seen from the figure. First of all, in healthcare supply chain mode selection results obtained by the basic selection method, the highest economic benefit is $44.19 billion, the highest social benefit is $1.78 billion, and the highest environmental benefit is $21 million, which is far from the result of the target mode. It can be seen that the application of this method in healthcare supply chain mode selection still has certain defects. This method cannot accurately screen out the optimal mode of healthcare supply chain. Secondly, in the results of healthcare supply chain mode based on the neural network selection method, the highest economic benefit is $44.37 billion, the highest social benefit is $1.9 billion, and the highest environmental benefit is $21 million, which is still a certain gap with the target mode, although the highest point is slightly higher than the target mode and the middle part is relatively close to the target mode. However, on the whole, healthcare supply chain mode selected by this method is relatively ideal, and it is difficult to implement in combination with the reality and complex situation. Finally, according to the results of healthcare supply chain mode obtained by the big data selection method, the highest economic benefit is $44.192 billion, the highest social benefit is $1.76 billion, and the highest environmental benefit is $21 million, which is also inconsistent with the results of the target mode. Although there are overlapping parts, the whole path still deviates from the path of the target mode.

To sum up, healthcare supply chain mode selected by the above three methods has certain errors with the target mode, that is to say, the existing commonly used research methods cannot effectively select the optimal healthcare supply chain mode, and it needs to be improved. In addition, after comparing the results in this part, it can be found that healthcare supply chain mode selected by artificial intelligence algorithm is the closest to the target mode with the smallest error. Therefore, the simulation experiment in this part can verify that artificial intelligence is more effective in healthcare supply chain mode selection, compared with other methods. Artificial intelligence has more advantages in the choice of healthcare supply chain mode, which can ensure the quality of healthcare supply chain mode selection.

5 Discussion

According to the results of this study, as shown in Figure 5, the path presented by healthcare supply chain mode based on DDPG algorithm is basically consistent with the path presented by the target mode, and the error between the two is very small, which indicates the effectiveness of the application of artificial intelligence technology. In the choice of healthcare supply chain mode, artificial intelligence technology can help healthcare supply chain make intelligent decisions and choose the best operation mode. However, the application of artificial intelligence technology in the selection of healthcare supply chain mode is blank, which limits the practical application of artificial intelligence in the medical field. In fact, there have been some achievements in the application of artificial intelligence technology in the healthcare field, including the use of artificial intelligence technology for disease diagnosis and surgery (55, 56), and the use of artificial intelligence technology in these fields has been proven to be effective, and its accuracy even exceeds that of human medical experts. The development of artificial intelligence has undoubtedly brought revolutionary innovation to the medical industry, and the

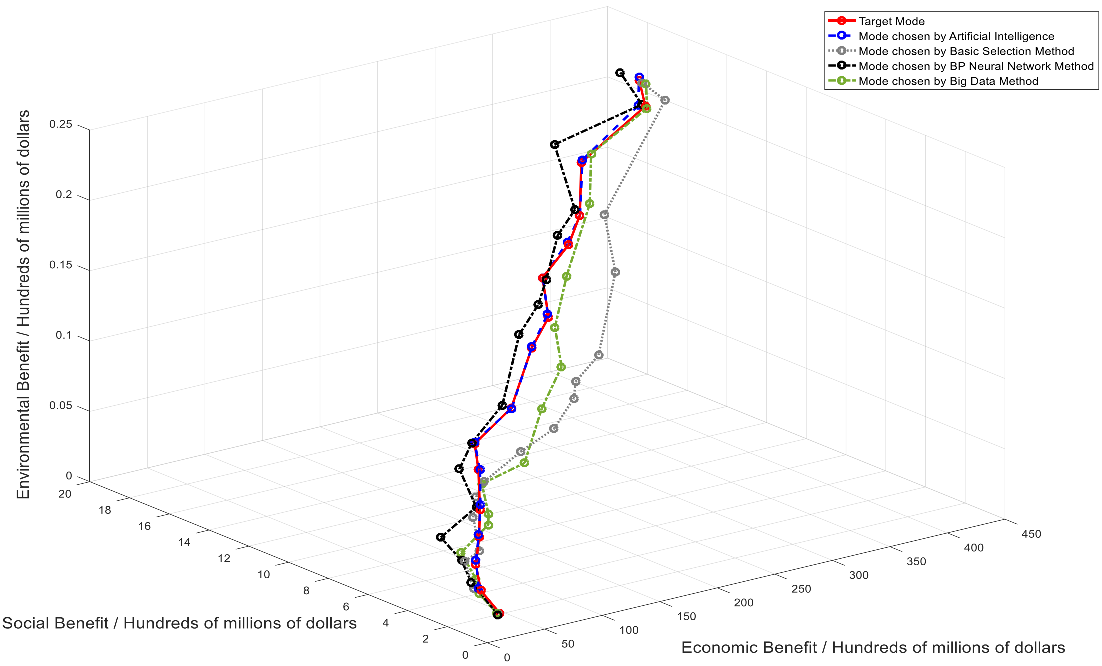


FIGURE 7
Comparison of target mode and multiple method selection results.

application of artificial intelligence algorithms in healthcare supply chain has also helped the medical industry to create a more intelligent and secure system, which is conducive to the improvement of the overall resilience level of healthcare supply chain, and is more conducive to the sustainable development of healthcare supply chain.

Therefore, this paper fills the gap in the application of artificial intelligence in the field of healthcare supply chain, and innovatively incorporates artificial intelligence technology into the choice of healthcare supply chain mode, which is not only conducive to the practical value of artificial intelligence technology, but also conducive to the improvement of the modern management level of healthcare supply chain. Specifically, firstly, healthcare supply chain model selection driven by artificial intelligence can better predict demand and manage inventory. AI algorithms can analyze historical patient data, disease trends, and other relevant factors to forecast the demand for medical supplies, pharmaceuticals, and equipment. Accurate demand forecasting can help healthcare organizations optimize inventory levels, reduce stockouts, and minimize excess inventory, leading to cost savings and improved operational efficiency. Secondly, the inclusion of artificial intelligence system in mode selection can provide decision support for managers. Artificial intelligence algorithm can analyze complex data sets and determine the most effective supply chain mode according to numerous indicators, help managers optimize the operation process of healthcare supply chain and improve the supply chain structure, so as to improve the efficiency of healthcare supply chain. Finally, the process of using AI for mode selection often involves supply chain risk management and resource allocation. Artificial intelligence can be used to identify risks in healthcare supply chain to ensure the continuity of supply chain operations and prevent supply chain interruptions. Artificial intelligence can also optimize resource allocation in healthcare supply chain and ensure the real-time supply of medical resources.

In addition, in order to further illustrate the advantages of artificial intelligence in healthcare supply chain mode selection and verify that artificial intelligence is superior to conventional methods, we selected three other commonly used methods for comparison. The first is the basic selection method, that is, to evaluate the key indicators of healthcare supply chain, and then select the corresponding mode according to the key indicator data of healthcare supply chain. This is the most common and commonly used selection method, but with the advent of the era of big data and the development of emerging technologies, this method has been gradually replaced by other methods. The second, neural network selection approach, is a computational model inspired by the structure and function of the human brain. It consists of interconnected nodes, or "neurons" that work together to process and analyze complex data. In a neural network, information is processed through layers of interconnected nodes, each of which applies a mathematical function to the input data and passes the results to the next layer. Neural network methods are popular for their ability to process large and complex data sets, learn from experience, and generalize patterns from input data to make predictions or decisions. They are used in a wide range of fields, including finance, healthcare, robotics, and many others (57). The third is the Big Data selection approach, which uses appropriate big data analysis methods and tools to effectively analyze large and complex data sets. This involves considering various factors such as the specific objectives of the analysis, the type and source of the data, scalability requirements, data processing tools, data storage and retrieval methods, and data security and privacy considerations. A big data selection approach is essential for organizations to gain valuable insights from the vast amount of available data and make informed decisions (58). As shown in Figure 6, we have drawn five paths, among which the path of the target mode is the reference object of this experiment. It can be intuitively seen from the figure that the path of healthcare supply chain mode selected by the basic selection method, BP neural network method and big data method is different from the path of the target mode to some extent. Healthcare supply chain mode obtained by the basic selection method has a slightly larger gap with the target model, while healthcare supply chain mode selected by the neural network method and big data method is close to the target mode, but there is still a certain gap compared with healthcare supply chain mode selected by artificial intelligence. Only the healthcare supply chain mode selected by artificial intelligence is more in line with the target mode. It can be judged that the accuracy rate of artificial intelligence application in healthcare supply chain mode selection is higher than that of existing common research methods. The use of artificial intelligence greatly improves the effectiveness of selection results, and artificial intelligence is the most effective and advantageous method.

6 Conclusion

The combination of artificial intelligence and healthcare field has become a hot spot at present, and the potential of artificial intelligence in the healthcare supply chain is the focus of existing research. However, few studies have explained the choice of healthcare supply chain mode, and the existing research methods are not systematic and effective. In order to make up for this shortcoming, the paper applies artificial intelligence to the choice of healthcare supply chain mode and realizes the intelligent choice of healthcare supply chain mode. Aiming at the intelligent choice of healthcare supply chain mode, firstly, this paper determines the indicators of evaluating healthcare supply chain mode, and defines the relevant calculation methods to provide support for effectively choosing the optimal healthcare supply chain mode. Secondly, combined with evaluation indicators, artificial intelligence technology is used to identify different healthcare supply chain modes, and the intelligent selection optimization method of healthcare supply chain modes is proposed based on deep reinforcement learning algorithm. Finally, the simulation experiment is constructed to carry out quality verification of the effect of artificial intelligence technology application, and the effectiveness of artificial intelligence application is demonstrated by comparing with different methods. The research results show that the path presented by the healthcare supply chain mode based on DDPG algorithm is basically consistent with the path presented by the target mode, which indicates the effectiveness of the application of artificial intelligence technology, and the path generated by the healthcare supply chain mode selected by the basic selection method, BP neural network method and big data method is different from the path of the target mode. This study not only makes a theoretical contribution, but also has practical significance. Further, we also point out the limitations of the study and the direction of future research.

6.1 Implications

In the current highly uncertain market environment, in order to promote the sustainable development of healthcare supply chain, this paper starts from the key point of healthcare supply chain mode selection, studies the intelligent choice of healthcare supply chain mode, applies artificial intelligence to the field of healthcare supply chain innovationally, and demonstrates the potential of artificial intelligence in the application of healthcare supply chain mode selection. The theoretical contribution of this study is to propose a more scalable and intelligent deep reinforcement learning algorithm to optimize the mode selection process of healthcare supply chain. By demonstrating the applicability of artificial intelligence in unknown environment and complex decision making, this algorithm demonstrates the rationality of artificial intelligence algorithm for intelligent decision making of supply chain mode. It makes up for the lack of objectivity of existing research methods, limited data analysis and inapplicability to dynamic environment, and enriches the existing research method system of healthcare supply chain. The healthcare supply chain can utilize the decision-making framework of deep reinforcement learning algorithm to optimize the choice of healthcare supply chain mode, which deepens the theoretical application of artificial intelligence in the healthcare supply chain and provides valuable insights for the integrated development of artificial intelligence and health economy. In addition, the practical significance of this study lies in the optimization of healthcare supply chain management, help medical enterprises to reduce costs while ensuring the quality and safety of drugs, and enable healthcare supply chain enterprises to better understand the shortcomings in the operation of the supply chain, so as to better improve the resilience of enterprises in the highly uncertain and highly risky market environment, and quickly respond to changes in consumer demand. In the end, gain competitive advantage in the market.

6.2 Limitations and future prospects

There are some limitations and deficiencies in this study. First, this study is limited by conditions, and the data collected in the simulation experiment is relatively limited, and the area involved is relatively narrow. Future studies can expand the scope of investigation and collect more samples. Second, the paper compares the artificial intelligence method with the basic selection method, BP neural network method and big data method. Although it has certain credibility, the existing other selection methods have not been fully considered. Therefore, in future research, the application of other models and methods can be considered to further improve the content of this part.

In summary, this paper combines artificial intelligence with healthcare supply chain, studies the intelligent selection process of healthcare supply chain mode, and proposes an intelligent selection optimization method of healthcare supply chain mode based on deep reinforcement learning algorithm. Finally, the simulation experiment verifies the effect of artificial intelligence technology in the application of healthcare supply chain mode selection, proving that the application of artificial intelligence in healthcare supply chain can also be extended to a wider range of research, providing innovative technical and method support for the future development of the health field.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

PL: Conceptualization, Writing – original draft. LL: Conceptualization, Investigation, Writing – original draft. QC: Investigation, Writing – review & editing. YC: Formal analysis, Writing – review & editing. CL: Formal analysis, Investigation, Writing – review & editing. XL: Formal analysis, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Guangxi Key R&D Plan (Project No: 2022AB34029): The Key Technologies and Industrialization of Intelligent Traceability in the Whole Link of High-quality Seedling Seed Supply Chain in Lijiang River Basin and by the Research Fund Project of Development Institute of Zhujiang-Xijiang Economic Zone, Key Research Base of Humanities and Social Sciences in Guangxi Universities (Project No: ZX2023051) and by the Innovation Project of Guangxi Graduate Education (JG2023018).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Beaulieu M, Bentahar O. Digitalization of the healthcare supply chain: a roadmap to generate benefits and effectively support healthcare delivery. *Technol Forecast Soc Chang.* (2021) 167:120717. doi: 10.1016/j.techfore.2021.120717

2. Bialas C, Bechtsis D, Aivazidou E, Achillas C, Aidonis D. Digitalization of the healthcare supply chain through the adoption of Enterprise resource planning (ERP) Systems in Hospitals: an empirical study on influencing factors and cost performance. *Sustainability.* (2023) 15:3163–3. doi: 10.3390/SU15043163

3. Senna P, Reis A, Dias A, Coelho O, Guimarães J, Eliana S. Healthcare supply chain resilience framework: antecedents, mediators, consequents. *Prod Plann Control*. (2023) 34:295–309. doi: 10.1080/09537287.2021.1913525

4. Ertz M, Patrick K. The future of sustainable healthcare: extending product lifecycles. *Resour Conserv Recycl*. (2020) 153:104589–10. doi: 10.1016/j.resconrec.2019.104589

5. Joseph MW, Shahram A. Managing the Blockchain-enabled digital transformation of the healthcare supply chain. *Manage Healthcare Peer Rev J*. (2021) 34:1111–28. doi: 10.1108/MEQ-02-2022-0025

6. Govindan K, Mina H, Alavi B. A decision support system for demand management in h-ealthcare supply chains considering the epidemic outbreaks: a case study of coronavirus diseas-e 2019 (COVID-19). *Transp Res E Logist Transp Rev*. (2020) 138:1–14. doi: 10.1016/j.tre.2020.101967

7. Hussain M, Ajmal MM, Gunasekaran A, Khan M. Exploration of social sustainability in healthcare supply chain. *J Clean Prod*. (2018) 203:977–89. doi: 10.1016/j.jclepro.2018.08.157

8. Tate W, Ellram L, Kirchoff J. Corporate social responsibility reports: a thematic analysis related to supply chain management. *J Supply Chain Manag*. (2010) 46:19–44. doi: 10.1111/j.1745-493X.2009.03184.x

9. Haraguchi N, Cheng CFC, Smeets E. The importance of manufacturing in economic development: has this changed? *World Dev*. (2017) 93:293–315. doi: 10.1016/j.worlddev.2016.12.013

10. Ilya L, Avi H. An optimization model for distribution of influenza vaccines through a green healthcare supply chain. *Oper Res Health Care*. (2023) 37:100387–15. doi: 10.1016/j.orhc.2023.100387

11. Kümmerer K, Hempel M. *Green and sustainable pharmacy*. Berlin, Heidelberg: Springer, pp. 37–59. (2010).

12. Borges FA, Laurindo JF, Spínola MM, Gonçalves RF, Mattos CA. The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. *Int J Inf Manag*. (2020) 57:102225. doi: 10.1016/j.ijinfomgt.2020.102225

13. Gupta H, Kumar A, Wasan P. Industry 4.0, cleaner production and circular economy: an integrative framework for evaluating ethical and sustainable business performance of manufacturing organizations. *J Clean Prod*. (2021) 295:126253–18. doi: 10.1016/j.jclepro.2021.126253

14. Bala H, Feng X. Success of small and medium enterprises in Myanmar: role of technological, organizational, and environmental factors. *J Glob Inf Technol Manag*. (2019) 22:100–19. doi: 10.1080/1097198X.2019.1603511

15. Ni D, Xiao Z, Lim KM. A systematic review of the research trends of machine learning in supply chain management. *Int J Mach Learn Cybern*. (2019) 11:1463–82. doi: 10.1007/s13042-019-01050-0

16. Al Kuwaiti A, Nazer K, Al-Reedy A, Al-Shehri S, Al-Muhanna A, Subbarayalu AV, et al. A review of the role of artificial intelligence in healthcare. *J Pers Med*. (2023) 13:951. doi: 10.3390/jpm13060951

17. Kumar A, Mani V, Jain V, Gupta H, Venkatesh VG. Managing healthcare supply chain through artificial intelligence (AI): a study of critical success factors. *Comput Ind Eng*. (2023) 175:108815. doi: 10.1016/j.cie.2022.108815

18. Smith BK, Nachtmann H, Pohl EA. Improving healthcare supply chain processes via data standardization. *Eng Manag J*. (2012) 24:3–10. doi: 10.1080/10429247.2012.11431924

19. Kim HK, Lee CW. Relationships among healthcare digitalization, social capital, and supply chain performance in the healthcare manufacturing industry. *Int J Environ Res Public Health*. (2021) 18:1417. doi: 10.3390/ijerph18041417

20. Senna P, Reis A, Marujo LG, Marujo LG, Ferro de Guimarães JC, Severo EA, et al. The influence of supply chain risk management in healthcare supply chains performance. *Prod Plann Control*. (2023) 1:1–16. doi: 10.1080/09537287.2023.2182726

21. Jamil F, Hang L, Kim K, Kim DH. A novel medical Blockchain model for drug supply chain integrity management in a smart hospital. *Electronics*. (2019) 8:505. doi: 10.3390/electronics8050505

22. Edoh TO, Teege G. Using information Technology for an Improved Pharmaceutical Care Delivery in developing countries. Study case: Benin. *J Med Syst*. (2011) 35:1123–34. doi: 10.1007/s10916-011-9717-y

23. Gao G, Che Y, Shen J. Path optimization for joint distribution of medical consumables under hospital SPD supply chain mode. *J Comb Optim*. (2021) 42:866–83. doi: 10.1007/s10878-019-00506-x

24. Chowdhury NR, Ahmed M, Mahmud P, Paul SK, Liza SA. Modeling a sustainable vaccine supply chain for a healthcare system. *J Clean Prod*. (2022) 370:133423–19. doi: 10.1016/j.jclepro.2022.133423

25. Kanokphanvanich C, Rattanawong W, Vongmanee V. A new model for a sustainable healthcare supply chain prioritizes patient safety: using the fuzzy Delphi method to identify healthcare workers' perspectives. *Sustainability*. (2023) 15:7123. doi: 10.3390/su15097123

26. Rajak S, Mathiyazhagan K, Agarwal V, Sivakumar K, Kumar V, Appolloni A. Issues and analysis of critical success factors for the sustainable initiatives in the supply chain during COVID-19 pandemic outbreak in India: a case study. *Res Transp Econ*. (2022) 93:101114. doi: 10.1016/j.retrec.2021.101114

27. Rajak S, Palanisamy P, Dhanalakshmi R. A DEA model for evaluation of efficiency and effectiveness of sustainable transportation system: a supply chain perspective. *Int J Logist Syst Manage*. (2020) 40:220. doi: 10.1504/IJLSM.2021.118737

28. Zhao F, Xu H. Applications and current status of AI in the medical field. *J Phys Conf Ser*. (2022) 2289:12030–7. doi: 10.1088/1742-6596/2289/1/012030

29. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chatbot for medicine. *N Engl J Med*. (2023) 388:1233–9. doi: 10.1056/NEJMsr2214184

30. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. (2023) 5:E107–8. doi: 10.1016/S2589-7500(23)00021-3

31. Sethuraman N. Artificial intelligence: a new paradigm for pharmaceutical applications in formulations development. *Indian J Pharm Educ Res*. (2020) 54:843–6. doi: 10.5530/ijper.54.4.176

32. Smalley E. AI-powered drug discovery captures pharma interest. *Nat Biotechnol*. (2017) 35:604–5. doi: 10.1038/nbt0717-604

33. Naz F, Agrawal R, Kumar A, Gunasekaran A, Majumdar A, Luthra S. Reviewing the applications of artificial intelligence in sustainable supply chains: exploring research propositions for future directions. *Bus Strategy Environ*. (2022) 31:2400–23. doi: 10.1002/bse.3034

34. Deveci M. Effective use of artificial intelligence in healthcare supply chain resilience using fuzzy decision-making model. *Soft Comput*. (2023) 1:1–14. doi: 10.1007/s00500-023-08906-2

35. Chen YJ, Moreira P, Liu WW, Monachino M, TLH N, Wang A. Is there a gap between artificial intelligence applications and priorities in health care and nursing management? *J Nurs Manag*. (2022) 30:3736–42. doi: 10.1111/jonm.13851

36. Sun R, Deutsch E, Fournier L. Artificial intelligence and medical imaging. *Bull Cancer*. (2022) 109:83–8. doi: 10.1016/j.bulcan.2021.09.009

37. Joshi S, Sharma M, Das RP, Rosak-Szyrocka J, Żywiołek J, Muduli K, et al. Modeling conceptual framework for implementing barriers of AI in public healthcare for improving operational excellence: experiences from developing countries. *Sustainability*. (2022) 14:11698. doi: 10.3390/su141811698

38. Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int J Inf Manag*. (2021) 57:101994. doi: 10.1016/j.ijinfomgt.2019.08.002

39. Verde LG, De Pietro A, Ghoneim M, Alrashoud M, Al-Mutib KN, Sannino G. Exploring the use of artificial intelligence techniques to detect the presence of coronavirus COVID-19 through speech and voice analysis. *IEEE Access*. (2021) 9:65750–7. doi: 10.1109/ACCESS.2021.3075571

40. Ho D. Addressing COVID-19 drug development with artificial intelligence. *Adv Intell Syst*. (2020) 2:2000070. doi: 10.1002/aisy.202000070

41. Balasubramanian S, Shukla V, Islam N, Upadhyay A, Duong L. Applying artificial intelligence in healthcare: lessons from the COVID-19 pandemic. *Int J Prod Res*. (2023) 1:1–34. doi: 10.1080/00207543.2023.2263102

42. Li H, Qin J, Liu Q, Yan C. An efficient deep reinforcement learning algorithm for Mapless navigation with gap-guided switching strategy. *J Intell Robot Syst*. (2023) 108:1–14. doi: 10.1007/s10846-023-01888-1

43. Zhu QL, Wu XQ, Lin QZ, Ma L, Li J, Ming Z. A survey on evolutionary reinforcement learning algorithms. *Neurocomputing*. (2023) 556:126628. doi: 10.1016/J.NEUCOM.2023.126628

44. Han D, Mulyana B, Stankovic V, Cheng S. A survey on deep reinforcement learning algorithms for robotic manipulation. *Sensors*. (2023) 23:3762. doi: 10.3390/s23073762

45. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deepreinforcement learning. *Nature*. (2015) 518:529–33. doi: 10.1038/nature14236

46. Hasselt HV, Guez A, Silver D. *Deep reinforcement learning with double q-learning*. In: AAAI Conference on Artificial Intelligence, pp. 2094–2100. (2015). doi: 10.48550/arXiv.1509.06461

47. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforce-ment learning. *arXiv*. (2015) 2015:2971. doi: 10.48550/arXiv.1509.02971

48. Ahmad F, Alnowibet KA, Alrasheedi AF, Adhami AY. A multi-objective model for optimizing the socio-economic performance of a pharmaceutical supply chain. *Socio Econ Plan Sci*. (2022) 79:101126–2. doi: 10.1016/J.SEPS.2021.101126

49. Low YS, Halim I, Adhitya A, Chew W, Sharratt P. Systematic framework for Design of Environmentally Sustainable Pharmaceutical Supply Chain Network. *J Pharm Innov*. (2016) 11:250–63. doi: 10.1007/s12247-016-9255-8

50. Hu H, Chen Y, Wang T, Feng F, Chen W. Research on the deep deterministic policy algorithm based on the first-order inverted pendulum. *Appl Sci*. (2023) 13:7594. doi: 10.3390/APP13137594

51. Sha ZX, Huo R, Sun C, Wang S, Huang T. A task-oriented hybrid routing approach based on deep deterministic policy gradient. *Comput Commun*. (2023) 210:183–93. doi: 10.1016/J.COMCOM.2023.07.040

52. Liu L, Ran W. Research on supply chain partner selection method based on BP neural network. *Neural Comput & Applic*. (2020) 32:1543–53. doi: 10.1007/s00521-019-04136-6

53. Cao H. Big data attribute selection method in distributed network fault diagnosis database. *J Intell Fuzzy Syst*. (2020) 38:7903–14. doi: 10.3233/jifs-179859

54. San Cristobal JR. Contractor selection using multicriteria decision-making methods. *J Constr Eng Manag*. (2012) 138:751–8. doi: 10.1061/(ASCE)CO.1943-7862.0000488

55. Zhang YM, Weng Y, Lund J. Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics*. (2022) 12:237–7. doi: 10.3390/DIAGNOSTICS12020237

56. Loftus JT, Upchurch RG, Bihorac A. Use of artificial intelligence to represent emergent systems and augment surgical decision-making. *JAMA Surg*. (2019) 154:791–2. doi: 10.1001/jamasurg.2019.1510

57. Li Y, Sun H, Chen XY, Xin LJ, Zhang X. Probe selection and power weighting in multiprobe OTA testing: a neural network-based approach. *Int J Antennas Propag*. (2019) 2019:1–8. doi: 10.1155/2019/1392129

58. Milton S, Pacheco NR. A big data approach to explore medical imaging repositories based on DICOM. *Proc Comput Sci*. (2023) 219:1224–31. doi: 10.1016/J.PROCS.2023.01.405

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 11 December 2023
DOI 10.3389/fpubh.2023.1259410

Check for updates

Forecasting daily COVID-19 cases with gradient boosted regression trees and other methods: evidence from U.S. cities

Anindya Sen[1]*, Nathaniel T. Stevens[2], N. Ken Tran[3], Rishav R. Agarwal[4], Qihuang Zhang[5] and Joel A. Dubin[2,3]

[1]Department of Economics, University of Waterloo, Waterloo, ON, Canada, [2]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, [3]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada, [4]Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada, [5]Department of Epidemiology, Biostatistics and Occupational Health, McGill College, Montreal, QC, Canada

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Héctor Armando Rincón León,
Mexican Social Security Institute (IMSS), Mexico
Amit Kumar Banerjee,
Indian Institute of Chemical Technology (CSIR), India

*CORRESPONDENCE
Anindya Sen
asen@uwaterloo.ca

RECEIVED 20 July 2023
ACCEPTED 13 November 2023
PUBLISHED 11 December 2023

CITATION
Sen A, Stevens NT, Tran NK, Agarwal RR, Zhang Q and Dubin JA (2023) Forecasting daily COVID-19 cases with gradient boosted regression trees and other methods: evidence from U.S. cities.
Front. Public Health 11:1259410.
doi: 10.3389/fpubh.2023.1259410

Introduction: There is a vast literature on the performance of different short-term forecasting models for country specific COVID-19 cases, but much less research with respect to city level cases. This paper employs daily case counts for 25 Metropolitan Statistical Areas (MSAs) in the U.S. to evaluate the efficacy of a variety of statistical forecasting models with respect to 7 and 28-day ahead predictions.

Methods: This study employed Gradient Boosted Regression Trees (GBRT), Linear Mixed Effects (LME), Susceptible, Infectious, or Recovered (SIR), and Seasonal Autoregressive Integrated Moving Average (SARIMA) models to generate daily forecasts of COVID-19 cases from November 2020 to March 2021.

Results: Consistent with other research that have employed Machine Learning (ML) based methods, we find that Median Absolute Percentage Error (MAPE) values for both 7-day ahead and 28-day ahead predictions from GBRTs are lower than corresponding values from SIR, Linear Mixed Effects (LME), and Seasonal Autoregressive Integrated Moving Average (SARIMA) specifications for the majority of MSAs during November-December 2020 and January 2021. GBRT and SARIMA models do not offer high-quality predictions for February 2021. However, SARIMA generated MAPE values for 28-day ahead predictions are slightly lower than corresponding GBRT estimates for March 2021.

Discussion: The results of this research demonstrate that basic ML models can lead to relatively accurate forecasts at the local level, which is important for resource allocation decisions and epidemiological surveillance by policymakers.

KEYWORDS

daily COVID-19 cases, epidemiological surveillance, Metropolitan Statistical Areas, Gradient Boosted Regression Trees, Seasonal Autoregressive Integrated Moving Average (SARIMA), Susceptible, Infectious, or Recovered (SIR), Linear Mixed Effects

1 Introduction

On May 5, 2023, the World Health Organization (WHO) officially ended the global COVID-19 emergency, referring to increased population immunity, fewer deaths, and reduced pressure on hospitals. The COVID-19 pandemic, which was first declared an international crisis by WHO on January 30, 2020, resulted in severe lockdowns, closure of international borders, devastating economic costs upheaval and the deaths of at least seven million people across the world.[1] Hundreds of published and working research papers have

1 Please see https://www.cbc.ca/news/health/canada-who-pandemic-no-longer-emergency (last accessed June 1st 2023) for further details.

attempted to evaluate the efficacy of traditional statistical models and Machine Learning (ML)/Artificial Intelligence (AI) methods in forecasting COVID-19 cases. Accurate short-term forecasts can inform government decision-making in terms of resource allocation to health practitioners and hospitals, as well as in deciding on the magnitude and severity of lockdowns and timing of re-openings.

The focus on ML and AI methods might be explained by the poor performance of Susceptible-Infected-Removed (SIR) models—traditionally used by epidemiologists to predict the spread of infectious diseases—in forecasting daily COVID-19 counts (1–3). However, many sophisticated statistical methods developed for COVID-19 modeling and forecasting, such as the models from the Institute of Health Metrics and Evaluations, the University of Texas at Austin, and the Los Alamos National Laboratory, have also yielded unsatisfactory results (4). Hence, there is value in identifying alternative models that are relatively easy to implement and interpret, and that are capable of producing accurate predictions. This study evaluates the efficacy of Gradient Boosted Regression Tree (GBRT), Susceptible, Infectious, or Recovered (SIR), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Linear Mixed Effects (LME) models in forecasting daily trends in COVID-19 cases across 25 cities in the U.S. (Albuquerque, Atlanta, Baltimore, Boston, Charlotte, Chicago, Cleveland, Dallas, Denver, Detroit, Houston, Indianapolis, Los Angeles, Louisville, Memphis, Miami, New York, Oklahoma, Phoenix, Pittsburgh, Portland, Sacramento, San Francisco, Seattle, Tampa). We forecast daily COVID-19 case rates one-week and four-weeks ahead over different testing periods. The models chosen are reasonably basic, but this choice is intentional and motivated by the desire to explore the efficacy of simpler models that are relatively easily interpretable and computationally efficient, while also crossing disciplinary boundaries to benchmark the performance of traditional methods employed by researchers across different fields.

The choice of these forecasting models and periods is consistent with other studies. For example, Chumachenko et al. (5) used Random Forest, K-Nearest Neighbors, and Gradient Boosting methods to forecast COVID-19 cases for Germany, Japan, South Korea, and Ukraine with respect to 3, 7, 10, 14, 21, and 30 days. The objective of the study is similar to ours, in terms of assisting public health agencies to identify models that could generate predictions to address various pandemic containment challenges. Krivtsov et al. (6) is a relatively recent example of research based on more complex Neural Network (NN) methods using data for the same countries. Mohammadi and Chumachenko (7) employ ARIMA methods to forecast cases for Ukraine. Other examples of ARIMA based studies at the country level include Dansana et al. (8), Singh et al. (9), and Sahai et al. (10). Devaraj et al. (11) use ARIMA and other deep learning methods. Fang et al. (12) compare the performances for XGBoost and ARIMA specifications using U.S. time-series data and Liu et al. (13) employ ARIMA models to study US national data as well.

Our focus on short-term predictions is guided by the importance of such forecasts for the allocation of local health resources, such as the supply of personal protective equipment (PPE), adequate testing infrastructure, and the availability of hospital care teams. Further, our use of data across several U.S. Metropolitan Statistical Areas (MSAs) or cities is a contribution to the literature as most studies rely on either cross-country, national, state/province, or county level data. While employing aggregated data has benefits, identifying models that are capable of relatively accurate forecasts at the local level can result in more targeted decisions by policymakers. In this respect, we are unaware of any study that has attempted to forecast COVID-19 case counts using a panel of MSAs in the United States. Our study is also a contribution to the literature given that a large fraction of studies, which have attempted to forecast the spread of COVID-19 in the U.S., have not provided any benchmarking of their forecasts against the truth, or stated their limitations (14). Constructing forecasting models at a local level is challenging, given the need to account for unobserved jurisdiction level heterogeneity, and corresponding volatility in daily COVID-19 cases which we observe for several U.S. cities. Such volatility often disappears when data are aggregated across jurisdictions, resulting in a smoother time-series of observations in training datasets, and hence, more accurate predictions. However, such predictions may not be very useful for policymakers interested in epidemic trends within a specific jurisdiction.

A primary objective of our study is to evaluate the performance of traditional SIR models employed by epidemiologists, given findings on the inaccuracy of COVID-19 case forecasts, relative to other methods. A possible explanation behind the poor performance of SIR models is because of the need for properly accounting for relevant geographical characteristics such as the number and distribution of outbreaks, and population size and density (15). The results from SIR modeling are compared to GBRT models intended to evaluate the efficacy of a common ML model. Besides being used by other research, our choice of GBRTs is also motivated by the ease in which models can be implemented by policymakers with limited knowledge of Machine Learning or Artificial Intelligence methods. SARIMA modeling is commonly used for time-series forecasting by economists and is especially useful when the modeled data has pronounced trend and seasonality. LME models are popular with researchers working at the intersection of statistics and health. They enable the researcher to flexibly control for the potentially confounding effects of unit-specific (in this paper, city-specific) heterogeneity through the accommodation of random effects, which themselves accommodate the correlation of within-city repeated measures over time. LME specifications also allow information to be borrowed across geographies, which might result in more accurate predictions, especially for cities that experience considerable within-city variation during the training phase and whose case rates may be spatially correlated with other cities. We do not explore the performance of Neural Network or Long Short-Term Memory (LSTM) models as we restrict our analysis of methods that are relatively straightforward to implement. Finally, given results from other studies [for example, (16–20)], which suggest benefits from the use of ML/statistical models in tandem with social mobility/internet data, we downloaded Apple mobility data for each MSA to evaluate their potential in generating more accurate forecasts.

Citing all papers that have employed such methods in forecasting daily COVID-19 case trends is beyond the scope of this study. Focusing on studies that have employed either county or state/province level data, Altieri et al. (21) and Liu et al. (22) construct one- and two-week-ahead forecasts (of case or death counts) using either different types of linear and exponential predictors or Bayesian methods. Chu and Qureshi (1), Chen et al. (2), Stevens et al. (23), and Sen et al. (24) find different autoregressive time series and ML models to be capable of comparable or superior short-term out-of-sample forecasts of daily cases relative to SIR models. In terms of cities, Wathore et al. (25) rely on deep learning models such as LSTM to forecast cases for 8 cities in India, U.S., and Sweden. Their study contains a summary of other LSTM based papers. Devaraj et al. (11) also contains a detailed discussion of deep learning-based papers. Zhang et al. (26) develop a hybrid predictive model of COVID-19 cases based on autoregressive and LSTM models, which they test for 8 counties in California and some countries.

The limited amount of research exploiting variation across cities over time is probably because of the lack of publicly available data. For U.S. cities, we surmount this difficulty by employing county level data collected by the New York Times and constructing corresponding MSA level daily case counts. As a result, we are able to clearly see patterns and differences in COVID-19 cases across some major U.S. metropolitan areas during the first and second waves of the pandemic. Another contribution of our study is that we produce forecasts during time periods that coincide with the peak of the second wave of infections, specifically during November 15 - December 12, 2020, along with subsequent time periods, which saw significant declines in cases counts for many MSAs. This is in contrast to many early published studies that are focused on forecasts of the first wave and Summer of 2020.

Our choice of this time period is also motivated by the considerable volatility in daily case counts observed across cities during November–December 2020, making the forecasting exercise more challenging. We also conduct forecasting exercises for January–March 2021 as well, to evaluate model efficacy during periods of significant declines in daily case counts. The need to ensure that training data for forecasting models capture dynamic changes in the spread of the virus has been noted by other studies (11, 25, 27). Employing data from these time periods is further justified given the rise in population vaccination rates from March 2021 onwards, and the widespread use of home testing kits from late 2021 onwards, which impacts the reliability of official statistics, given the possibility of under-reporting of positive tests to health authorities.

For most MSAs, GBRT and SARIMA models produce forecasts for November 15–December 12, 2020, with lower Median Absolute Percentage Errors (MAPEs) than corresponding one-week ahead predictions produced by LME and SIR models and are consistent with other studies that find SIR models to produce inaccurate forecasts of the incidence and spread of COVID-19. Apple mobility data do not make a significant difference for the forecast accuracy of SARIMA models. With respect to 7-day ahead forecasts, GBRTs produce MAPEs lower than SARIMA models for most MSAs for the November-December 2020, January, and February 2021 testing periods. On the other hand, SARIMA MAPEs are lower for the March 2021 testing period. Likewise, the 28-day ahead forecasts produced by SARIMA models generate lower MAPE values in March 2021. However, for the other months considered, 28-day ahead GBRT forecasts tend to be associated with lower MAPE values.

2 Methods and materials

2.1 Data

Daily COVID-19 case data at the county level were downloaded from the Github repository maintained by the New York Times[2] We note that the COVID-19 data maintained by the New York Times and John Hopkins University[3] have been widely employed by researchers. Using Federal Information Processing System (FIPS) codes, county-level data were mapped to Metropolitan Statistical Areas (MSAs), which include city cores and adjoining suburbs, to provide daily case counts at the MSA level. We herein use the terms city and MSA interchangeably. Our choice of MSAs was based on investigating COVID-19 trends in the largest cities across the U.S, while ensuring representation across different regions.

Figure 1 shows significant variation in COVID-19 daily cases across MSAs and over time. Some MSAs such as Miami, Phoenix, Oklahoma City, Atlanta, Dallas, Charlotte, Tampa, Houston, San Francisco, and Sacramento had much higher daily case counts per 100,000 of population during July and August of 2020, relative to the first wave in March and April of the same year. In contrast, New York, Boston, and Chicago had much higher case counts per 100,000 of population or per capita daily cases during the first wave. Most cities experienced a decline in COVID-19 cases during September, which was succeeded by a rapid increase during November and December that coincided with intensive campaigning associated with the U.S. Presidential Election. Increases in daily case counts during this time period were succeeded by declines that began sometime during December 2020 or January 2021 for almost all MSAs, and that continued through March 2021.

Five other points should be noted. First, some cities such as Atlanta, Boston, Cleveland, and Houston, have peaks in daily case counts that are clearly anomalous with previous and successive trends. This can be traced to errors in data collection and reporting. The most common explanation we were able to find is that spikes are attributable to the reporting of a significant backlog of cases that, for some reason, were not reported correctly earlier.[4] Second, most cities seem to experience considerable volatility in daily COVID-19 case counts during November and December 2020. Third, days with zero cases reported are likely due to batch reporting in previous days or non-working days and were changed to 0.5 to prevent our performance metrics from having nonsensical results when dividing by zero, as well as to allow us to model the natural logarithm of the counts (specifically, the log of standardized

2 https://github.com/nytimes/covid-19-data

3 https://coronavirus.jhu.edu/map.html

4 See, for example, https://www.click2houston.com/news/local/2020/09/23/explained-why-nearly-15000-new-coronavirus-cases-were-reported-in-the-houston-area-in-one-day/ and https://www.wsbtv.com/news/local/georgia-records-highest-daily-increase-coronavirus-cases-since-bef7IAY2EPRURABPICVYA3PO7AEJ4/.

case counts per 100,000) for the LME and SARIMA models. Fourth, unlike other MSAs, Detroit experienced a significant rise in daily case counts during March 2021. Daily COVID-19 cases also rose modestly during March 2021 in Pittsburgh and Boston. Fifth, across all MSAs, case rates are clearly non-stationary with pronounced trend and seasonality. Effective modeling and forecasting should account for this correlation structure. Table 1 contains means and standard deviations of standardized daily case counts per 100,000 for each MSA for our training period of March 22, 2020–November 14, 2020.

2.2 Methods

We begin by introducing notation that will be used for all the different methods introduced in this section. Let $t = 0, 1, 2, \ldots, T$ represent time, which here we take to be measured in days. Thus, for a given analysis of an MSA's daily case counts, we will observe a maximum of $T + 1$ time points. Each city is represented by the index $i = 1, 2, \ldots, N$, where $N = 25$ in this paper's analyses. The response variable, denoted by $Y_{i,t}$, is documented COVID-19 case counts per 100,000 people, standardized by the city population size. In particular,

$$Y_{i,t} = (\#\text{ new cases in city } i \text{ on day } t/\text{Population size of city i}) \times 100,000$$

represents the per capita daily case count (per 100,000 people) in city i at time point t. Note that this standardization is necessary to ensure that case rates are comparable across cities of different sizes. For models where lagged case values are included as predictor variables, the earliest first day used for the response will be constrained by the number of lagged terms; for example, if we include predictors Y_{t-1} and Y_{t-2}, then the first response in the model will be at $t = 2$. In our models specified in the following subsections, the first date to be used for Y_t is March 22, 2020, since

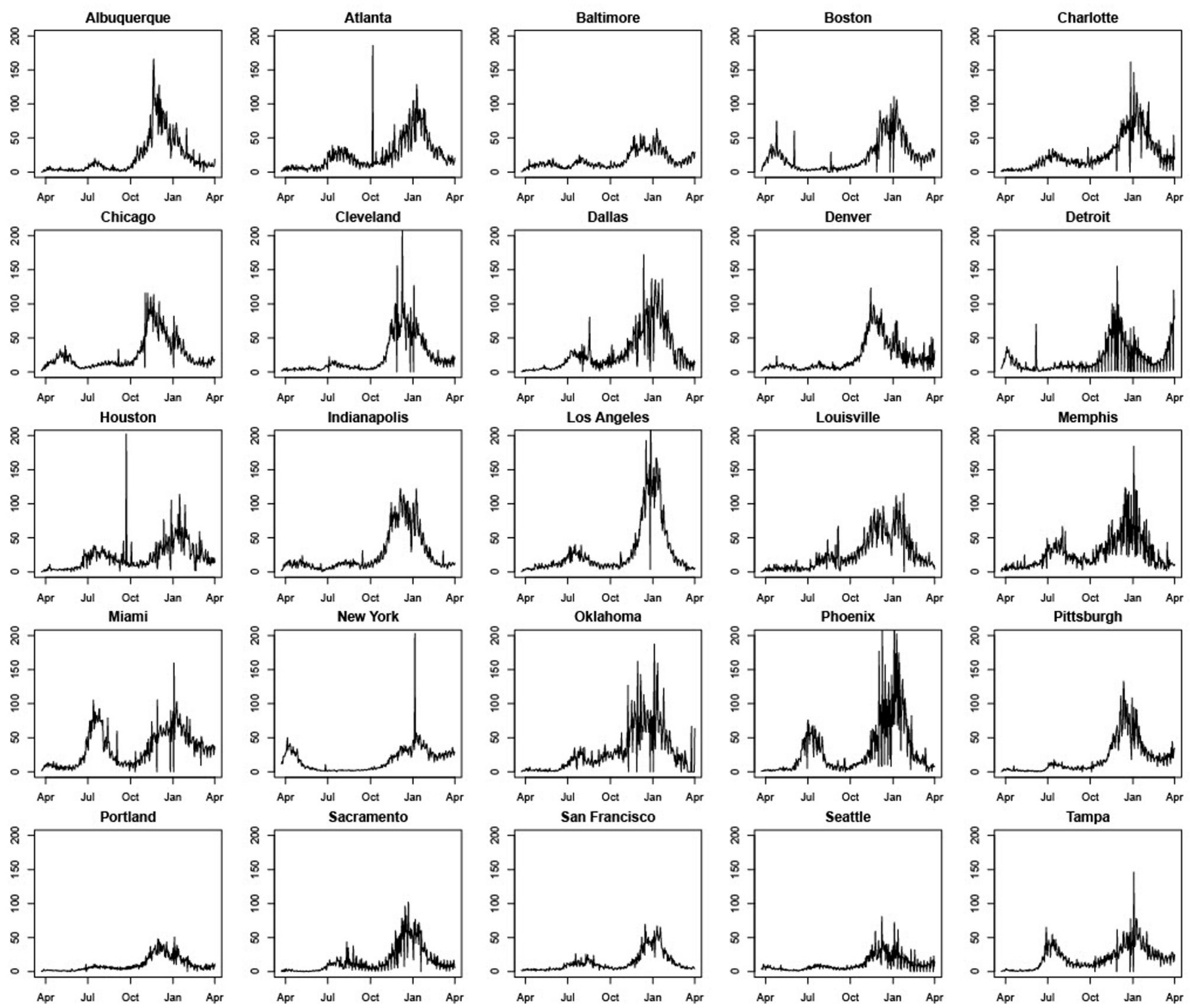


FIGURE 1
Standardized daily case counts per 100,000 people in 25 U.S. cities (March 22, 2020–March 31, 2021).

TABLE 1 Sample statistics.

Sample mean (s.d.) of daily COVID-19 standardized cases per 100K (March 22–Nov 14, 2020)				
Albuquerque	Atlanta	Baltimore	Boston	Charlotte
9.65	14.36	10.98	11.69	13.50
(12.53)	(14.69)	(5.47)	(11.02)	(8.41)
Chicago	Cleveland	Dallas	Denver	Detroit
18.84	8.59	14.59	13.03	11.74
(19.39)	(9.81)	(11.77)	(16.58)	(12.71)
Houston	Indianapolis	Los Angeles	Louisville	Memphis
14.30	13.52	12.81	15.17	19.54
(15.95)	(12.82)	(7.97)	(15.02)	(13.49)
Miami	New York	Oklahoma	Phoenix	Pittsburgh
23.81	8.45	15.42	16.06	6.49
(23.42)	(10.05)	(15.54)	(17.72)	(6.47)
Portland	Sacramento	San Francisco	Seattle	Tampa
5.52	7.16	7.06	6.23	12.87
(5.06)	(7.18)	(4.85)	(4.54)	(12.18)

this is the date that we start to see (at least some) non-zero case counts in all 25 cities.

As discussed in Section 1, some studies have used social media or cellular data to model population mobility and forecast the incidence and spread of COVID-19. Therefore, we use three separate social mobility indices released by Apple, that measure walking, driving, and transit use, respectively, for certain geographies. While most studies have relied on Google mobility data, Google does not offer publicly available data at the MSA level. The Apple data are collected from Apple Maps app usage through individual iPhones. The indices show changes in relative volume of directions requests per country/region, sub-region, or city compared to a baseline volume on January 13th, 2020. While we do not claim these indices necessarily capture population-wide mobility trends, using these data at least allows us to evaluate the effects of social mobility on daily COVID-19 cases counts, and assess the usefulness of this information in constructing daily predictions of COVID-19 cases in MSAs. Moreover, the walking mobility indicator is highly correlated with temperature and therefore serves as a useful proxy for weather, which is known to be an important factor in transmission (28).

In terms of quantifying the forecasting performance of different models, we consider the Median Absolute Error (MAE) and Median Absolute Percentage Error (MAPE) of predicted daily case counts for each Metropolitan Statistical Area. The MAE is the median of the absolute value of the difference between actual and predicted daily case counts over the chosen testing period, while the MAPE is the median of the absolute value of the corresponding percentage difference. While MAEs are reported, our comparisons are primarily based on the MAPE as it is a standardized metric interpreted equivalently across all cities. Hence, for a city i, if the actual observed daily case count is Y_t and the predicted value from a chosen model is $\hat{Y}_t$, the MAE is the median absolute value of their difference or $\left|Y_t - \hat{Y}_t\right|$, calculated from daily values over the testing period. The MAPE is therefore the median of $\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right| \times 100$ over the testing sample. For further sensitivity analysis, we also report the proportion of daily forecasts with an absolute percentage forecast error $<20\%$.

2.2.1 Linear Mixed Effects (LME) models

To assess the statistical importance of Apple mobility indicators, we ran LME regressions using $Y_{i,t}$ (i.e., standardized case count per 100,000 for city i at time j) as the dependent variable and based on data pooled across cities and over days for the training period. As a sensitivity analysis, we also ran the LME regressions for the time period March 22 - August 31, 2020, before the beginning of the steady rise in case counts for most cities. The LME models considered contain the seven-day lag of the walking index, one and two day lags of the dependent variable along with a weekend indicator variable.[5] Estimation results are reported in Table 2. First, while these specific results are not reported, we note that lagged dependent variables of a higher order than 3 days were statistically insignificant in these models. Second, one and two-day lags in the dependent variable are statistically significant at the 1% level. Third, the seven-day

5 Only the Walking Index was employed in these specifications because of a high correlation with the Driving Index ($r = 0.85$) and the fact that the Transit Index was consistently insignificant in preliminary regressions that contained all three mobility indicators. More complicated forms of autocorrelation are handled differently in the various models, but generally through the inclusion of lagged case rates as predictor variables.

TABLE 2 Baseline linear mixed effect model summaries (standard errors in parentheses).

	Mar 22–Aug 31, 2020	Mar 22–Nov 14, 2020	Mar 22–Aug 31, 2020 with Apple Mobility	Mar 22–Nov 14, 2020 with Apple Mobility
Estimation results				
Log cases$_{t-1}$	0.49 (0.018)	0.444 (0.012)	0.489 (0.019)	0.435 (0.013)
Log cases$_{t-2}$	0.38 (0.015)	0.407 (0.011)	0.371 (0.015)	0.401 (0.012)
Seven day lag Apple Walking Mobility index			0.001 (0.0003)	0.002 (0.0002)
Weekend dummy	Yes	Yes	Yes	Yes
AIC	7,137.503	13,607.74	7,136.557	13,582.3
Log-likelihood	−3,560.75	−6,795.87	−3,559.28	−6,782.16

lagged Apple mobility Walking Index is also statistically significant at the 1% level. These results offer some support to the inclusion of a mobility index in the models used to forecast daily case counts. On the other hand, we note that the inclusion of the Apple mobility index does not significantly improve the fit of the LME models as measured through the AIC and Log-Likelihood, and so its inclusion may not drastically improve a model's forecasting capability.

In terms of the specific LME model:

$$\log(Y_{i,t}) = \beta_0 + b_{0,\mathrm{i}} + \beta_1 \log(Y_{i,t-1}) + \beta_2 \log(Y_{i,t-2}) + \beta_3 WE_t + \varepsilon_{i,t}, \quad (1)$$

where:

- β_0 is the population-level intercept;
- $b_{0,i}$ is a city-specific random intercept, with $b_{0,i} \sim \mathrm{N}(0^2, \sigma_{b0})$;
- β_1 is the fixed-effect parameter connected with one-day lagged (log) count for city i;
- β_2 is the fixed-effect parameter connected with two-day lagged (log) count for city i;
- β_3 is the fixed-effect parameter connected with the weekend indicator, common to all cities. Specifically, if day t falls on a weekend, then $WE_\mathrm{t} = 1$, otherwise $WE_\mathrm{t} = 0$;
- $\varepsilon_{i,t}$ is the model error term, with $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})^t \sim \mathrm{N}_T(\mathbf{0}, \sigma^2 \mathbf{I}_T)$, and $\boldsymbol{\varepsilon}_i$ independent of $b_{0,i}$.

Note we have initially assumed that the model error terms are conditionally independent, with the random intercept being the only term to induce correlation between the repeated measures of daily counts within the same city. We also see that all cities have a common T repeated measures observed. We found that a modified version of (1) provides a better fit to the standardized city-level daily COVID-19 case count data. Specifically, allowing for heterogeneity in both time-varying predictors, i.e., attaching city-specific random effects to each, and properly handling existing heteroskedasticity among the model error terms was also required, leading to:

$$\log(Y_{i,t}) = \beta_0 + b_{0,\mathrm{i}} + (\beta_1 + b_{1,\mathrm{i}}) \log(Y_{i,t-1}) + (\beta_2 + b_{2,\mathrm{i}}) \log(Y_{i,t-2}) + \beta_3 WE_t + \varepsilon_{i,t}, \quad (2)$$

where the changes relative to (1) are:

- $b_{1,i}$ is a city-specific random effect connected to the first-order lagged (log) counts, with $b_{1,i} \sim \mathrm{N}(0^2, \sigma_{b1})$;
- $b_{2,i}$ is a city-specific random effect connected to the second-order lagged (log) counts, with $b_{2,i} \sim \mathrm{N}(0^2, \sigma_{b2})$; we allow for all three random effects to be correlated, though each are independent of $\boldsymbol{\varepsilon}_i$;
- $\boldsymbol{\varepsilon}_i \sim \mathrm{N}_T(\mathbf{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ diagonal, but weighted in such a way to remove the original heteroskedasticity in the model error terms.

From (2), we can forecast one time point ahead (e.g., one day ahead), then use that forecast as the first-order lagged term to forecast one additional day ahead (where the prior first-order lagged term is now the second-order lagged term), and so on, until we have forecasted forward the desired number of days. Note that linear mixed effect models have units borrow strength from other units for a given unit's trajectory prediction. Though not all cities are aligned in time in terms of model dynamics, not nearly as much borrowing of strength from other city's predictions will affect a given city's predictions, due to the extensive collection of repeated measures per city paired with the relatively reasonable within-city variation. Based on the results in Table 2, the Apple Walking Index is employed as a predictor for some specifications. Estimation of LME models and forecasts were conducted using the *lme4* package from the R programming language.

2.2.2 Susceptible-Infected-Removed (SIR) model

SIR models are the dominant methodology to model the spread of epidemics; see, for example, Tolles and Luong (29) for further details. The SIR model uses differential equations to describe the dynamic status of an individual switching between three compartments in a population at time t: susceptible (S_t), infected (I_t), and removed (R_t) (including recovered and deceased individuals) and is a standard approach employed by epidemiologists to forecast disease spread. N_t is the total population at time t and is identified by $N_\mathrm{t} = S_\mathrm{t} + I_\mathrm{t} + R_\mathrm{t}$, where β is the average number of contacts per infectious person per time unit, γ is the transition rate from I_t to R_t, and R_t includes recovered and deceased individuals. The SIR model is then given by the following ordinary differential equations:

$$\frac{\partial S_t}{\partial t} = -\frac{\beta I_t S_t}{N} \quad (3)$$

$$\frac{\partial I_t}{\partial t} = \frac{\beta I_t S_t}{N} - \gamma I_t \quad (4)$$

$$\frac{\partial R_t}{\partial t} = \gamma I_t \quad (5)$$

where β is the average number of contacts per infectious person per time unit and γ is the transition rate from I_t to R_t. While the SIR model is based on the modeling of S_t, I_t, and R_t, our focus here is the daily infection numbers and aims to conduct predictions such that the prediction errors are minimized. With the daily number of confirmed cases on day t considered as the difference of S_t and S_{t-1}, we calculate the predicted number of daily confirmed cases on day t, denoted $\hat{Y}_t(\beta, \gamma)$, as follows:

$$\hat{Y}_t(\beta, \gamma) = S_{t-1} - S_t = I_t - I_{t-1} + R_t - R_{t-1} \quad (6)$$

for $t = 1, \ldots, T$, where T represents the end of the study period. We implicitly assume that N, β, and γ are constant over time.

Then, the parameters β and γ can be obtained by minimizing the squared prediction error

$$PE(\beta, \gamma) = \sum_{t=1}^{T} \{Y_t - \hat{Y}_t(\beta, \gamma)\}^2 \quad (7)$$

for β and γ. Y_t is the total number of daily cases in a province at time t and is not in per capita terms. We convert the forecasts into per capita terms. Our forecasting procedures model predictions of total daily cases as these are the numbers that are reported and are of interest to policymakers in tracking the incidence and spread of COVID-19. This is consistent with the approach employed by Chen et al. (2). Estimation of the SIR models and construction of forecasts were done using the R programming language. Specifically, we utilized the *EpiDynamics* and *bbmle* packages to estimate the parameters of the SIR model.

2.2.3 Box-Jenkins time series modeling

In this paper we also employ the Box-Jenkins class of time series models referred to herein as SARIMA (Seasonal Autoregressive Integrated Moving Average) models and we use the notation $\log(Y_{i,t}) \sim \text{SARIMA}(p, d, q)(P, D, Q)[m]$, where, again, $Y_{i,t}$ is the standardized case count per 100,000 in city i on day t. Thus, for each city $i = 1, 2, \ldots, N$, we separately fit a model of the form

$$\phi(B)\Phi(B^m)^D(1-B)^d log(Y_{i,t}) = \theta(B)\Theta(B^m)^D \varepsilon_{i,t}, \quad (8)$$

where B is the backshift operator[6] and m is the period of the seasonal component, which here we take to be 7 given the daily data. We also assume that $\varepsilon_{i,t} \overset{iid}{\sim} N(0, \sigma^2)$ and we define $\phi(z)$, $\theta(z)$, $\Phi(z)$, and $\Theta(z)$ respectively to be the following p^{th} order, q^{th} order, P^{th} order, and Q^{th} order polynomials:

$$\phi(z) = 1-\phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$$
$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_p z^q$$
$$\Phi(z) = 1-\Phi_1 z - \Phi_2 z^2 - \ldots - \Phi_p z^P$$
$$\Theta(z) = 1 + \Theta_1 z + \Theta_2 z^2 + \ldots + \Theta_p z^Q$$

Simple algebra shows that the response for city i on day t, $\log(Y_{i,t})$, is therefore a weighted sum of historical log case counts. Note that different values of the non-seasonal (p, d, q) and seasonal (P, D, Q) orders give rise to different configurations of the model, accounting for different forms of correlation structure in the observed time series. We choose the values of p, d, q, P, D, Q that minimize the corrected Akaike information criterion to ensure the model fits the observed data well. The models themselves are fit using maximum likelihood estimation. As with the LME model, the 7-day lag of the Apple Walking Index is also employed as an exogenous variable in some specifications. These time-series models were estimated employing the *forecast* package in the R programming language.

2.2.4 Gradient Boosted Regression Trees (GBRTs)

Gradient Boosted Regression Trees are a commonly used ML algorithm based on decision trees to produce forecasts of particular outcomes. The algorithm sequentially tests the predictive power of different trees to reduce forecast errors, until no further improvements can be made. These predictions are then combined through a weighted average of regressions to produce a final prediction. Although there are other more sophisticated ML methods that can be employed to generate predictions, our choice of GBRTs was motivated by their relative simplicity and interpretability, as well as their ease of implementation through software such as R and Python. This makes them an attractive choice for policymakers with limited resources and fairly limited experience with machine learning.

We assume again our sample contains $T + 1$ observations for each MSA i. In particular, we assume that we observe the response variable $Y_{i,t}$ (standardized case count per 100,000 for city i at time j) and a vector of predictors given by $\mathbf{X}_{i,t}$ for $i = 1, 2, \ldots, N$ and $t = 0, 1, 2, \ldots, T$. The model that forecasts $Y_{i,t}$ based on $\mathbf{X}_{i,t}$ is a weighted additive model of the form

$$Y_{i,t} = \sum_{k=1}^{K} \alpha_k f_k(\mathbf{X}_{i,t}) + \varepsilon_{i,t}, \quad (9)$$

where $f_k(\cdot)$ for $k = 1, 2, \ldots, K$ are regression trees, α_k are weights, and $\varepsilon_{i,t}$ is an error term. The algorithm estimates both the weights α_k for $k = 1, 2, \ldots, K$ and $f_k(\cdot)$ by sequentially minimizing a penalized differentiable convex loss function of $Y_t - \sum_{k=1}^{K} \alpha_k f_k(\mathbf{X}_k)$ with respect to both α_k and $f_k(\cdot)$ over K boosting iterations for $k = 1, 2, \ldots, K$. These estimates can then be used to produce forecasts of $Y_{i,t}$. For the $\mathbf{X}_{i,t}$ variables, we employ: the weekend dummy variable; city-specific fixed effects; one- to seven-day lags of the dependent variable; and seven- to ten-day lags of all the Apple mobility variables. GBRTs enable the use of all Apple mobility variables and more lagged dependent variables without concern of collinearity issues, as weak predictor variables are given less weight in constructing forecasts. The *XGBoost* package in Python was used

6 B is a notational operator that shifts subscripts t back by 1: $BY_t = Y_{t-1}$. And when raised to the k^{th} power, it shifts subscripts t back by k: $B^k Y_t = Y_{t-k}$.

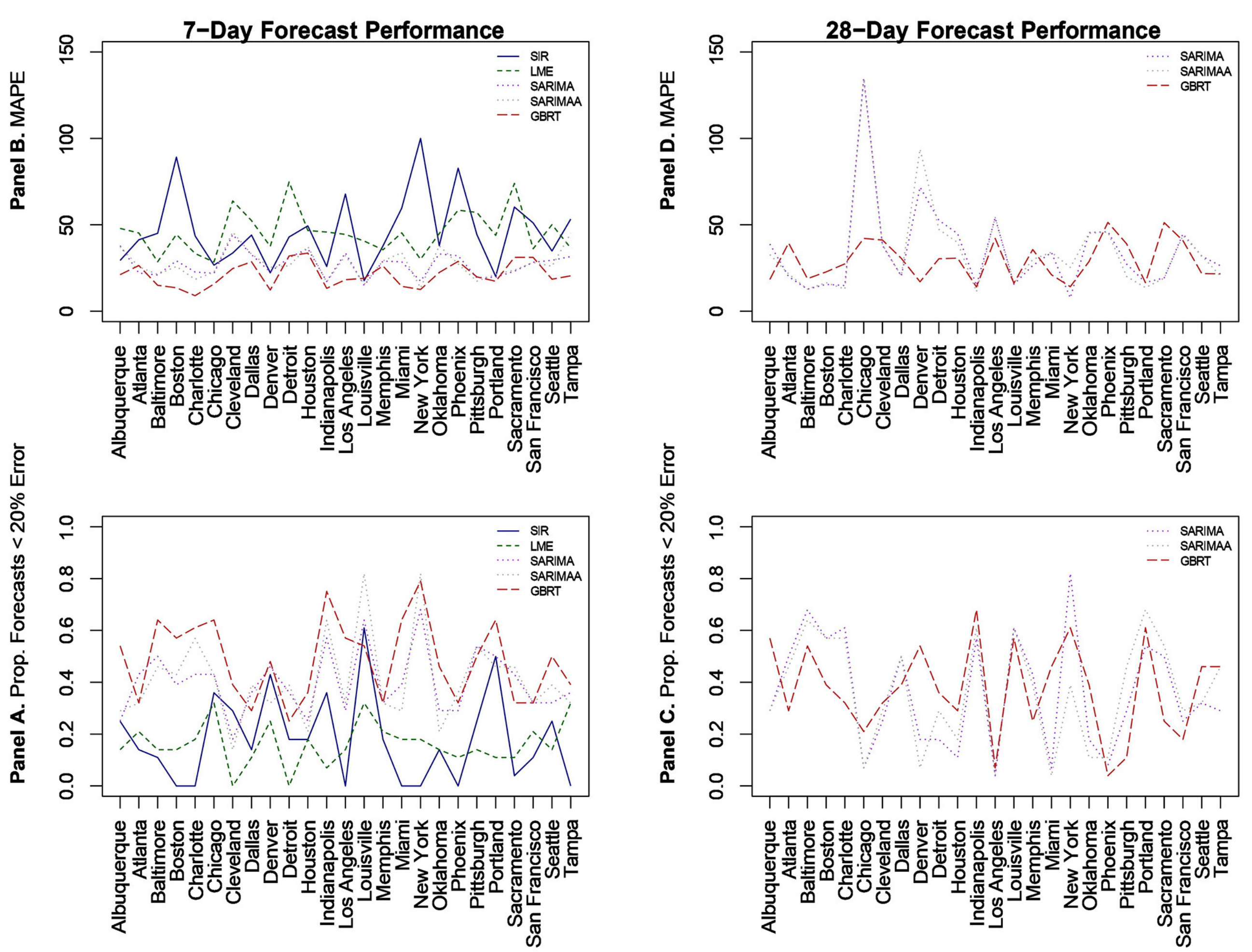


FIGURE 2
Forecast performance as a function of MSA for 7-day **(A, B)** and 28-day **(C, D)** forecasting strategies for November 15-December 12, 2020. **(B, D)** Quantify forecast performance by MAPE (smaller values are better); **(A, C)** do so using the proportion of daily predictions with a forecast error lower than 20% (larger values are better).

to estimate the GBRT models and construct predictions. Code for all of the above models and algorithms are available upon request.

3 Results

We begin with predictions that are 1 week or 7-day ahead forecasts for each of the 25 MSAs for November 15–December 12, 2020 based on all the models. Given the poor performance of SIR and LME models, we then evaluate the sensitivity of our findings by constructing forecasts from SARIMA and GBRT models for different time periods in January, February, and March of 2021. Figure 2 visualizes MAPE values and the proportion of daily predictions with a forecast error lower than 20% for daily standardized COVID-19 cases (per 100,000 of population) based on LME, SIR, SARIMA, SARIMA models with the Apple Walking Index (referred to as SARIMAA), and GBRTs for each MSA. Note the values displayed in these plots are also tabulated in Table 3. In Figure 2, Panels A and C report the proportion of daily predictions with a forecast error lower than 20% with respect to 7-day and 28-day forecast performances, respectively. Panels B and D contain corresponding visualizations for MAPE values based on 7-day and 28-day forecast errors. For Panels B and D we seek to identify the methodology that has the lowest average MAPE values across cities. On the other hand, for Panels A and C we are interested in the methodology which has the highest average of the proportion of daily predictions with a forecast error lower than 20%, across cities.

3.1 LME and SIR 7-day ahead

With respect to 7-day ahead predictions, LME models produce lower MAPEs for 9 MSAs, relative to SIR models. However, the LME MAPEs are also quite high, with Baltimore having the lowest one at 28.63%. SIR models also do not offer consistently accurate forecasts across cities. With the exception of Denver, Louisville, and Portland, SIR MAPE values are always above 25% for such models. Further, SIR models perform poorly on the basis of the proportion of daily forecasts falling within 20% of

10.3389/fpubh.2023.1259410

TABLE 3 7-day forecasting performance from LME, SIR, and SARIMA models November 15-December 12, 2020.

City	LME			SIR			SARIMA			SARIMAA			GBRT		
	MAE	MAPE	Prop <20% FE	MAE	MAPE	Prop <20% FE	MAE	MAPE	Prop <20% FE	MAE	MAPE	Prop <20% FE	MAE	MAPE	Prop <20% FE
Albuquerque	43.97	47.93	0.14	25.91	29.31	0.25	28.12	37.97	0.25	32.79	37.21	0.29	18.4	21.69	0.46
Atlanta	13.89	45.27	0.21	12.72	41.31	0.14	6.59	22.2	0.43	8.99	25.63	0.32	9.79	21.66	0.43
Baltimore	9.3	28.63	0.14	18.71	45.09	0.11	7.08	21.09	0.5	7.1	21.18	0.46	6.18	18.12	0.57
Boston	19.01	44.44	0.14	33.23	89.18	0	12.02	29.27	0.39	9.93	25.78	0.43	6.21	15.72	0.64
Charlotte	14.67	33.43	0.18	19.29	43.59	0	9.51	22.8	0.43	8.85	18.38	0.57	5.66	12.61	0.64
Chicago	22.26	28.47	0.32	20.03	26.61	0.36	13.99	22.01	0.43	14.17	22.47	0.43	11.81	14.45	0.61
Cleveland	41.95	63.77	0	22.28	33.65	0.29	33.32	44.39	0.18	33.38	45.73	0.14	19.85	27.21	0.39
Dallas	27.69	52.43	0.11	20.67	44.03	0.14	17.89	32.81	0.36	17.65	32.84	0.39	17.6	30.49	0.36
Denver	26.57	37.55	0.25	15.43	22.19	0.43	17.13	22.81	0.46	18.29	25.31	0.32	10.76	13.81	0.68
Detroit	43.82	74.78	0	28.71	42.99	0.18	20.33	31.05	0.36	13.64	26.16	0.39	25.04	39.86	0.18
Houston	11.51	46.72	0.18	14.1	49.37	0.18	11.08	36.37	0.25	11.15	37.35	0.21	9.1	33.18	0.36
Indianapolis	38.5	45.82	0.07	25.18	25.84	0.36	15.7	17.32	0.57	14.03	16.85	0.64	6.89	8.77	0.71
Los Angeles	19.88	44.47	0.14	31.99	67.81	0	14.56	33.86	0.29	14.81	32.15	0.32	10.76	22.53	0.39
Louisville	28.24	40.63	0.32	12.21	17.63	0.61	10.03	15.03	0.64	9.77	15.42	0.82	8.95	11.99	0.64
Memphis	15.07	35.55	0.21	18.69	37.78	0.18	14.02	29.11	0.32	14.15	29.03	0.32	15.64	28.53	0.36
Miami	23.62	45.33	0.18	32.43	59.62	0	15.35	28.69	0.39	17.83	33.77	0.29	11.6	22.33	0.43
New York	6.53	30.08	0.18	25.95	99.97	0	4.28	17.56	0.68	3.42	13.41	0.82	3.27	14.04	0.75
Oklahoma	31.79	44.99	0.14	26.61	37.92	0.14	25.81	33.28	0.29	26.26	37.77	0.21	16.72	25.09	0.32
Phoenix	34.4	58.45	0.11	42.71	82.78	0	17.74	32.23	0.29	16.28	27.28	0.36	18.23	32.29	0.32
Pittsburgh	28.53	57.17	0.14	21.81	44.38	0.25	9.83	19.45	0.54	9.86	17.32	0.54	10.4	21.2	0.43
Portland	14.38	43.91	0.11	6.2	20.05	0.5	5.46	19.77	0.5	6.15	21.26	0.46	4.9	15.83	0.57
Sacramento	27.5	73.97	0.11	20.56	60.25	0.04	9.14	23.27	0.43	9.27	23.72	0.46	12.75	27.41	0.32
San Francisco	7.13	36.19	0.21	9.87	51.06	0.11	5.8	28.45	0.32	5.13	28.16	0.32	6.88	5.97	30.58
Seattle	15.39	50.07	0.14	9.89	34.83	0.25	8.44	29.59	0.32	7.4	25.92	0.39	6.22	20.1	0.5
Tampa	12.3	36.56	0.32	15.69	53.29	0	9.59	31.74	0.36	11.48	44.97	0.32	5.63	19.52	0.57

TABLE 4 28-day forecasting performance from GBRT and SARIMA models with and without Apple Mobility November 15- December 12, 2020.

City	GBRT			SARIMAA			SARIMA		
	MAE	MAPE	Prop <20% FE	MAE	MAPE	Prop <20% FE	MAE	MAPE	Prop <20% FE
Albuquerque	17.01	18.41	0.57	31.17	32.77	0.29	34.61	38.91	0.29
Atlanta	11.9	39.49	0.29	6.59	21.61	0.46	6.46	20.07	0.5
Baltimore	7.48	18.86	0.54	4.34	12.79	0.64	4.33	12.79	0.68
Boston	9.39	22.77	0.39	8.04	16.56	0.57	7.19	15.56	0.57
Charlotte	11.22	27.48	0.32	5.07	12.84	0.57	5.62	14.86	0.61
Chicago	30.41	42.21	0.21	88.57	134.28	0.07	88.29	135.05	0.07
Cleveland	29.51	41.2	0.32	32.69	38.41	0.29	32.45	38.17	0.25
Dallas	19.71	30.59	0.39	11.22	20.43	0.5	11.22	20.4	0.5
Denver	13.83	17.08	0.54	72.34	93.8	0.07	57.8	71.89	0.18
Detroit	19.8	30.41	0.36	25.42	47.38	0.29	36.66	52.67	0.18
Houston	8.8	30.78	0.29	11.4	39.78	0.18	11.82	44.9	0.11
Indianapolis	11.99	14.09	0.68	10.63	11.38	0.61	12.84	14.99	0.57
Los Angeles	18.97	42.17	0.07	26.59	54.23	0.04	26.15	54.11	0.04
Louisville	10.59	16.21	0.57	10.79	16.29	0.61	10.75	15.68	0.61
Memphis	20.26	35.76	0.25	14.9	29.93	0.39	13.73	26.5	0.43
Miami	11	21.13	0.46	17.01	34.17	0.04	18.27	34.01	0.07
New York	3.43	14.21	0.61	5.66	25.32	0.39	1.79	7.73	0.82
Oklahoma	22.32	28.81	0.39	33.3	45.14	0.11	34.27	45.5	0.18
Phoenix	26.42	51.49	0.04	20.12	45.83	0.11	22.38	46.04	0.08
Pittsburgh	21.08	39.06	0.11	10.9	20.19	0.46	14.11	27.3	0.29
Portland	5.42	16.52	0.61	4.46	13.7	0.68	4.86	16.61	0.54
Sacramento	19	51.31	0.25	6.23	19.18	0.54	4.33	19.37	0.5
San Francisco	7.8	40.96	0.18	8.58	43.84	0.29	8.65	44.77	0.25
Seattle	6.73	21.82	0.46	10.99	32.15	0.32	10.62	31.88	0.32
Tampa	5.23	21.65	0.46	5.88	20.36	0.46	7.52	26.29	0.29

the actual value, as there are 7 cities for which the value of this measure is zero. It is also important to note the extremely high MAPE values of approximately 90% and 100% for Boston and New York, respectively, as these two cities experienced extremely high per capita daily case counts during the first wave of infections. Further, the MAPE results indicate that even the use of a long training period is not sufficient to enable the SIR models to acknowledge the subsequent downward trend in daily infections and readjust to generate more accurate forecasts for the testing period.

3.2 SARIMA 7-day ahead

Relative to SIR and LME forecasts, SARIMA models produce forecasts with MAPEs that are lower for the majority of cities. SARIMA forecasts yield MAPEs smaller than 20% for four cities and smaller than 25% for six more cities. Another observation is that SARIMA modeling generates predictions that are within tighter bounds for almost each city. Specifically, there is only one MSA (Cleveland) with a MAPE >40% with SARIMA modeling. On the other hand, the number of MSAs with MAPEs >40% based on LME and SIR models are 17 and 15, respectively. Another measure that supports the notion that SARIMA models produce superior forecasts to SIR and LME methods, is the fact that no MSA has a proportion of absolute forecast errors lower than 20% equal to zero, based on SARIMA predictions. Further, for 11 MSAs the proportion of such observations, generated from SARIMA models, is >0.4. Based on these findings, we conclude that SARIMA is a superior forecasting strategy for the vast majority of MSAs. Employing the Apple Walking Index does not make much of a difference in SARIMA forecasts. As we can see in Figure 2, with the exception of a few MSAs, SARIMA MAPEs with and without the Apple Walking Index are comparable. For seventeen cities the difference between MAPEs is <3 percentage points. The proportion of daily forecasts falling within 20% of the actual values are also quite comparable between SARIMA forecasts with and without Apple mobility data.

10.3389/fpubh.2023.1259410

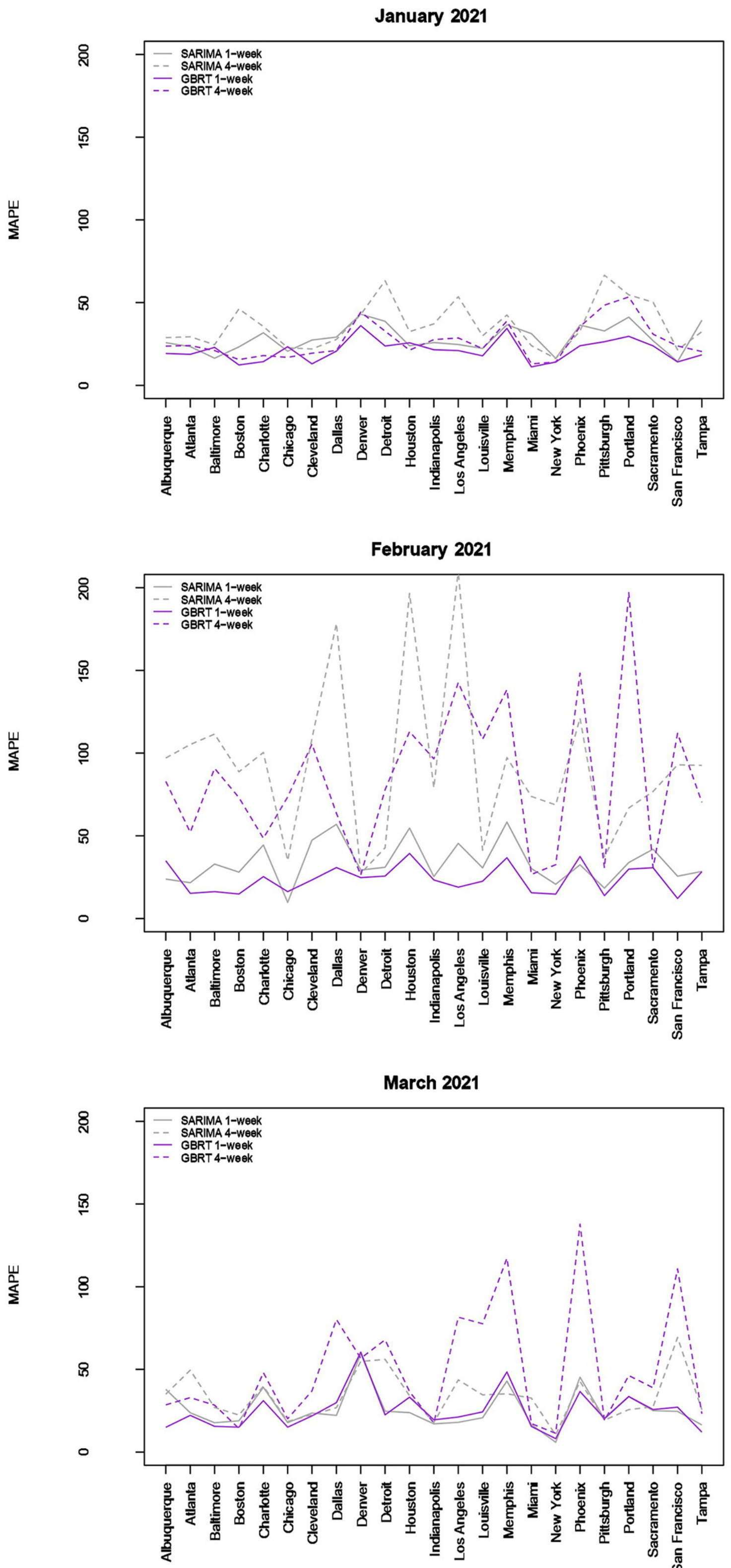


FIGURE 3
Forecast performance as a function of MSA for January, February, and March 2021. Forecast performance is quantified by MAPE for two different model types and two different forecasting strategies.

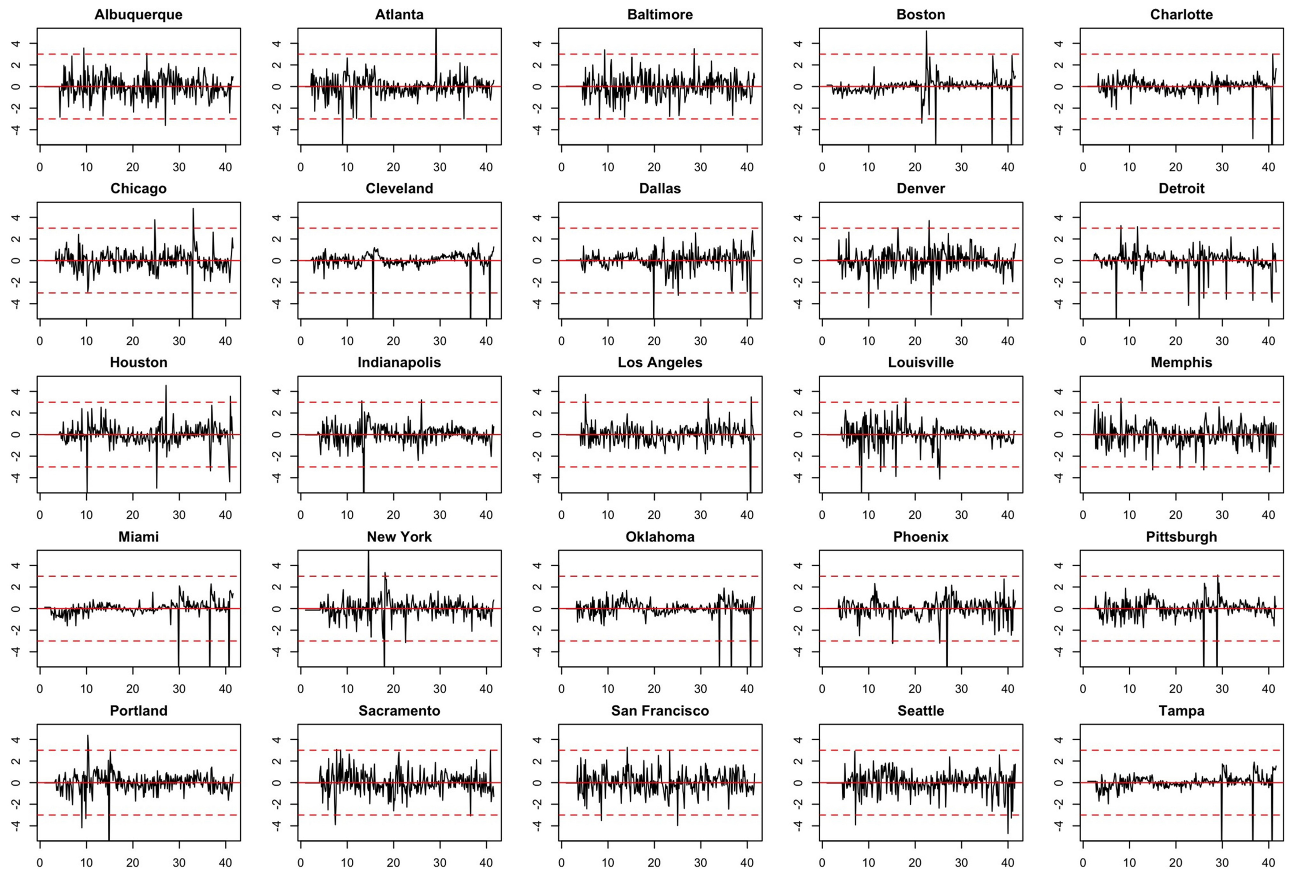


FIGURE 4
Model residuals associated with the SARIMA model fit to March 2020-December 2020 data.

3.3 GBRT 7-day ahead

For most MSAs, GBRTs produce more accurate forecasts of the methods considered. Specifically, GBRTs generate predictions with lower MAPEs for 17 MSAs, relative to corresponding forecasts from SARIMA models with and without the Apple Mobility Index. Eleven MSAs have GBRT forecasts with MAPEs lower than 20%, with Baltimore, Boston, Charlotte, Chicago, Denver, Indianapolis, Miami, and New York having the lowest values. These results suggest that GBRTs are capable of producing relatively accurate 7-day ahead forecasts even during time periods of steep increases and some volatility in daily case counts. The GBRT results are driven by the lagged dependent variables, and not the Apple mobility variables. Specifically, the feature importance scores of the one- to five-day lagged dependent variables range from 0.01 to 0.29. Among the Apple mobility indicators, only the 8-day lagged Driving Index has a feature score above zero, but which is still low at 0.01. These results are consistent with the satisfactory performance of SARIMA specifications without the Apple Walking Index for some MSAs, as the predictors in these models are lagged dependent variables.

3.4 SARIMA, SARIMAA, and GBRT 28-day ahead

Based on the noticeably poorer performance of the LME and SIR models, the 28-day ahead forecasts (Figures 2C, D) are based on SARIMA, SARIMAA, and GBRT models only. The predictions are constructed from a single sequence of 28-day ahead forecasts (without model updating) for each of the 25 MSAs for November 15–December 12, 2020. Note the values displayed in these plots are also tabulated in Table 4. Here, we note that SARIMA models produce lower MAPEs for a majority of MSAs, as there are 14 MSAs for which SARIMA MAPEs are lower than corresponding GBRT values. However, when the SARIMA MAPEs are larger than the GBRT MAPEs, they are often *much* larger, and when SARIMA MAPEs are smaller than GBRT MAPEs, they tend only to be *marginally* smaller. Consequently, across all MSAs the average SARIMA MAPE is *higher* than the corresponding GBRT value, despite individual SARIMA MAPEs being smaller for a larger number of cities. As such, GBRTs appear to be the best choice for forecasting both 7-day ahead and 28-day ahead daily standardized COVID-19 case counts for November 15–December 12, 2020.

TABLE 5 MAPE values—7-day- and 28-day forecasts from SARIMA models January—March 2021.

City	January 1–28 2021		February 1–28 2021		March 1–28 2021	
	7-day	28-day	7-day	28-day	7-day	28-day
Albuquerque	25.78	28.84	24	97.09	37.89	34.88
Atlanta	23.5	29.37	21.82	105.21	23.73	49.47
Baltimore	16.49	24.58	33.04	111.57	17.67	26.88
Boston	23.32	46.21	28.2	88.76	18.85	22.33
Charlotte	31.8	35.54	44.53	100.47	39.12	39.82
Chicago	20.68	22.69	9.91	35.04	17.8	18.46
Cleveland	27.41	21.98	47.47	108.83	23.59	22.42
Dallas	29.13	27.76	57.02	178.42	22.12	26.81
Denver	42.92	42.75	29.45	27.35	59.24	54.72
Detroit	38.63	63.27	31.11	42.85	24.69	56
Houston	24.01	32.47	54.79	196.61	23.84	34.01
Indianapolis	25.9	37.15	25.49	79.09	17.07	17.75
Los Angeles	24.75	53.61	45.55	210.15	17.94	43.55
Louisville	22.38	29.89	30.8	41.36	20.7	34.48
Memphis	36.55	42.52	58.42	97.32	42.85	35.16
Miami	31.29	23.99	30.34	73.81	16.62	32.52
New York	16.44	16.18	20.92	68.7	5.87	10.22
Phoenix	36.39	32.84	32.62	121.21	45.25	42.64
Pittsburgh	32.89	66.6	18.56	37.4	19.95	19.22
Portland	41.2	54.7	33.94	66.91	33.52	25.63
Sacramento	26.99	50.22	42.02	76.86	24.97	27.27
San Francisco	14.49	21.44	25.74	92.98	24.58	69.25
Tampa	39.33	32.47	28.56	92.65	16.36	25.48

3.5 SARIMA and GBRT 7-day and 28-day ahead for January, February, and March 2021

A relevant question is whether these findings are robust to the use of other time periods. Figure 3 visualizes the MAPEs associated with 7-day and 28-day ahead SARIMA and GBRT forecasts for the following testing periods: January 1–28 2021; February 1–28 2021; and March 1–28 2021. For the sake of brevity, we only report MAPEs. The proportion of daily forecasts falling within 20% of the actual values for these time periods are available upon request. As discussed, daily case counts dropped significantly for most MSAs during these months, making the evaluation of the predictive abilities an interesting contrast to our previous exercise of constructing forecasts during a period of rising daily COVID-19 cases during November-December 2020. We do not construct forecasts for Oklahoma City and Seattle, given the presence of a significant number of zeros in daily case values that seem anomalous to case counts in other days.

The results visualized in Figure 3 (and tabulated in Tables 5, 6) demonstrate that for almost all MSAs, with respect to 7-day ahead forecasts, GBRTs offer superior predictions relative to SARIMA models for the month of January. GBRT MAPEs are lower than 20% for 9 MSAs, with another 8 MSAs having MAPEs ranging from 20 to 25%. Therefore, for a number of MSAs, GBRTs were able to recognize the change in the trend in daily case counts, from increases to a steady decline. Like the November–December 2020 results, GBRTs are also able to produce more accurate 28-day predictions relative to SARIMA models. GBRT MAPEs are lower than 20% for 6 MSAs, with another 7 MSAs having MAPEs between 20% and 25%. In terms of specific MSAs, Albuquerque, Atlanta, Baltimore, Boston, Cleveland, Louisville, Miami, New York, San Francisco, and Tampa all have GBRT-generated 7-day ahead MAPEs lower than 20% for January 2021, with most of the MSAs (with the addition of Chicago) also possessing low GBRT MAPEs for 28-day ahead forecasts. Including Denver and Indianapolis, these are the same MSAs with low GBRT MAPEs during November - December 2020.

Some of the same trends are visible for February forecasts with GBRT 7-day ahead MAPEs lower than corresponding SARIMA values for 15 MSAs. Further, 9 MSAs have GBRT MAPEs lower than 20% and 4 MSAs possess GBRT MAPEs for from 20 to 25%. On the other hand, both GBRTs and SARIMA models produce low-quality forecasts with high 28-day ahead MAPEs for the same February time period; 17 or more cities possessing MAPE values >50%, and both methods result in MAPEs of over 100% for

TABLE 6 MAPE values—7-day- and 28-day forecasts from GBRT models January—March 2021.

City	January 1–28 2021		February 1–28 2021		March 1–28 2021	
	7-day	28-day	7-day	28-day	7-day	28-day
Albuquerque	19.34	23.89	35.19	83.11	14.94	28.43
Atlanta	18.84	24.18	15.52	52.28	22.12	32.84
Baltimore	23.07	21.2	16.49	91.01	15.54	28.31
Boston	12.4	15.76	15.1	73.26	15.01	14.14
Charlotte	14.41	18.21	25.56	48.66	31.05	47.94
Chicago	23.47	17.08	16.49	73.65	15.04	20.08
Cleveland	13.07	19.62	23.55	105.68	21.81	36.84
Dallas	20.8	21.2	31.01	63.77	29.89	80.1
Denver	36.26	44.47	24.97	26.39	60.43	56.56
Detroit	23.86	32.95	25.91	78.08	22.45	67.72
Houston	25.81	21.41	39.58	113.14	33.14	36.39
Indianapolis	21.59	27.82	23.56	96.7	19.48	17.45
Los Angeles	21.12	28.82	19.2	142.85	21.11	81.42
Louisville	17.9	22.39	22.82	108.54	24.27	77.49
Memphis	34.52	39.05	36.97	138.54	48.39	117.08
Miami	11.18	13.1	15.84	26.81	15.43	17.1
New York	14.14	14.26	14.96	32.69	8.03	11.24
Phoenix	24.08	35.68	37.75	148.52	36.63	137.68
Pittsburgh	26.5	48.61	14.07	30.69	20.29	18.92
Portland	29.73	53.59	30.08	197.17	33.57	46.15
Sacramento	23.97	31.12	30.93	30.71	25.67	38.8
San Francisco	14.27	23.95	12.39	112.16	27.09	110.62
Tampa	18.5	20.65	28.56	70.38	12.01	23.09

8 MSAs. These results suggest that both GBRT and SARIMA models were unable to accurately predict the downward trend in daily case rates that occurred in February for most MSAs. While we cannot confirm this, a possible reason for the decline in daily case rates might be the increased availability and uptake in vaccinations during and preceding February 2021. To the best of our knowledge, data on city-level vaccination rates are unavailable and, therefore, cannot be included as predictors to test this hypothesis.

In contrast to the February 2021 results, both GBRT and SARIMA March 2021 MAPEs become much lower for 7-day ahead forecasts and are comparable to previous months. GBRT and SARIMA performance are comparable in the sense that the number of MSAs are roughly split between the two methods in terms of which approach has the lowest MAPE. With both methods, 10 MSAs have 7-day ahead MAPEs lower than 20%, while another 8 MSAs possess MAPEs between 20 and 25%. On the other hand, SARIMA seems to be a superior method for generating 28-day ahead predictions in March 2021. Specifically, SARIMA forecasts yield MAPEs that are either comparable or much lower than corresponding MAPE values from GBRT models for most MSAs.

To summarize, the GBRT and SARIMA models are the most reliable in terms of producing 7-day ahead daily predictions with the lowest MAPE values for November-December 2020. With respect to 28-day ahead forecasts for the same time-period, while SARIMA and GBRT MAPES are comparable for several MSAs, there are MSAs where SARIMA MAPEs are much higher. Using Apple Mobility data does not considerably improve forecasts from SARIMA models. GBRT models similarly outperform SARIMA models for 7-day and 28-day ahead predictions for January, but neither model produces high quality predictions for February 2021, with MAPE values from GBRT models, on average, being slightly lower. For March 2021, 7-day ahead forecasts from GBRT and SARIMA models are comparable, with 28-day ahead SARIMA MAPE values being somewhat lower than corresponding GBRT estimates.

4 Discussion

Findings from previous studies suggest that the standard SIR model used by epidemiologists for disease predictions resulted in inaccurate forecasts for multiple jurisdictions. Further, we are

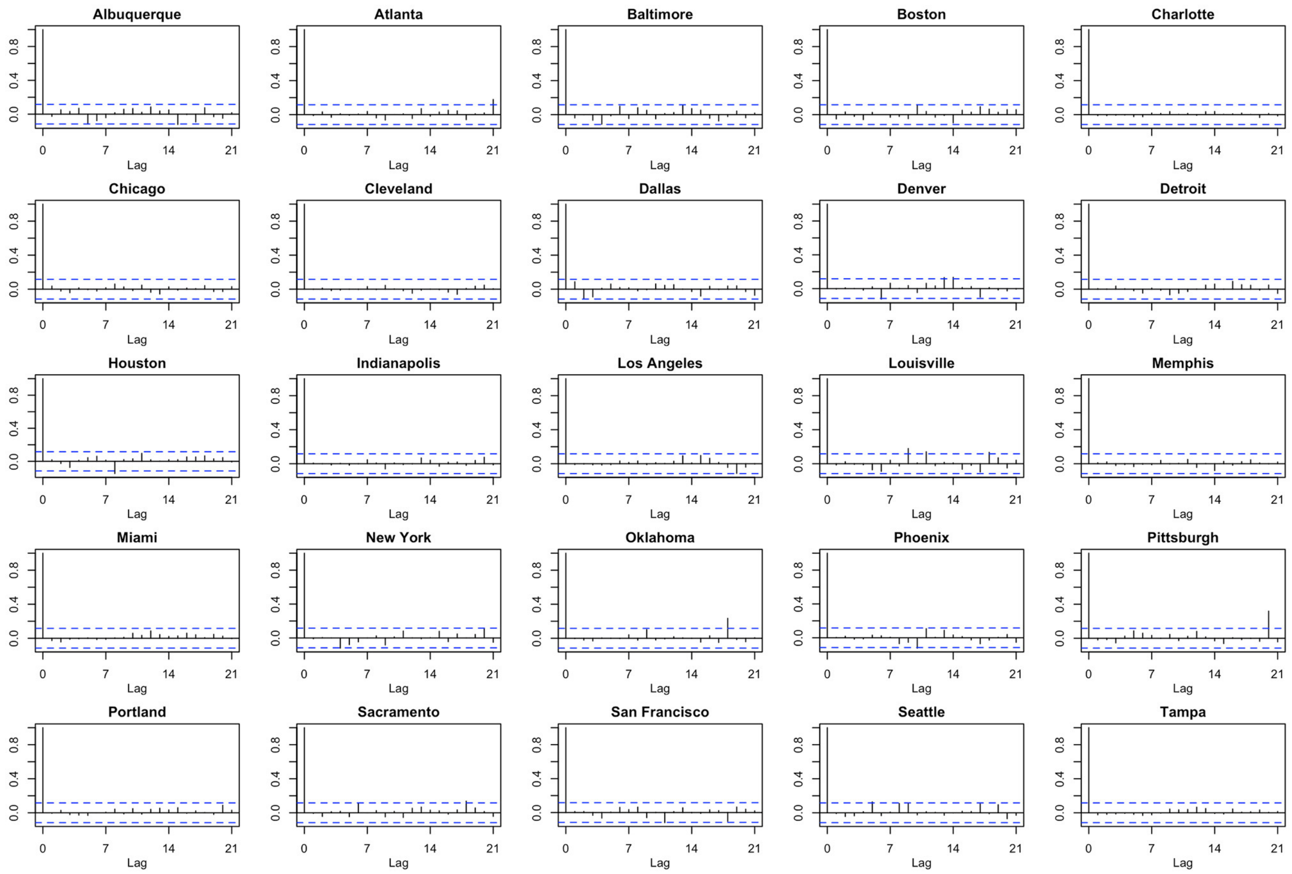


FIGURE 5
Autocorrelation function (ACF) plots of model residuals associated with the SARIMA model fit to March 2020-December 2020 data.

unaware of any other study that has attempted to construct daily forecasts of a panel of U.S. cities. While there is an abundance of research based on country-level data, research attempting to predict daily COVID-19 cases across cities is scarce (11). This is unfortunate, as cities or MSAs have often been the center of significant occurrences and spread of COVID-19 infections. Besides difficulties associated with obtaining city level data, forecasting daily cases is extremely challenging given the considerable heterogeneity in daily case trends across cities. Perhaps even more important are factors such as differences in testing rates, lack of uniformity in data collection protocols, and inaccurate processing of data, that make comparisons of cases across jurisdictions difficult (30).

This study explores the efficacy of SIR, LME, SARIMA, and GBRT models in producing one and four-week ahead predictions. The choice of these models is premised on different advantages. GBRTs are an example of supervised ML methods that have proved to be effective in different cases (12, 24) and can be developed with less knowledge than more sophisticated machine learning methods, such as deep learning. LME models accommodate city-level heterogeneity, enabling the researcher to borrow information across geographies, and hence do not rely exclusively on time-series variation within geographies. SIR models are the conventional workhorses in epidemiology, but do not have the flexibility to accommodate time-specific changes in external factors, such as population mobility, that can plausibly impact spread of infections. SARIMA models are relatively simple to implement and as noted earlier, are useful when the modeled data have pronounced trend and seasonality. These models have been employed by other studies (11, 12).

We evaluate model performance primarily through MAPEs and find that the prediction accuracy associated with LME and SIR forecasts tends to be inferior relative to that of SARIMA and GBRT forecasts. Compared to SARIMA, GBRTs generate 7-day ahead and 28-day forecasts with lower MAPEs for a vast majority of MSAs for most months. While SARIMA produces lower MAPEs for 28-day ahead predictions for March 2021, the good performance of GBRTs more broadly make it a suitable choice for modeling and forecasting daily COVID-19 case counts during time periods of both increase and decrease in infections. From a policy perspective, these results are important as they imply the availability of superior forecasting methods relative to conventional epidemiological methods. The performance of GBRTs should be noted, given that they are relatively straightforward to use through available statistical packages. The performance of these models is comparable to findings from other studies that use more disaggregated data than at the country level. For example, using provincial level data from Canada, Sen et al. (24) find MAPEs from GBRTs ranging from 8 to 30% for 2 week ahead daily forecasts. Zhang et al. (26) obtain Mean Absolute Percentage Errors ranging from 4 to 8% for 8 counties in California, but not for time-periods of 4 weeks or over.

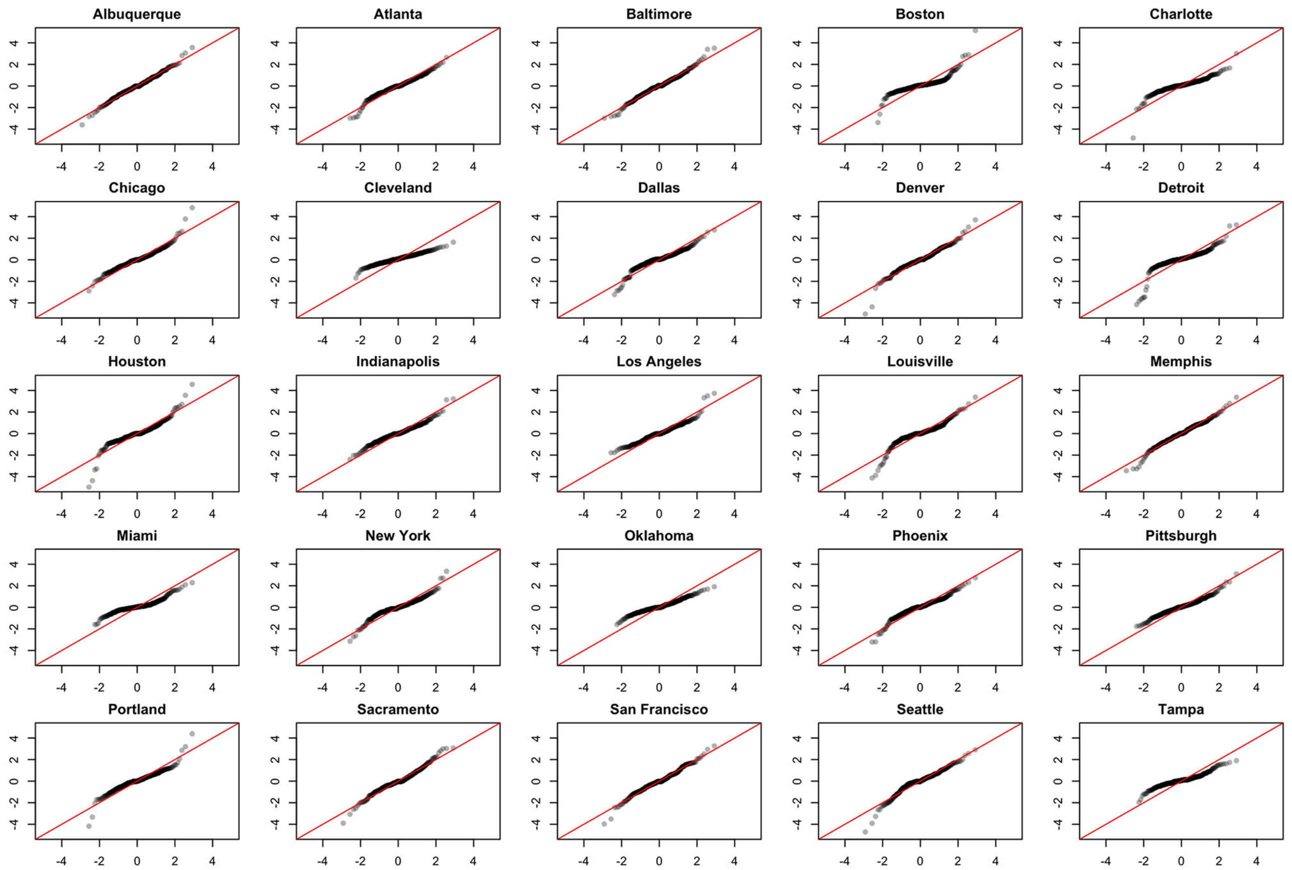


FIGURE 6
Normal QQ-plots of model residuals associated with the SARIMA model fit to March 2020-December 2020 data.

We note the following limitations. It is important to acknowledge that several of the chosen models make strong and potentially restrictive assumptions including, but not limited to, errors that are independently and identically distributed normal random variables. Such assumptions have been evaluated using a battery of residual diagnostic tools. Although many models and many specifications of these models were considered, we present results only for those whose assumptions were reasonably satisfied. To demonstrate this, we include Figures 4–6 which depict residuals, autocorrelation function and normal quantile-quantile plots of the residuals of the SARIMA model fit to the March 2020-December 2020. As is evident, we do not observe serious violations of stationarity or normality.

5 Conclusion

Along with other papers that have emerged over the past 2 years, the results of this study suggest that GBRTs can also be used for predicting the spread of highly infectious diseases on a daily basis. These findings suggest that the relatively basic ML modeling can lead to vital insights for government resource allocation and decision-making, and result in superior disease surveillance relative to conventional epidemiological methods. Future work will investigate the benefits of employing more complex deep learning and neural network-based methods, which have the trade-off of being more complex and possibly more accurate but also more difficult to interpret, especially for a wider audience of policy practitioners.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://github.com/kentranz/socialMobilityCOVID/blob/master/data/all.csv.

Author contributions

AS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing—original draft, Writing—review & editing. NS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing—original draft, Writing—review & editing. NT: Data curation, Formal analysis, Investigation, Methodology, Software, Writing—original draft. RA: Data curation, Formal analysis, Investigation, Methodology, Software, Writing—original draft. QZ: Data curation, Formal analysis, Investigation, Methodology, Software, Writing—original draft. JD: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this study was obtained to: NT gratefully acknowledges financial assistance from NSERC Discovery Grant RGPIN-2019-04212. JD acknowledges financial support from NSERC Discovery Grant RGPIN-2020-04382.

Acknowledgments

The authors acknowledge helpful comments from the two referees.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Chu B, Qureshi S. Predicting the covid-19 pandemic in Canada and the U.S. *Econ Bull.* (2020) 40:2565–85.

2. Chen LP, Zhang Q, Yi GY, He W. Model-based forecasting for Canadian covid-19 data. *PLoS ONE.* (2021) 16:e0244536. doi: 10.1371/journal.pone.0244536

3. Moein S, Nickaeen N, Roointan A, Borhani N, Heidary Z, Javanmard SH, et al. Inefficiency of sir models in forecasting covid-19 epidemic: a case study of isfahan. *Sci Rep.* (2021) 11:1–9. doi: 10.1038/s41598-021-84055-6

4. Ioannidis JP, Cripps S, Tanner MA. Forecasting for covid-19 has failed. *Int J Forecast.* (2022) 38:423–38. doi: 10.1016/j.ijforecast.2020.08.004

5. Chumachenko D, Meniailov I, Bazilevych K, Chumachenko T, Yakovlev S. Investigation of statistical machine learning models for covid-19 epidemic process simulation: random forest, k-nearest neighbors, gradient boosting. *Computation.* (2022) 10:86. doi: 10.3390/computation10060086

6. Krivtsov S, Meniailov I, Bazilevych K, Chumachenko D. Predictive model of covid-19 epidemic process based on neural network. *Radioelect Comp Syst.* (2022) 4. doi: 10.32620/reks.2022.4.01

7. Mohammadi A, Chumachenko D. COVID-19 epidemic process simulation using ARIMA model. In: *International Scientific and Technical Conference on Integrated Computer Technologies in Mechanical Engineering-Synergetic Engineering*, ICTM 2021 (2022). p. 353–63.

8. Dansana D, Kumar R, Das Adhikari J, Mohapatra M, Sharma R, Priyadarshini I, et al. Global forecasting confirmed and fatal cases of covid-19 outbreak using autoregressive integrated moving average model. *Front Public Health.* (2020) 8:580327. doi: 10.3389/fpubh.2020.580327

9. Singh S, Sundram BM, Rajendran K, Law KB, Aris T, Ibrahim H, et al. Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. *J Infect Dev Ctries.* (2020) 14:971–6. doi: 10.3855/jidc.13116

10. Sahai AK, Rath N, Sood V, Singh MP. ARIMA modelling and forecasting of covid-19 in top five affected countries. *Diabetes Metab Syndr Clin Res Rev.* (2020) 14:1419–27. doi: 10.1016/j.dsx.2020.07.042

11. Devaraj J, Madurai Elavarasan R, Pugazhendhi R, Shafiullah GM, Ganesan S, Jeysree AK, et al. Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant? *Results Phys.* (2021) 21:103817. doi: 10.1016/j.rinp.2021.103817

12. Fang ZG, Yang SQ, Lv CX, An SY, Wu W. Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study. *BMJ Open.* (2022) 12:e056685. doi: 10.1136/bmjopen-2021-056685

13. Liu Q, Fung DLX, Lac L, Hu P. A novel matrix profile-guided attention LSTM model for forecasting covid-19 cases in USA. *Front Public Health.* (2021) 9:741030. doi: 10.3389/fpubh.2021.741030

14. Nixon K, Jindal S, Parker F, Reich NG, Ghobadi K, Lee EC, et al. An evaluation of prospective covid-19 modelling studies in the USA: from data to science translation. *Lancet Digital Health.* (2022) 4:e738–47. doi: 10.1016/S2589-7500(22)00148-0

15. Bisin A, Moro A. JUE insight: learning epidemiology by doing: the empirical implications of a spatial-sir model with behavioral responses. *J Urban Econ.* (2022) 127:103368. doi: 10.1016/j.jue.2021.103368

16. Wang H, Yamamoto N. Using a partial differential equation with google mobility data to predict covid-19 in Arizona. *Math Biosci Eng.* (2020) 17:4891–904. doi: 10.3934/mbe.2020266

17. Garcia-Cremades S, Morales-Garcia J, Hernandez-Sanjaime R, Martinez-Espan∼a R, Bueno-Crespo A, Hernandez-Orallo E, et al. Improving prediction of covid-19 evolution by fusing epidemiological and mobility data. *Sci Rep.* (2021) 11:1–16. doi: 10.1038/s41598-021-94696-2

18. Guan G, Dery Y, Yechezkel M, Ben-Gal I, Yamin D, Brandeau ML. Early detection of covid-19 outbreaks using human mobility data. *PLoS ONE.* (2021) 16:1–17. doi: 10.1371/journal.pone.0253865

19. Ilin C, Annan-Phan S, Tai XH, Mehra S, Hsiang S, Blumenstock JE. Public mobility data enables covid-19 forecasting and management at local and global scales. *Sci Rep.* (2021) 11:1–11. doi: 10.1038/s41598-021-92892-8

20. Amusa LB, Twinomurinzi H, Okonkwo CW. Modeling covid-19 incidence with Google Trends. *Front Res Metrics Anal.* (2022) 7:1003972. doi: 10.3389/frma.2022.1003972

21. Altieri N, Barter RL, Duncan J, Dwivedi R, Kumbier K, Li X, et al. Curating a covid-19 data repository and forecasting county-level death counts in the United States. *Harvard Data Sci Rev.* (2020). doi: 10.1162/99608f92.1d4e0dae

22. Liu L, Moon HR, Schorfheide F. Panel forecasts of country-level covid-19 infections. *J Econ.* (2021) 220:2–22. doi: 10.1016/j.jeconom.2020.08.010

23. Stevens NT, Sen A, Kiwon F, Morita PP, Steiner SH, Zhang Q. Estimating the effects of non-pharmaceutical interventions (npis) and population mobility on daily covid-19 cases: evidence from Ontario. *Can Public Policy.* (2022) 48:144–61. doi: 10.3138/cpp.2021-022

24. Sen A, Baker JD, Zhang Q, Agarwal RR, Lam JP. Do more stringent policies reduce daily COVID-19 case counts? Evidence from Canadian provinces. *Econ Anal Pol.* (2023) 78:225–42. doi: 10.1016/j.eap.2023.03.006

25. Wathore R, Rawlekar S, Anjum S, Gupta A, Bherwani H, Labhasetwar N, et al. Improving performance of deep learning predictive models for covid-19 by incorporating environmental parameters. *Gondwana Res.* (2023) 114:69–77. doi: 10.1016/j.gr.2022.03.014

26. Zhang Y, Tang S, Yu G. An interpretable hybrid predictive model of covid-19 cases using autoregressive model and LSTM. *Sci Rep.* (2023) 13:6708. doi: 10.1038/s41598-023-33685-z

27. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals.* (2020) 135:109864. doi: 10.1016/j.chaos.2020.109864

28. Wang J, Tang K, Feng K, Lin X, Lv W, Chen K, et al. Impact of temperature and relative humidity on the transmission of covid-19: a modelling study in China and the United States. *BMJ Open.* (2021) 11:e043863. doi: 10.1136/bmjopen-2020-043863

29. Tolles J, Luong T. Modeling epidemics with compartmental models. *JAMA.* (2020) 323:2515–6. doi: 10.1001/jama.2020.8420

30. Alvarez E, Bielska IA, Hopkins S, Belal AA, Goldstein DM, Slick J, et al. Limitations of COVID-19 testing and case data for evidence-informed health policy and practice. *Health Res Pol Syst.* (2020) 21. doi: 10.1186/s12961-023-00963-1

frontiers | Frontiers in Public Health

TYPE Review
PUBLISHED 15 December 2023
DOI 10.3389/fpubh.2023.1338215

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Anjan Pal,
University of York, United Kingdom
Paulo Bandiera-Paiva,
Federal University of São Paulo, Brazil

*CORRESPONDENCE
Muhammad Hasnain
✉ drhasnain.it@leads.edu.pk

RECEIVED 14 November 2023
ACCEPTED 04 December 2023
PUBLISHED 15 December 2023

CITATION
Alotaibi SS, Rehman A and Hasnain M (2023) Revolutionizing ocular cancer management: a narrative review on exploring the potential role of ChatGPT.
Front. Public Health 11:1338215.
doi: 10.3389/fpubh.2023.1338215

Revolutionizing ocular cancer management: a narrative review on exploring the potential role of ChatGPT

Saud S. Alotaibi[1], Amna Rehman[2] and Muhammad Hasnain[2]*

[1]Information Systems Department, Umm Al-Qura University, Makkah, Saudi Arabia, [2]Department of Computer Science, Lahore Leads University, Lahore, Pakistan

This paper pioneers the exploration of ocular cancer, and its management with the help of Artificial Intelligence (AI) technology. Existing literature presents a significant increase in new eye cancer cases in 2023, experiencing a higher incidence rate. Extensive research was conducted using online databases such as PubMed, ACM Digital Library, ScienceDirect, and Springer. To conduct this review, Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines are used. Of the collected 62 studies, only 20 documents met the inclusion criteria. The review study identifies seven ocular cancer types. Important challenges associated with ocular cancer are highlighted, including limited awareness about eye cancer, restricted healthcare access, financial barriers, and insufficient infrastructure support. Financial barriers is one of the widely examined ocular cancer challenges in the literature. The potential role and limitations of ChatGPT are discussed, emphasizing its usefulness in providing general information to physicians, noting its inability to deliver up-to-date information. The paper concludes by presenting the potential future applications of ChatGPT to advance research on ocular cancer globally.

KEYWORDS

eye cancer, ChatGPT, challenges, oncology, artificial intelligence

1 Introduction

Ocular cancer is not only an issue for the population in developing countries, as developed countries have also reported cases of eye cancer. According to the American Society of Clinical Oncology, more than 3,400 new cases of eye cancer are expected to be diagnosed in 2023 (1). However, in developing countries such as Pakistan, this number is significantly higher. Based on statistics provided by Al-Shifa Trust Eye Hospital, out of the 150,000 individuals diagnosed with cancer each year, 2,200 of them have ocular cancer, resulting in the unfortunate loss of 220 lives annually (2023) (2). It is important to note that there are likely many unreported cases from remote or rural areas. Currently, there is scarce literature addressing ocular cancer, making us pioneers in exploring this issue. Recent work states that cancer is rising worldwide but cancer prevalence data is limited. Several factors contribute to the causes of ocular cancer (3). In order to accomplish this task, we are seeking support from emerging applications of artificial intelligence (AI) and relevant literature on the occurrence of ocular cancer in developing countries. Cancer screening programs have been launched that led to the improvement in survival of patients. However, patient selection and risk stratification are real challenges for caretakers. Moreover, work force and recently prevailing COVID-19 pandemic situation have increased the concerns (4).

The role of AI has been recognized in diagnosis and clinical management of certain cancer diseases (5). AI is excelled at performing well-organized tasks such as image recognition, brain tumor and skin cancer detection (6). AI is playing a clear role in giving support to doctors in their workplaces (7). Since last five years, AI has shown promising role in prospective and clinical trials (8). Open AI application, ChatGPT has shown better performance to examine the European Board of Ophthalmology test, education and knowledge assessment (9).

Emerging literature on ChatGPT and its applications in several fields have motivated us to present an overview of ocular cancer from the literature, shedding light on the challenges faced by common people worldwide. To the best of our knowledge, there currently exists a significant gap in the literature, as no comprehensive studies have yet focused on providing an overview of ocular cancer within the context of generative AI technology. Before this study, a previous review was conducted on the modern treatment of retinoblastoma and its treatment algorithms (10). However, the review focused on data up until 2020, and numerous research studies have been published since then. Another review on the topic of "ocular cancer" specifically addressed common ocular tumors, as well as their diagnosis and treatment technique (11). The most recent review article focuses on the role of deep learning models in cancer care. Deep learning significantly reduces costs in baseline imaging. Ethical issues related to the use of deep learning models were also studied in the review article (12). However, these review articles did not discuss the role of generative AI models and their applications in diagnosing ocular cancer. In the field of oncology and particular ocular cancer, the specific role and impact of ChatGPT application remains relatively unexplored. ChatGPT offers to address challenges related to ocular cancer by assisting in early detection using data analysis and offers tailored treatment recommendations.

To understand the versatility and widespread applicability of AI in healthcare, particularly in the context of ocular cancer, is an interesting topic to explore. The promising use of ChatGPT applications in ophthalmology for question-answering has been previously outlined, assessing the enhanced performance of large language models in healthcare (13). Additionally, a recent research highlights the application of ChatGPT in diagnosing intraocular tumors and ophthalmic pathologies (14). Moreover, the improved competency of generative AI applications in diagnosing diverse ocular conditions has further underscored its potential in the field of ophthalmology (15).

To conduct a narrative review on the role of ChatGPT application in ocular cancer is necessary and timely. Earlier mentioned studies highlight the rapid advancement of generative AI applications in healthcare, understanding the ChatGPT's capabilities in diagnosing and treating the ocular cancer becomes crucial. Insights drawn from existing literature can complement the existing practices and contribute to the management of ocular cancer diagnosis and treatment. Therefore, the main aim of this study is to bridge the gap in the literature by evaluating the specific contributions of AI applications, like ChatGPT, in diagnosing and treating ocular cancer. This includes identifying challenges faced by patients and examining how ChatGPT can help overcome these obstacles. Additionally, our review seeks to consolidate emerging information on the potential benefits of integrating ChatGPT in oncology. Furthermore, this narrative review outlines significant implications for future research in this rapidly evolving field.

Overall this narrative review contributes to the literature as follows:

- This paper presents an overview of the important types of ocular cancer based on literature and resources related to ChatGPT applications.
- This paper highlights the potential role of ChatGPT applications in improving the health outcomes of individuals with the ocular conditions.
- This paper identifies several challenges of ocular cancer and suggests ways to overcome them.
- This paper presents some important research implications for future studies in the area of AI applications and their use in the context of ocular cancer.

Section 2 outlines the methodologies employed in this paper. In Section 3, we provide the profound results and engage in in-depth discussions concerning the crucial role and applications of ChatGPT in addressing ocular cancer. Lastly, Section 4 succinctly summarizes the cornerstone findings of this research, thereby underscoring its significance.

2 Methods

Methodological approach adopted in this study is based on the narrative review that presents a qualitative interpretation of publications aimed at discussing the literature on the issues that may increase the debate from scientific community (16).

2.1 Search strategy and selection criteria

To conduct this review, relevant references were identified through the comprehensive searches in PubMed, Springer, ACM Digital Library and ScienceDirect databases. The search terms utilized for literature retrieval encompassed "Ocular cancer" OR "Eye cancer," in combination with "ChatGPT application" OR "Open AI application". Table 1 illustrates the strategic employment of search keywords across digital libraries.

Data extraction was performed by authors in three steps as follows:

1. All articles covering the topic were selected for the final inclusion in the review.
2. Authors read the articles and summarized the main features such as ocular cancer types, causes, treatment, future directions and references.
3. Authors read the articles and removed the overlapping and redundancy among titles.

Authors (SA, AR, and MH) contributed to the data extraction from the full-length articles. The search scope of this study was limited to the AI application and ocular cancer. Other articles were manually searched and added to the review to provide additional knowledge on the topic. Overall, 20 articles were selected for the inclusion in this review (Figure 1). For this purpose, we followed the PRISMA guidelines (17).

TABLE 1 Search keywords used on digital libraries.

Search keyword	Digital library	URL
(((Ocular Cancer) AND (ChatGPT)) OR (Eye Cancer)) AND (ChatGPT)	PubMed	https://pubmed.ncbi.nlm.nih.gov/37266720/
(ocular AND cancer AND ChatGPT AND application)	Springer	https://link.springer.com/advanced-search
(Ocular cancer or eye cancer And ChatGPT application)	ScienceDirect	https://www.sciencedirect.com/search?qs=Ocular%20cancer%20or%20eye%20cancer%20And%20ChatGPT%20application

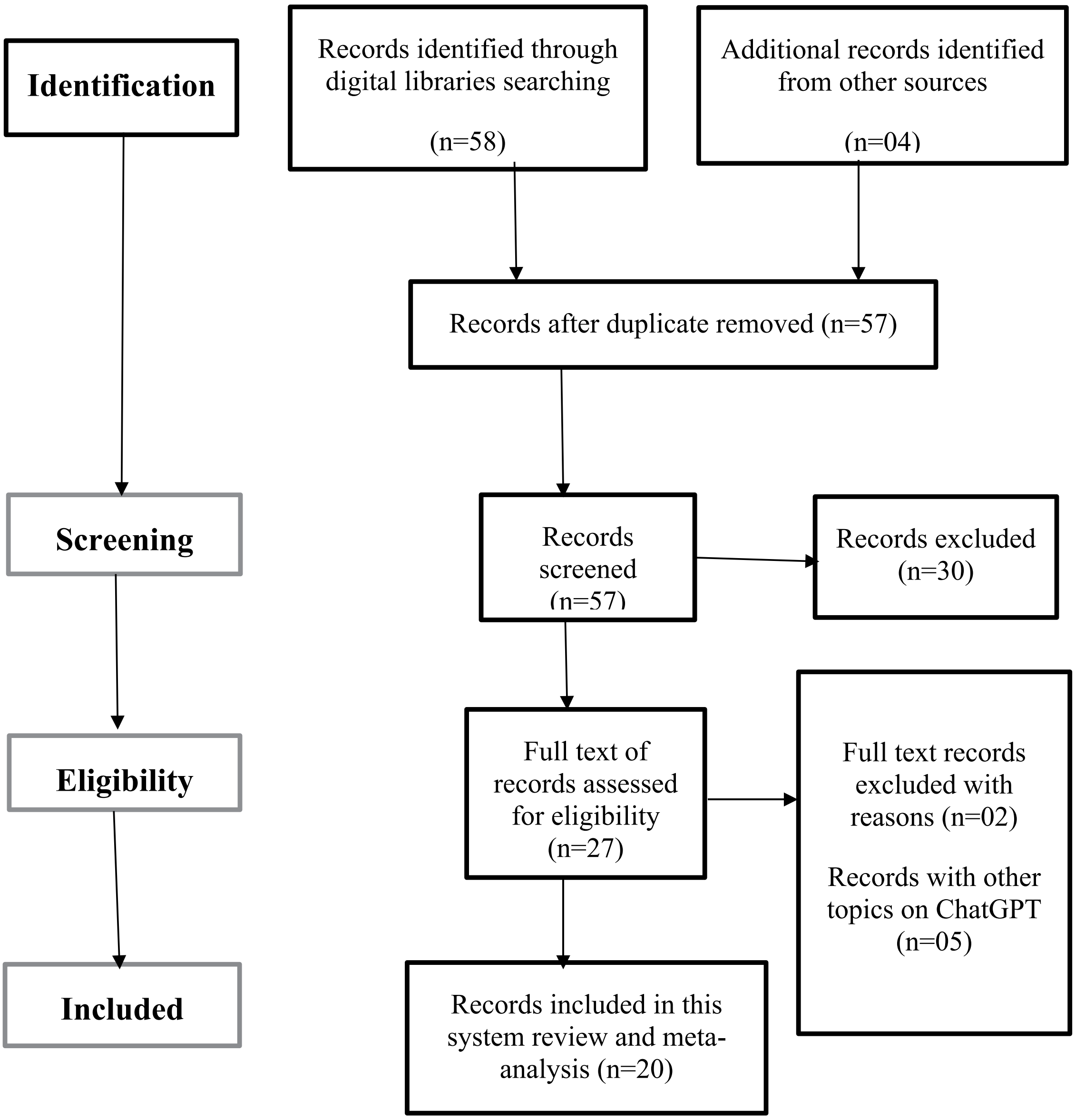


FIGURE 1
PRISMA flowchart.

2.2 Inclusion and exclusion criteria

Studies' inclusion and exclusion criteria are given as follows:

Full-text research articles published in English language were considered in this narrative review. Specifically, articles published within the last 4 years (from 2020 to 2023) were eligible for inclusion. To ensure the integrity of the review, duplicate articles were excluded. Additionally, grey literature, and book chapters were omitted from this review study. Moreover, research articles that did not relate to ocular cancer and ChatGPT were also excluded.

TABLE 2 Summary of ocular cancer types.

Ocular cancer	Causes	Treatment	Future directions	References
Retinoblastoma	RB1 gene mutations	Chemotherapy, radiation, and surgery	Targeted therapies, gene therapies, and immunotherapies	Dimaras et al. (19) and Ancona-Lezama et al. (10)
Ocular melanoma	Genetic mutations, UV exposure	Surgery, radiation, and immunotherapy	Personalized medicine, targeted therapies, and immunotherapies	van Poppelen et al. (20) and Ruan et al. (21)
Conjunctival melanoma	UV exposure, genetic mutations	Surgery, radiation, and immunotherapy	Targeted therapies, and immune checkpoint inhibitors	Brouwer et al. (22)
Ocular Surface Squamous Neoplasia (OSSN)	HPV infection, UV exposure	Surgery, topical chemotherapy, and cryotherapy	HPV vaccination, targeted therapies, and immunotherapies	Höllhumer et al. (23) and Kozma et al. (24)
Intraocular lymphoma	Unknown (possibly viral infections)	Chemotherapy, and radiation	Immunotherapies, targeted therapies, and understanding viral links	Ghesquieres et al. (25) and Hearne et al. (26)
Lacrimal gland tumors	Unknown (genetic and environmental)	Surgery, and radiation	Improved understanding of tumor biology, and targeted therapies	Emerick et al. (27)
Choroidal metastasis	Breast and lung cancers	Did not reveal	ChatGPT may help enhance human experience in healthcare	Au (28)

2.3 Document screening and data extraction

Titles and abstracts of studies were checked for de-duplication. Researchers of this study independently reviewed each document against the eligibility criteria. Following the initial screening, full-text records were collected to confirm the eligibility. No disagreement was observed between reviewers.

3 Results and discussion

Based on the search strategy in four databases, 62 records were collected (Figure 1). After removal of duplicate records, 57 articles were further screened for eligibility. Of the remaining records, 30 were excluded using the exclusion criteria. Resulting 27 records were evaluated based on the full-length articles, after that 7 records were excluded. A total of 20 articles were selected in this narrative review. Subsequently results have been discussed in the following.

3.1 Important types of ocular cancer

To investigate the strength of ChatGPT in the identification of ocular cancer types, we present the comparison of literature and ChatGPT-generated text on ocular cancer types. Prominent venues published articles on various types of ocular cancer. They reported Retinoblastoma and Ocular Surface Squamous Neoplasia (OSSN) in the study and found that the former type was more common than the latter one (18). In the following table, we outline ocular cancer types, their respective causes, treatments, future directions and relevant references.

The data presented in Table 2 varies due to differences in treatment oprions, location, stage and individual patient characterization. Seven types of ocular cancer have been identified from the ChatGPT responses. UV exposure is mainly pointed out as the main cause of different ocular cancer types. The table outlines future directions, offering potential areas where ongoing research projects could be efficiently implemented. Aside from the causes of ocular cancer, literature reports suggest that genetic mutation, ultraviolet radiation, age, environmental exposures, viral infections and hereditary conditions may also contribute to eye diseases (29, 30). Notably, Ocular Adnexal Malignancies has been missed by the ChatGPT application (31). Regarding treatments as listed in Table 2, the shift towards personalized medicine, especially in conjunctival melanoma and ocular melanoma, reflects a growing trend in oncology, moving away from generalized treatments towards more individualized techniques. However, the treatment for Choroidal Metastasis has not been widely explored in the literature. This gap is exemplified by one of the studies by Au (28), as listed in Table 2, which is marked with a 'Did not reveal' status. This indicates a significant omission of a treatment strategy in the study. Furthermore, the potential role of ChatGPT in enhancing healthcare experiences, particularly for choroidal metastasis, signals an increasing recognition of AI's role in healthcare. However, this appears as an isolated mention, indicating an early stage of AI integration in this field.

A detailed analysis of the data in the Table 2 reveals that while ChatGPT is a valuable resource, it is not without limitations as a source of information for health data, particularly in the context of ocular cancer types. Several important aspects of ocular cancer have been overlooked by the ChatGPT application. This discrepancy may stem from the data used to train ChatGPT, which remains current only until 2021. The recent advancements in ocular cancer research have not been incorporated into the training of the OpenAI application. Consequently, we are unable to access updated and highly accurate information regarding the causes and treatment of ocular cancer.

3.2 Ocular cancer challenges

This section presents some challenges associated with the ocular cancer.

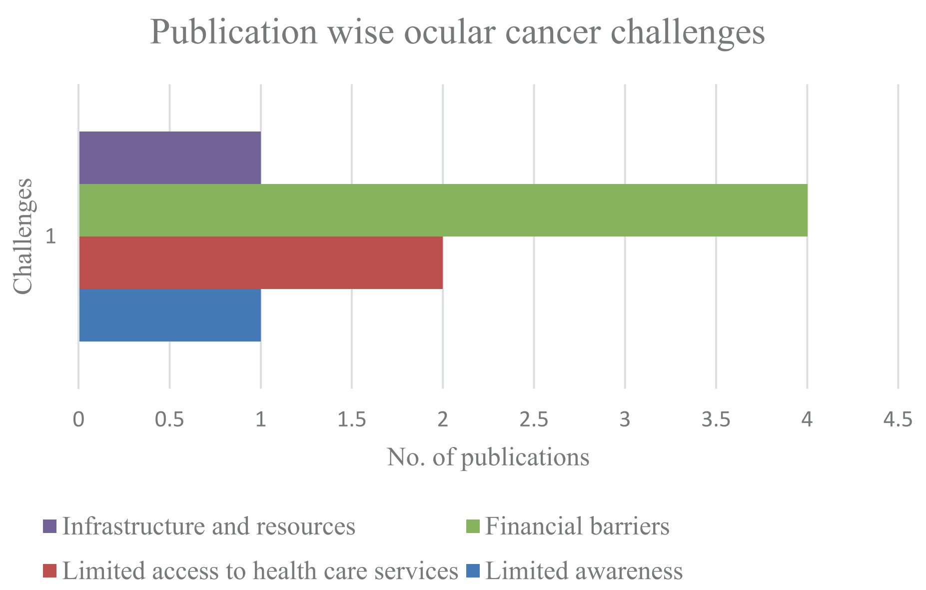


FIGURE 2
Publication wise ocular cancer challenges.

3.2.1 Limited awareness

In many developing countries, a critical lack of awareness about the signs and symptoms of eye cancer persists, often leading to delayed diagnosis and treatment. This issue is particularly true in rural areas where access to healthcare and health education is frequently limited. Targeted public health campaigns and educational initiatives are essential to increase awareness among the general population, healthcare professionals, and policymakers. Research has shown that individuals in resource-limited economies often have limited awareness and knowledge about the ocular cancer (32). Therefore, education is a key factor in elevating public awareness about eye cancer, especially in these vulnerable populations.

3.2.2 Limited access to health care services

The availability of specialized ophthalmic oncology services, including expert ophthalmologists and oncologists, can be markedly limited in certain regions of the world. This scarcity significantly hinders the ability of people with eye cancer to receive timely and appropriate care. A study conducted in the Middle East, West Asia, and North African regions investigated the treatment capabilities concerning retinoblastoma. This study further analyzed resources such as diagnostics, advanced treatment, focal therapy and chemotherapy (33). A significant disparity in the availability of resources among these countries was evident. Consistent with previous studies, recent research also showed that Retinoblastoma care facilities were notably limited in Ethopia, predominantly restricted to urban areas (34).

3.2.3 Financial barriers

The expense of diagnostic procedures, surgical procedures, radiation therapy, and post-operative care may be prohibitive for patients due to the high cost of eye cancer treatment. Cost is one of the leading factors that increases difficulties in receiving better treatment modalities (35). The literature overlooks this issue associated with the high cost of ocular cancer treatment. One study solely presents the use of convolutional neural networks for detection of ocular abnormalities (36). Similarly, another study examines the ocular melanoma in Polish population and does not present cost associated with this type of ocular cancer (37).

Government initiatives, insurance coverage, and support programmers are just a few examples of accessible and cheap healthcare choices that can assist ease the financial strain on impacted people and their families. There is a high cost on travelling and referrals for the treatment of eye cancer (38). Delays in care may cause the serious concerns and challenges in receiving the definitive treatment.

3.2.4 Infrastructure and resources

Effective management of eye cancer depends on the accessibility of cutting-edge medical infrastructure, including diagnostic facilities, treatment tools, and skilled healthcare personnel. Diagnosis, therapy, and overall patient outcomes can all be improved by enhancing the infrastructure and resources devoted to eye cancer care.

Several other challenges contribute to the rise of eye cancer in different settings. United Kingdome healthcare system was impacted by the COVID-19 pandemic. There was a considerable decrease in melanoma referral patterns in UK during the first wave of COVID-19 in 2020 (39). Since then recovery plan has been initiated by the ocular oncology services.

As seen in Figure 2, financial barriers have been widely studied as ocular cancer challenges in the literature. The second most examined ocular cancer challenge is limited access to healthcare services, followed by infrastructure and resource constraints, and limited awareness, each documented in a single publication.

3.3 Potential role of ChatGPT to overcome ocular cancer challenges

ChatGPT is designed to converse with users in natural language, responding to their inquiries and assertions in a manner that simulates

TABLE 3 Potential role of ChatGPT application.

Problem	Methods	Advantages	Disadvantages	Recommendation	References
Operative notes documenting	ChatGPT application	Supports to write operative notes	Sufficient in fully comprehending the requirements	ChatGPT can be employed with human expertise.	Waisberg et al. (42)
Inherent complexity of operative notes	ChatGPT application	Plays promising role in healthcare	Certain privacy concerns are overlooked	Crucial consideration to overcoming the data privacy of patients	Lawson McLean (43)
Answering question related with ophthalmology	ChatGPT application	Provides 50% right answers to questions	Accuracy is not up to the required standard	Updated information can be used to train ChatGPT and receive more accurate results	Mihalache et al. (44)
Accuracy in answering the ocular symptoms	ChatGPT application	Shows better accuracy ChatGPT-3.5 (59.5%) and ChatGPT-4 (89.2%)	ChatGPT exhibits weaknesses in its medical acumen compared to human experts.	Continuous improvement in ChatGPT could minimize the inaccuracy in its performance	Pushpanathan et al. (45)

online discussion (40). Capable for addressing various topics, from general knowledge questions to more specialized inquiries spanning healthcare, travel, technology, and many other topics, the Chatbot can comprehend and react to a wide range of topics.

Here are some potential ways that GPT-based chatbots could be used to overcome ocular cancer challenges:

- **Education and awareness:** patients can get information from chatbots about ocular cancer symptoms, risk factors, and treatment options. Cancer patients continuously require online sources to educate themselves. ChatGPT can provide them information about ocular cancer. This may aid in raising awareness and promoting early detection of ocular cancer, both of which are essential for effective treatment outcomes. Recent work evaluate the role of ChatGPT in providing recommendations about breast, lungs, and prostate cancers (41). However, this study overlooks the role of ChatGPT in giving information to patients on ocular cancer.
- **Support for patients and caregivers:** ChatGPT might offer help and direction to patients and caregivers by responding to their frequent inquiries and sharing details about available resources and support services. ChatGPT helps patients cope with the emotional impact of cancer, while patient caregivers assist them in managing the distress caused by the disease. ChatGPT reassures patients and caregivers that they are not alone (28). It offers valuable social support to patients in their personal lives.
- **Telemedicine and remote consultations:** ChatGPT might make it easier for patients and medical professionals to interact and communicate online during telemedicine and remote consultations. Patients who reside in distant or underdeveloped locations, where access to specialized healthcare services may be constrained, may find this to be of great use.
- **Research and development:** ChatGPT could support ocular cancer research and development by enabling access to information and insights that could guide the creation of novel therapies and treatments.

In developing countries, dealing with the effects of ocular cancer requires a comprehensive approach. This entails disseminating knowledge about the condition, boosting patient and family support networks, expanding access to specialized care, and promoting early identification through routine eye exams. Additionally, putting regulations in place to reduce the cost of ocular cancer treatments will aid in more effective management and treatment for affected individuals. Beyond these ChatGPT applications, existing literature also emphasizes the potential applications of ChatGPT in ocular cancer. Table 3 presents a summary of the literature on ocular cancer along with its ChatGPT applications.

The current literature on cancer and related topics regarding ChatGPT applications is rapidly emerging. Numerous published works have highlighted the role of ChatGPT applications in healthcare, with a particular focus on cancer (46, 47). However, only a limited number of articles have been published on the potential role of ChatGPT applications in enhancing the health conditions of ocular patients. These studies primarily concentrated on generating operative notes during surgical procedures (42, 43). This demonstrates a clear trend towards leveraging the ChatGPT application to enhance efficiency, as highlighted by its use in answering specialized questions and documenting operative notes. While ChatGPT is emerging as a useful tool in oncology, its effectiveness is tempered by concerns over accuracy and comprehension. For instance, ChatGPT achieves only 50% accuracy in some cases (44), and shows limitations in understanding complex medical treatments. In addition, Data privacy has not been adequately addressed in the process of generating these operative notes (44, 45). The data utilized in surgical operations is highly sensitive and must be shared with the surgical staff in a cautious and meticulous manner.

ChatGPT application demonstrates its widespread use in many areas. GPT's role in identifying the rare eye disease has been explored in a research study (48). ChatGPT works as a consultation assistant for patients and family physicians for referral suggestions. In addition, ChatGPT application helps junior ophthalmologist in diagnosing the rare eye diseases. All ChatGPT users need to be careful and cautious to acknowledge the adequate referrals and verifications in clinical settings. ChatGPT application is being applied in the world of intelligent diagnostics, Literature reveals that ChatGPT has shown potential in radiology and interpretation of the clinical images (49). It has remarkably improved the clinical workflow and responsible utilization of radiology services. Higher accuracy in clinical decision making has been seen using the ChatGPT application (50). Trusting the ChatGPT application is crucial for its adoption in healthcare, considering that ChatGPT wasn't primarily developed for healthcare purposes. Overreliance on this technology might potentially result in disseminating false information and health risks (51). Therefore, our efforts should focus on further enhancing the ChatGPT application to differentiate between queries that it can manage effectively and those that should be redirected to human experts.

Overall, these insights indicate a future in which AI could play a supporting and integral role in ocular cancer, necessitating continuous advancements.

3.4 Research implications

Currently, AI applications have demonstrated a significant role in ophthalmology, particularly in the areas of imaging and data measurement. Conventional methods to diagnose ophthalmic disease depend on the clinical assessment and image capturing devices for the different modalities (52). These diagnostic methods are time consuming and costly, and make ophthalmology one of the areas suited to the recent deep learning models. ChatGPT application is widely used in training curricula in medicine. Although ChatGPT application is still in the research stage, it has shown promising results, particularly in ocular cancer, where the potential role of AI is extensively explored in the literature (53). In this context, ChatGPT serves as a valuable tool to assist researchers and practitioners in successfully transitioning to AI-based approaches in the field of ocular cancer. By leveraging the capabilities of ChatGPT, professionals can benefit from its insights and assistance, ultimately leading to improved outcomes for patients. By sharing insights and potential data on the AI platform, physicians may promote greater communication among other doctors and patients (54). This way, physicians can make better decisions by accessing the patients' data and opinions from colleague doctors.

Recent work explores the use of ChatGPT in the simulated "Ophthalmic Knowledge Assessment Program" OKAP examination. ChatGPT achieved an accuracy of 59.4% on the OpthoQuestions testing set compared to a 74% accuracy by human on the Basic and Clinical Science Course (BCSC) test set (13). Another research study reported that ChatGPT attained 46% accuracy in answering OpthoQuestions for the preparation of board examinations (44). Compared to human performance on the testing set, ChatGPT exhibited slightly lower performance (55). However, ChatGPT's performance is noteworthy and could be further enhanced with the use of updated versions of ChatGPT. The latest versions of ChatGPT can be trained with the most recent oncology data, ensuring higher accuracy and relevance in healthcare. The recent version of ChatGPT (v. 3.0) demonstrated limitations in controlling a question's difficulty as well as its cognitive level. This limitation should be considered for future versions of ChatGPT applications.

In a most recent study, it was revealed that ChatGPT's performance in offering recommendations for cancer treatment deviated from NCC guidelines (56). This is a major concern that needs to be addressed in the near future. Although, ChatGPT largely adhered to the NCC guidelines, instances of partial deviation should be addressed in upcoming versions of the ChatGPT application.

ChatGPT could be future of the AI technology due to its features. However, we need to evaluate and monitor it while using for the communication of the eye cancer. Therefore, potential bias must be eliminated to provide equal information to all populations. ChatGPT shows limitations to answer the queries on the updated information (57). ChatGPT is trained on the data collected prior to 2021, and emerging information on eye cancer is not accurate and its accuracy may be increased due to its training on the data collected to date. While ChatGPT version 4.0 has addressed this deficiency to some extent, there is a continued need for consistent enhancements in future versions. A recent study reports that none of the existing studies explored the proposal of AI model to diagnose the Retinoblastoma (58). Common types of Retinoblastoma diagnosis were based on the clinical examination and special signs (59). One of the leading works used AI and Machine learning ML models to analyze the images and numerical data (60). At this stage, the interpretation of results is a very sensitive task in cancer research. Therefore, ChatGPT could be incorporated to make interpretation easier in studies involving the data regarding ocular cancer. Additionally, AI models may be used in the future to detect and refer ocular oncologies, specifically focusing on the lesions where treatment is required.

The emphasis on telemedicine applications extends to ocular cancer. Recent work highlights the considerations for implementing teleophthalmology in Brazil as a future research direction (61). Incorporating ChatGPT application with other technologies may give rise to favourable approaches for physicians and patients in oncology.

A large volume of unstructured information in the form of electronic health records (EHRs) is rather difficult to analyze using traditional techniques (62). ChatGPT has the potential to revolutionize oncology by analyzing vast amounts of information, uncovering trends and patterns that might elude human detection. This innovative tool could enable physicians in oncology to make more precise decisions regarding diagnosis, treatment, and cancer prevention, leading to highly personalized treatment plans tailored to individual patients' needs. Moreover, the strategic integration of ChatGPT with EHRs could significantly reduce the workload of frontline healthcare workers in oncology, providing crucial support and alleviating workforce shortages, particularly during times of crisis (63). Due to increased concerns about data security and privacy among AI ChatGPT users, it could be integrated with blockchain technology to ensure data consistency and protect medical data, thereby providing traceable and secure usage in oncology medicine (64). Furthermore, for the advancement of medical research in oncology, ChatGPT could become a faster and more reliable source for searching medical information and tracking the progress in the research field.

Teleconsultants benefit from ChatGPT application as they can receive timely information regarding medical conditions, symptoms, treatment and medication (49). This may help teleconsultants understand patients' concerns, needs and expectations. In telemedicine, routine tasks include scheduling appointments, sending reminders, and refilling prescriptions. ChatGPT helps explore the virtual patient-physician interaction. A virtual assistant based on ChatGPT may be developed to assistant healthcare professionals in telemedicine to triage patients, and provide remote guidance for home care (65). However, we have observed that participants in such studies may encounter several issues while using the AI application. For example, Caruccio et al. (49) reported that during traditional consultations, patients communicate with physicians who highly prioritize the privacy and confidentiality of their interactions. However, a physician using the ChatGPT application may inadvertently create ambiguity for patients, who have a legitimate right to know whether they are interacting with an AI application or a teleconsultant. Moreover, patients' privacy and security could potentially be compromised when physicians input sensitive information such as conditions, age, sex, and other factors into the AI system for suggestions. Consequently, significant ethical concerns regarding data privacy, security, accountability, and liability may impose limitations on the utilization of ChatGPT in telemedicine. During the COVID-19 pandemic, ocular oncology faced new challenges with a focus on preventing SARS-CoV-2 exposure. Telemedicine stands out as a major option, setting new standards for ocular treatment during unprecedented times (66). However, a major challenge is to implement the safety measures for healthcare workers and patients to enhance ocular cancer diagnosis and treatment practices.

We noted "unknown" causes for several ocular cancers (25, 27), highlighting areas for future research. Additionally, the 'Did not reveal' status under the treatment for choroidal metastasis (28), suggests a lack of consensus or insufficient information regarding treatments, identifying a critical area for further research and development.

4 Conclusion

This paper presents an overview of literature detailing the challenges associated with the ocular cancer and explores ways to overcome them using information obtained from the ChatGPT application. This narrative review identified several significant types of ocular cancer, exploring their causes, treatments, and future directions. The strengths and limitations of the ChatGPT application were thoroughly examined in this paper. Furthermore, this review highlighted the challenges associated with ocular cancer. Public health education, telemedicine and remote support through ChatGPT are recommended approaches. Additionally, further research and development in ChatGPT can aid in the prevention of complex eye diseases, such as malignant tumors inside the human eye. Early detection and standardized ocular cancer treatment significantly reduce the risk of developing this disease.

ChatGPT has promising role in managing ocular cancer, offering advantages such as early detection, patient education, treatment plans and research support. However, the successful implementation of ChatGPT in ocular cancer care requires careful addressing of challenges such as data availability, cultural and linguistic sensitivity, regulation and infrastructure support. Addressing these challenges is necessary in future endeavors.

Author contributions

SA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Writing – original draft. AR: Formal analysis, Project administration, Resources, Software, Visualization, Writing – original draft. MH: Methodology, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors extend their appreciation to the Deanship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number: IFP22UQU4210118DSR037.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Durden F. L. (2023). Melanoma in the eye. Available at: https://www.cancercenter.com/cancer-types/melanoma/types/intraocular-melanoma (Accessed).

2. Al-Shifa Trust eye hospital [online]. (2023). Available at: https://alshifaeye.org/al-shifa-center-for-eye-cancer.php.

3. Saeed S, Khan JA, Iqbal N, Irfan S, Shafique A, Awan S. Cancer and how the patients see it; prevalence and perception of risk factors: a cross-sectional survey from a tertiary care Centre of Karachi, Pakistan. *BMC Public Health*. (2019) 19:1–7. doi: 10.1186/s12889-019-6667-7

4. Hunter B, Hindocha S, Lee RW. The role of artificial intelligence in early cancer diagnosis. *Cancers*. (2022) 14:1524. doi: 10.3390/cancers14061524

5. Soons E, Siersema P, Van Lierop L, Bisseling T, Van Kouwen M, Nagtegaal I, et al. Laboratory variation in the grading of dysplasia of duodenal adenomas in familial adenomatous polyposis patients. *Familial Cancer*. (2023) 22:177–86. doi: 10.1007/s10689-022-00320-1

6. Lan Y, Xie L, Cai X, Wang L. A many-objective evolutionary algorithm based on integrated strategy for skin cancer detection. *Ksii Transac Internet Inform Syst*. (2022) 16:80–96. doi: 10.3837/tiis.2022.01.005

7. Kaplan A, Cao H, Fitzgerald JM, Iannotti N, Yang E, Kocks JW, et al. Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and Copd diagnosis. *J Allergy Clin Immunol Pract*. (2021) 9:2255–61. doi: 10.1016/j.jaip.2021.02.014

8. Kang J, Chowdhry AK, Pugh SL, Park JH. Integrating artificial intelligence and machine learning into Cancer clinical trials. *Semin Radiat Oncol*. (2023) 33:386–94. doi: 10.1016/j.semradonc.2023.06.004

9. Panthier C, Gatinel D. Success of Chatgpt, an Ai language model, in taking the French language version of the European Board of Ophthalmology examination: a novel approach to medical knowledge assessment. *J Fr Ophtalmol*. (2023) 46:706–11. doi: 10.1016/j.jfo.2023.05.006

10. Ancona-Lezama D, Dalvin LA, Shields CL. Modern treatment of retinoblastoma: a 2020 review. *Indian J Ophthalmol*. (2020) 68:2356. doi: 10.4103/ijo.IJO_721_20

11. Maheshwari A, Finger PT. Cancers of the eye. *Cancer Metastasis Rev*. (2018) 37:677–90. doi: 10.1007/s10555-018-9762-9

12. Chakrabarty N, Mahajan A. Imaging analytics using artificial intelligence in oncology: a comprehensive review. *Clin Oncol*. (2023). doi: 10.1016/j.clon.2023.09.013 [in press].

13. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of chatgpt in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmol Sci*. (2023) 3:100324. doi: 10.1016/j.xops.2023.100324

14. Sensoy E, Citirik M. A comparative study on the knowledge levels of artificial intelligence programs in diagnosing ophthalmic pathologies and intraocular tumors evaluated their superiority and potential utility. *Int Ophthalmol*. (2023):1–5. doi: 10.1007/s10792-023-02893-x

15. Jiao C, Edupuganti NR, Patel PA, Bui T, Sheth V, Edupuganti N. Evaluating the artificial intelligence performance growth in ophthalmic knowledge. *Cureus*. (2023) 15:e45700. doi: 10.7759/cureus.45700

16. Mesce M, Ragona A, Cimino S, Cerniglia L. The impact of media on children during the Covid-19 pandemic: a narrative review. *Heliyon*. (2022) 8:e12489. doi: 10.1016/j.heliyon.2022.e12489

17. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the Prisma statement. *Int J Surg*. (2010) 8:336–41. doi: 10.1016/j.ijsu.2010.02.007

18. Maurya RP, Prajapat MK, Singh VP, Roy M, Todi R, Bosak S, et al. Serum malondialdehyde as a biomarker of oxidative stress in patients with primary ocular carcinoma: impact on response to chemotherapy. *Clin Ophthalmol*. (2021) 15:871–9. doi: 10.2147/OPTH.S287747

19. Dimaras H, Corson TW, Cobrinik D, White A, Zhao J, Munier FL, et al. Retinoblastoma. *Nat Rev Dis Primers*. (2015) 1:1–23. doi: 10.1038/nrdp.2015.21

20. Van Poppelen NM, De Bruyn DP, Bicer T, Verdijk R, Naus N, Mensink H, et al. Genetics of ocular melanoma: insights into genetics, inheritance and testing. *Int J Mol Sci*. (2020) 22:336. doi: 10.3390/ijms22010336

21. Ruan J, Li F, Tian H, Yu J, Deng H, Ge S, et al. A cascade fret photosensitizer that enhances photodynamic therapy for ocular melanoma. *Nano Today*. (2022) 47:101684. doi: 10.1016/j.nantod.2022.101684

22. Brouwer NJ, Verdijk RM, Heegaard S, Marinkovic M, Esmaeli B, Jager MJ. Conjunctival melanoma: new insights in tumour genetics and immunology, leading to new therapeutic options. *Prog Retin Eye Res*. (2022) 86:100971. doi: 10.1016/j.preteyeres.2021.100971

23. HÖllhumer R, Williams S, Michelow P. Ocular surface squamous neoplasia: management and outcomes. *Eye*. (2021) 35:1562–73. doi: 10.1038/s41433-021-01422-3

24. Kozma K, Dömötör ZR, Csutak A, Szabó L, Hegyi P, Erőss B, et al. Topical pharmacotherapy for ocular surface squamous neoplasia: systematic review and meta-analysis. *Sci Rep*. (2022) 12:14221. doi: 10.1038/s41598-022-18545-6

25. Ghesquieres H, Chevrier M, Laadhari M, Chinot O, Choquet S, Molucon-Chabrot C, et al. Lenalidomide in combination with intravenous rituximab (Revri) in relapsed/refractory primary Cns lymphoma or primary intraocular lymphoma: a multicenter prospective 'proof of concept'phase ii study of the French Oculo-cerebral lymphoma (Loc) network and the lymphoma study association (Lysa). *Ann Oncol*. (2019) 30:621–8. doi: 10.1093/annonc/mdz032

26. Hearne E, Netzer OT, Lightman S. Learning points in intraocular lymphoma. *Eye*. (2021) 35:1815–7. doi: 10.1038/s41433-021-01500-6

27. Emerick C, Mariano FV, Vargas PA, Nör JE, Squarize CH, Castilho RM. Adenoid cystic carcinoma from the salivary and lacrimal glands and the breast: different clinical outcomes to the same tumor. *Crit Rev Oncol Hematol*. (2022) 179:103792. doi: 10.1016/j.critrevonc.2022.103792

28. Au SCL. Patient with cancer who found support and care from Chatgpt. *Cancer Res Stat Treatment*. (2023) 6:305–7. doi: 10.4103/crst.crst_113_23

29. Lucas R, Yazar S, Young A, Norval M, De Gruijl F, Takizawa Y, et al. Human health in relation to exposure to solar ultraviolet radiation under changing stratospheric ozone and climate. *Photochem Photobiol Sci*. (2019) 18:641–80. doi: 10.1039/c8pp90060d

30. Parsa N. Environmental factors inducing human cancers. *Iran J Public Health*. (2012) 41:1–9.

31. Nittmann M, Margo CE. Age conditional probability of ocular and ocular adnexal malignancies. *Ocul Oncol Pathol*. (2021) 7:70–3. doi: 10.1159/000511364

32. Ayanniyi AA, Jamda AM, Badmos KB, Adelaiye RS, Mahmoud AO, Kyari F, et al. Awareness and knowledge of ocular cancers in a resource-limited economy. *N Am J Med Sci*. (2010) 2:526–31. doi: 10.4297/najms.2010.2526

33. Burges M, Qaddoumi I, Brennan RC, Krull L, Sahr N, Rodriguez-Galindo C, et al. Assessment of retinoblastoma capacity in the Middle East, North Africa, and West Asia region. *Jco Global Oncol*. (2020) 6:1531–9. doi: 10.1200/GO.20.00321

34. Sherief ST, Wu F, O'banion J, Teshome T, Dimaras H. Health services needs assessment for retinoblastoma in Ethiopia. *JCO Global Oncol*. (2023) 9:e2200445. doi: 10.1200/GO.22.00445

35. Zeng L, Gowda BJ, Ahmed MG, Abourehab MA, Chen Z-S, Zhang C, et al. Advancements in nanoparticle-based treatment approaches for skin cancer therapy. *Mol Cancer*. (2023) 22:10. doi: 10.1186/s12943-022-01708-4

36. Santos-Bustos DF, Nguyen BM, Espitia HE. Towards automated eye cancer classification via Vgg and ResNet networks using transfer learning. *Eng Sci Technol Int J*. (2022) 35:101214. doi: 10.1016/j.jestch.2022.101214

37. Nowak MS, Romanowska-Dixon B, Grabska-Liberek I, Żurek M. Incidence and survival of ocular melanoma in National Cancer Registry of Poland in 2010–2017. *Adv Clin Exp Med*. (2022) 31:615–21. doi: 10.17219/acem/146581

38. Sherief ST, Wu F, O'banion J, Teshome T, Dimaras H. Referral patterns for retinoblastoma patients in Ethiopia. *BMC Health Serv Res*. (2023) 23:172. doi: 10.1186/s12913-023-09137-9

39. Wang H, Elsheikh M, Gilmour K, Cohen V, Sagoo MS, Damato B, et al. Impact of Covid-19 pandemic on eye cancer care in United Kingdom. *Br J Cancer*. (2021) 124:1357–60. doi: 10.1038/s41416-021-01274-4

40. Haleem A, Javaid M, Singh RP. An era of Chatgpt as a significant futuristic support tool: a study on features, abilities, and challenges. *Bench Council Transact Benchmarks Stand Eval*. (2022) 2:100089. doi: 10.1016/j.tbench.2023.100089

41. Chen S, Kann BH, Foote MB, Aerts HJ, Savova GK, Mak RH, et al. The utility of Chatgpt for cancer treatment information. *medRxiv*. (2023):1–9 doi: 10.1101/2023.03.16.23287316

42. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. Gpt-4 and ophthalmology operative notes. *Ann Biomed Eng*. (2023) 51:1–3. doi: 10.1007/s10439-023-03263-5

43. Lawson Mclean A. Artificial intelligence in surgical documentation: a critical review of the role of large language models. *Ann Biomed Eng*. (2023) 51:1–2. doi: 10.1007/s10439-023-03282-2

44. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA* ophthalmology. (2023) 141:589. doi: 10.1001/jamaophthalmol.2023.1144

45. Pushpanathan K, Lim ZW, Yew SME, Chen DZ, Lin HAHE, Goh JHL, et al. Popular large language model Chatbots' accuracy, comprehensiveness, and self-awareness in answering ocular symptom queries. *iScience*. (2023) 30:108163. doi: 10.1016/j.isci.2023.108163

46. Gebrael G, Sahu KK, Chigarira B, Tripathi N, Mathew Thomas V, Sayegh N, et al. Enhancing triage efficiency and accuracy in emergency rooms for patients with metastatic prostate Cancer: a retrospective analysis of artificial intelligence-assisted triage using Chatgpt 4.0. *Cancers*. (2023) 15:3717. doi: 10.3390/cancers15143717

47. Kothari A. Chatgpt, large language models, and generative Ai as future augments of surgical Cancer care. *Ann Surg Oncol*. (2023) 30:1–3. doi: 10.1245/s10434-023-13442-2

48. Hu X, Ran AR, Nguyen TX, Szeto S, Yam JC, Chan CK, et al. What can Gpt-4 do for diagnosing rare eye diseases? A pilot study. *Ophthalmol Therapy*. (2023):1–8. doi: 10.1007/s40123-023-00789-8

49. Caruccio L, Cirillo S, Polese G, Solimando G, Sundaramurthy S, Tortora G. Can Chatgpt provide intelligent diagnoses? A comparative study between predictive models and Chatgpt to define a new medical diagnostic bot. *Expert Syst Appl*. (2024) 235:121186. doi: 10.1016/j.eswa.2023.121186

50. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the utility of Chatgpt throughout the entire clinical workflow: development and usability study. *J Med Internet Res*. (2023) 25:e48659. doi: 10.2196/48659

51. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of Chatgpt: survey analysis. *J Med Internet Res*. (2023) 25:e47184. doi: 10.2196/47184

52. Li J-PO, Liu H, Ting DS, Jeon S, Chan RP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. *Prog Retin Eye Res*. (2021) 82:100900. doi: 10.1016/j.preteyeres.2020.100900

53. Akkara JD, Kuriakose A. Role of artificial intelligence and machine learning in ophthalmology. *Kerala J Ophthalmol*. (2019) 31:150–60. doi: 10.4103/kjo.kjo_54_19

54. Alanzi TM. Impact of Chatgpt on teleconsultants in healthcare: perceptions of healthcare experts in Saudi Arabia. *J Multidiscip Healthc*. (2023) 16:2309–21. doi: 10.2147/JMDH.S419847

55. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of Chatgpt on Usmle: potential for Ai-assisted medical education using large language models. *PloS Digital Health*. (2023) 2:e0000198. doi: 10.1371/journal.pdig.0000198

56. Waters M, Aneja S, Hong J. Unlocking the power of Chatgpt, artificial intelligence, and large language models: practical suggestions for radiation oncologists. *Pract Radiat Oncol*. (2023) 13:e484–90. doi: 10.1016/j.prro.2023.06.011

57. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using Chatgpt to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spect*. (2023) 7:pkad015. doi: 10.1093/jncics/pkad015

58. Gu X, Xie M, Jia R, Ge S. Publication trends of research on retinoblastoma during 2001–2021: a 20-year bibliometric analysis. *Front Med*. (2021) 8:675703. doi: 10.3389/fmed.2021.675703

59. Rao R, Honavar SG. Retinoblastoma. *Indian J Pediatr*. (2017) 84:937–44. doi: 10.1007/s12098-017-2395-0

60. Kaliki S, Vempuluru VS, Ghose N, Patil G, Viriyala R, Dhara KK. Artificial intelligence and machine learning in ocular oncology: retinoblastoma. *Indian J Ophthalmol*. (2023) 71:424–30. doi: 10.4103/ijo.IJO_1393_22

61. Filho ARG, Singh N, Belfort RN. The role of telemedicine in ocular oncology In: *Global perspectives in ocular oncology*. BV Chawla, ME Aronow editors. New York, USA: Springer (2023)

62. Xue VW, Lei P, Cho WC. The potential impact of Chatgpt in clinical and translational medicine. *Clin Transl Med*. (2023) 13:e1216. doi: 10.1002/ctm2.1216

63. Chow JC, Sanders L, Li K. Impact of Chatgpt on medical chatbots as a disruptive technology. *Front Artifi Intell*. (2023) 6:1166014. doi: 10.3389/frai.2023.1166014

64. Li M, Zheng X. Identification of ancient Chinese medical prescriptions and case data analysis under artificial intelligence Gpt algorithm: a case study of song dynasty medical literature. *Ieee Access*. (2023) 11:131453–64. doi: 10.1109/ACCESS.2023.3330212

65. Liu J, Wang C, Liu S. Utility of Chatgpt in clinical practice. *J Med Internet Res*. (2023) 25:e48568. doi: 10.2196/48568

66. Villegas VM, Murray TG. Ocular oncology practice during the Covid-19 pandemic. *Global Perspect Ocular Oncol*. (2023). 405–13. doi: 10.1007/978-3-031-08250-4_40

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 19 December 2023
DOI 10.3389/fpubh.2023.1333776

Check for updates

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Daniele Spoladore,
Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, Italy
Azira Khalil,
Universiti Sains Islam Malaysia, Malaysia

*CORRESPONDENCE
Seung-Chan Kim
✉ seungk@g.skku.edu
Soo-Chul Lim
✉ limsc@dongguk.edu

RECEIVED 06 November 2023
ACCEPTED 28 November 2023
PUBLISHED 19 December 2023

CITATION
Yoon K-I, Jeong T-S, Kim S-C and Lim S-C (2023) Anonymizing at-home fitness: enhancing privacy and motivation with virtual reality and try-on.
Front. Public Health 11:1333776.
doi: 10.3389/fpubh.2023.1333776

Anonymizing at-home fitness: enhancing privacy and motivation with virtual reality and try-on

Kang-Il Yoon[1], Tae-Soo Jeong[1], Seung-Chan Kim[2]* and Soo-Chul Lim[1]*

[1]Department of Mechanical, Robotics, and Energy Engineering, Dongguk University, Seoul, Republic of Korea, [2]Machine Learning Systems Lab, College of Sports Science, Sungkyunkwan University, Suwon, Republic of Korea

Introduction: This study aimed to address privacy concerns associated with video conferencing tools used in home-based exercise training. To that end, a method that could anonymize participants' appearances and exercise environments during at-home fitness sessions was proposed.

Methods: This method combines virtual reality for 3-D human-model rendering using key-points tracking with a virtual try-on system enhanced by UV mapping and instance segmentation. To validate the proposed method, we conducted a user study by recruiting participants to assess effectiveness of virtual reality and virtual try-on in terms of privacy protection, self-confidence, and coaching satisfaction.

Results: Experimental results demonstrated the effectiveness and improved user experience of using virtual reality or virtual try-on in remote fitness, particularly in enhancing privacy protection and self-confidence with statistical significance. However, no significant differences were noted in coaching satisfaction.

Discussion: These findings confirmed the efficacy of our proposed approach. We believe that the proposed approach can significantly contribute to the future of remote fitness training, offering a more secure and engaging environment for users, thereby potentially increasing adherence to fitness regimens and overall physical wellbeing.

KEYWORDS

smart applications, fitness apps, virtual reality exercise, virtual try-on, physical health, physical activity

1 Introduction

Exercise can improve physical appearance. It is also widely recognized as an essential factor for enhancing and maintaining health in the long term. However, the proportion of individuals engaging in insufficient physical activity remains high worldwide, rendering them more susceptible to non-communicable diseases (1, 2). Furthermore, people in most countries faced restrictions of physical activity during the COVID-19 pandemic, resulting in significant reduction of overall physical activity levels (3–5). Although these restrictions have started to ease, physical activity levels of people have not recovered fully (6, 7).

In light of these concerns, home-based exercise programs have gained attention as potential alternatives as they can improve accessibility and save time. Home-based exercise is a useful solution for busy individuals who struggle to balance work and physical activity. For instance, home-based exercise has already been proven to be effective for various groups, such as female caregivers with limited time for exercise (8), frail individuals (9), and geriatric patients with cognitive impairments (10). In response to this demand, various online platforms have emerged and shared exercise videos that can be followed at home, benefiting many users. There is evidence that patients who undergo training using recorded exercise videos following COVID-19 hospitalization show encouraging cardiovascular, respiratory, and functional outcomes (11). In addition, home-based exercises can provide clinical benefits to individuals with severe SARS-CoV-2 infection (12). However, participants in these home-based exercise programs show a tendency to experience difficulty when first attempting such exercises.

Video conferencing technology has gained attention in home-based exercise training as it can provide real-time feedback coaching and supervised exercise programs. Exercise training using video conferencing can also provide motivation and offer personalized exercise programs that overcome distance and time constraints. Online exercise programs can provide guidance for independent workouts and monitor continuous exercise, enabling individuals to exercise conveniently from their own residences. Moreover, exercise becomes possible anytime and anywhere with internet connectivity, thus greatly mitigating limitations of time and location.

Home-based exercise training programs often rely on screen sharing for delivering exercise instructions. However, due to privacy concerns, this method may not be considered ideal. This is because exercise videos can contain personal information and sharing videos from home-based training sessions could potentially lead to unintended exposure of such details. In light of the growing popularity of video conferencing tools, recent research has focused on understanding vulnerabilities and risks associated with data privacy in this context (13, 14). Furthermore, other recent studies (15, 16) have highlighted privacy concerns among participants in home-based fitness training, which can lead to reluctance in video sharing.

Individuals participating in home-based exercise training often need to expose their clothing and appearance to others, potentially evoking Social Physique Anxiety (SPA). SPA is an emotional response that arises from concerns about others scrutinizing or judging their physical appearance (17). This anxiety is often triggered when individuals feel their physique does not match societal ideals of an "ideal body" (17). Elevated levels of SPA often correlate with lower adherence to physical activity and decreased participation rates in such programs (18). These individual psychological factors, including privacy concerns, social pressure, and self-consciousness, can make exercise participants hesitant to share their videos, with many preferring not to share their videos for these reasons (16). However, interventions aimed at reducing SPA might significantly enhance motivation and promote behaviors that encourage physical activity (17). Additionally, the concept of self-presentational efficacy (SPE) expectancy, defined as a subjective likelihood of successfully of conveying desired impressions to others, has been identified as a key factor influencing both exercise behavior and SPA (19). Elevating SPE expectancy could potentially be an effective strategy to encourage engagement in physical activity. It directly addresses SPA challenges in home-based exercise programs and aligns with findings that link SPA and SPE to physical activity participation (20).

In support of these findings, Lee et al. (21) have provided compelling evidence on benefits of using avatars in video conferencing. Their study demonstrated that avatars, rather than actual appearances, could enhance participant confidence and engagement. Their findings reveal that using avatars can foster better interactions, alleviate anxiety about appearances and behaviors, and promote increased concentration and willingness to engage in virtual communication. Similarly, observable changes in physical activity (PA) have been noted in virtual environments based on the selection of avatars. For example, it has been found that overweight children randomly allocated to avatars depicting a typical body size demonstrate enhanced performance in a PA-centric video game (22). This cohort also displayed heightened exercise motivation and a more favorable disposition toward PA compared to their counterparts assigned to obese avatars. Likewise, women assigned to slender avatars during engagement in a tennis exergame exhibited heightened PA compared to those assigned to obese avatars (23). These findings are reproducible within men cohorts, wherein an association exists between increased PA and allocation to slim rather than overweight avatars (24). This aligns with the notion that modifying virtual self-representation, such as through avatars, could be an innovative approach to manage SPA and improve exercise adherence. From this perspective, this paper presents a method to explore potential solutions for promoting exercise participation and addressing privacy concerns associated with home-based exercise training.

This study primarily focused on privacy concerns associated with video conference technology used in home-based exercise training programs. Additionally, we aimed to improve privacy protection while exploring approaches to reduce the burden caused by exposing the exercise environment and process, with the goal of enhancing exercise motivation. To achieve this, we first identified vulnerabilities of current video conference technologies used and proposed alternatives to address privacy concerns. Furthermore, we identified factors contributing to perceived burden in an exercise environment and process to suggest mitigation strategies. By doing so, we seek to find safe and effective methods to utilize video conference technologies in home-based exercise training while simultaneously ensuring privacy protection and promoting user engagement during exercise.

This study used a novel approach that could leverage VR technology (3D human-model rendering based on 3D key-points tracking) and virtual try-on network (UV mapping-enhanced instance-segmentation-based virtual try-on network: UVI-VTON) to mitigate constraints associated with user privacy protection effectively while alleviating the burden experienced by users in the context of screen sharing for educational purposes. The remainder of this manuscript is organized as follows. Section 2 presents a review of related research from perspectives of VR-based conference and VR-based exercise. Section 3 introduces applications of VR and UVI-VTON in home-based exercise training. Section 4 presents an empirical study evaluating the usability of the proposed approach. Sections 5 describes statistical analysis results and discussion of the usability survey. Section 6 then summarizes key findings of this study and suggests some future research directions.

2 Related work

This study was inspired by existing research incorporating VR technologies, which could be broadly categorized into two groups: VR and VTON.

2.1 Virtual reality

Currently, VR technology is gaining attention as a means of promoting and assisting physical activity. The use of VR in exercise has potential to generate more favorable effects on physiological, psychological, and rehabilitative outcomes of individuals than conventional exercise methodologies (25). Prior to the COVID-19 pandemic, these studies were already underway, showcasing the efficacy of virtual reality interventions. For instance, a comparative study between virtual reality game exercises and ball exercises has highlighted superior benefits of VR games in enhancing balance abilities of older adults (26). Arlati et al. have introduced VR-based SocialBike, tailored to improve clinical outcomes in older adults (27). Furthermore, a study focusing on individuals with cardiovascular diseases has demonstrated significant enhancements in body composition, lipid profile, and eating patterns through cardiac rehabilitation within a virtual reality environment in comparison with conventional rehabilitation methods (28).

Since the onset of the COVID-19 pandemic, there has been a resurgence of interest in this research domain. Previous research has indicated that virtual reality exercise and fitness apps can be used as effective strategies for enhancing physical and mental wellbeing during the pandemic period (29). Additionally, during COVID-19 lockdowns, empirical studies have demonstrated that the use of smart applications, including live streaming exercise classes and virtual reality fitness programs, has a positive impact by promoting physical activity (30). The necessity of replacing outdoor exercise with a viable alternative has been increasingly emphasized since the COVID-19 outbreak. VR has stood out as a pivotal alternative to outdoor exercise, demonstrating a substantial capacity to improve cognitive function and mitigate motor disabilities (31). In particular, there have been active discussions on the necessity and effectiveness of alternative exercises in academic circles for individuals facing difficulties with outdoor activities, such as older adults (32), patients with post-COVID-19 condition (33), patients recovering from COVID-related pneumonia (34), physically inactive individuals during the COVID-19 lockdown (30), and patients diagnosed with Parkinson's disease (35). In addition, both specialized and gaming VR can be effective in treating upper-extremity impairments, with specialized VR showing potential for improving balance in patients with neurological conditions (36). In contrast to studies focusing on individuals with specific physical limitations, the study by Ng et al. (37) investigated the impact of VR- or augmented reality (AR)-enhanced training on physical activity in healthy individuals through a meta-analysis to. It demonstrated that VR intervention had substantial effects on physical activity levels, minor effects on performance, but no effects on psychological outcomes.

2.2 Virtual try-on network

VTONs leverage advanced computer-vision and machine-learning techniques to simulate the process of virtually trying on clothing items. By overlaying virtual garments on a user's body, individuals can visualize how different clothing items would look on them without actually wearing them. Recent advancements in deep-learning-based VTONs have shown significant improvements in clothing synthesis performances. For instance, context-driven virtual try-on network (C-VTON) (38) employs discriminators specific to different types of contextual information, allowing enhanced clothing synthesis quality. The dual-branch collaborative transformer (DBCT) (39) utilizes a transformer-based architecture and cross-modal information to improve the virtual try-on process. To address the alignment of target garments to corresponding body parts, a novel global appearance flow estimation model has been proposed, which warps clothing spatially (40). The Dress Code dataset (41) containing images of diverse categories of clothes has also been introduced. This dataset enables the generation of high-resolution try-on images. These virtual try-on technologies have been utilized for virtual fitting in online shopping. Several studies have investigated their impacts. Research has shown that attitudes of consumers toward virtual try-on can influence their online purchasing behaviors (42). Furthermore, a review of literature has shown a wide range of psychological (cognitive, emotional, and social) and behavioral consequences associated with virtual try-on in the context of shopping (43).

3 Methods

The aim of this study was to develop a methodology utilizing deep-learning-based models to safeguard the privacy of individuals engaging in home-based exercises and to encourage exercise participation. Specifically, the methodology consisted of two key features: (1) VR (3D human-model rendering based on 3D key-points tracking); and (2) virtual try-on network (UVI-VTON). The 3D human model rendering was designed for real-time motion tracking and avatar rendering based on images of individuals engaged in movement. It was generated by predicting joint positions of users using deep-learning algorithms and mapping them to a prebuilt human model. The UVI-VTON was developed to enhance video generation speed compared to the conventional virtual try-on network, enabling real-time video generation while improving video quality. It could generate an image of the user wearing different clothing, allowing them to exercise without being self-conscious about their appearance.

To evaluate the effectiveness and usability of this methodology, we conducted experiments and a usability survey. Experiments involved testing performances of deep-learning models used in real-world scenarios. The usability survey involved evaluating user experiences of the methodology, including user satisfaction and feedback assessments on coaching effectiveness from the perspectives of privacy protection and motivation. In this section, we will describe each of these components in detail, including methods used, results obtained, and implications for future development.

3.1 VR: 3D human-model rendering

The skinned multi person linear (SMPL) model widely used in computer vision was employed as the 3D human model to construct human shapes. The SMPL model comprised a 3D body mesh that could transform shapes of the joints and muscles of a person to resemble their actual appearance depending on body shape and key points. The SMPL model could be rendered from estimated parameters based on real-time image frames. Estimated SMPL parameters could then be used with deep-learning-based 3D key-points estimation models. This instance 3D key points estimation model has been discussed in various ways in the field of computer vision. As shown in Figures 1A, B, HRNet was adopted (44) to detect and track 3D key points of a person from a single image frame.

The SMPL model represents the shape of a person without clothing. Using the generated model as an avatar can potentially make users feel uncomfortable and unpleasant. In this research, to create a more comfortable and confident representation of the user, the avatar representing the user tried on clothing. Furthermore, various clothing designs were implemented with the ability for color selection to reflect individuality of the user to achieve effects discussed in a previous study (45). This demonstrate that increased avatar personalization can lead to greater body ownership, presence, and emotional responses. As shown in Figures 1C, D, the SNUG model (46) was adopted to try on clothes over vertices of the SMPL mesh. This model utilizes physical loss to depict dynamic aspects of clothing by taking into account movement speed of the human body, resulting in a more realistic representation of garments. Three types of top and two types of bottom apparel were applied to the SMPL model (47) using this approach.

3.2 UVI-VTON: user image regenerated with reference garments using generative adversarial network

For training the UVI-VTON, we captured movements of subjects performing yoga and weight training. These two exercise types were used in our home-based training application. One experimental group composed of 14 males and 7 females who performed weight training. Another experimental group consisted of 6 males and 14 females engaged in yoga, resulting in the collection of exercise videos from a total of 41 participants. Each participant executed 10 prescribed movements specified in Table 1 for a single exercise category, i.e., either yoga or weight training. For each exercise movement, participants alternated between wearing four sets of upper- and lower-body clothing and performed exercise movements iteratively. This process resulted in a total of 40 sets of exercises (1 exercise category × 10 movements × 4 clothing sets) performed by each participant. As shown in Figure 2, these scenes were captured using a setup of eight synchronized cameras (GoPro Hero5) positioned to face the subject, with each video recorded at 30 frames per second. The resulting dataset contained a total of 20,606,496 image frames (8 cameras × 2,575,812 image frames extracted from the exercise videos of 41 people) that captured activities of the 41 participants across the 8 cameras. This dataset was used to train the model.

UVI-VTON consisted of three main components: a segmentation predictor, a try-on network for the top (TON-T), and a try-on network for the bottom (TON-B). The segmentation predictor generated a segmentation image from the raw image, where areas for wearing top and bottom garments were indicated explicitly. This segmentation image was utilized by the two try-on networks to try on respective clothing items.

As shown in Figure 3A, the UV field was extracted from the raw image (I^{raw}) using Densepose (48), enabling pixel-level recognition of human body parts in the image and estimation of their 3D positions and orientations, which were then visualized as density maps. Subsequently, the segmentation predictor applied segmentation maps to this UV field, generating a segmentation image I^{seg}) that delineated a total of 12 distinct body parts: top, bottom, head, neck, left arm, right arm, left hand, right hand, left leg, right leg, left foot, and right foot.

As shown in Figure 3B, the top garment (G^{top}) was processed through the garment warping network (GWN), resulting in a warped top-garment image ($I^{top}_{garment}$) that was fed to TON-T. Additionally, the raw image (I^{raw}) was segmented to produce the semantic segmentation image (I^{seg}), generating parts for preservation ($I^{top}_{preserv}$) and skin parts for regeneration (I^{top}_{skin}) in the process of trying on the top. Both these images were used as inputs to TON-T. Applied inputs were then synthesized through TON-T, resulting in the predicted top-replaced image (I^{top}_{pred}). This process could simulate a person trying on the top garment to produce I^{top}_{pred} wherein the person appeared to be wearing the replaced top.

To try on G^{top}, semantic segmentation was utilized and divided into three masks: a mask for G^{top} ($M^{top}_{garment}$), a mask for upper-body skin (M^{top}_{skin}) consisting of both arms and neck, and a preservation mask ($M^{top}_{preserv}$) consisting of the bottom, head, both hands, both legs, and both feet. The GWN was used to warp G^{top} to match the shape of $M^{top}_{garment}$, resulting in $I^{top}_{garment}$. I^{raw} was divided into preserved parts ($I^{top}_{preserv}$) determined by $M^{top}_{preserv}$ and regenerated skin parts (I^{top}_{skin}) of the upper body defined by M^{top}_{skin}. Consequently, $I^{top}_{preserv}$, I^{top}_{skin}, and $I^{top}_{garment}$ were combined as inputs to TON-T to predict I^{top}_{pred}.

TON-B had the same pipeline as TON-T, with the only difference being the reference garments (top or bottom) that users tried on. Similarly, the bottom garment (G^{bottom}) was processed through the GWN, resulting in a warped bottom-garment image ($I^{bottom}_{garment}$) that was then fed to TON-B. Additionally, I^{top}_{pred} was segmented by I^{seg} to generate parts for preservation ($I^{bottom}_{preserv}$) and skin parts for regeneration (I^{bottom}_{skin}) in the process of trying on the bottom. Both these images were then used as inputs to TON-B. These inputs were then synthesized through TON-B, resulting in the predicted top-and-bottom-replaced image ($I^{top,\ bottom}_{pred}$). This process simulated a person trying on the top and bottom garments, resulting in $I^{top,\ bottom}_{pred}$ wherein the person appeared to be wearing the replaced top and bottom apparel.

To try on G^{bottom}, semantic segmentation was utilized and divided into three masks: a mask for G^{bottom} ($M^{bottom}_{garment}$), a mask for lower-body skin (M^{bottom}_{skin}) consisting of both legs, and a

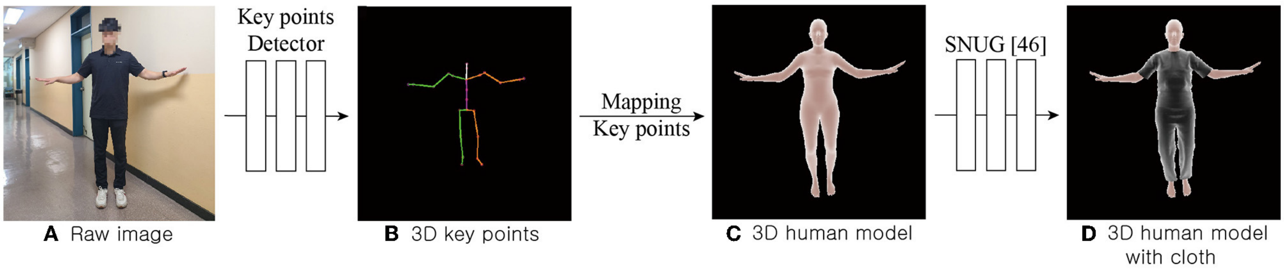


FIGURE 1
Procedure for 3D human modeling with cloth. Given the raw frame **(A)**, the detected 3D key points **(B)** are mapped to a 3D human model **(C)** constructed using SMPL model (47). Subsequently, the SNUG model (46) is used to apply clothing to the corresponding 3D human model to render a 3D human model with cloth **(D)**.

TABLE 1 **Order of exercises in home-based exercise training.**

	Week	Training poses
Weight training	Week 1	Squat—Lunge—Push-up—Bird Dog—Plank
	Week 2	Bridge—Shoulder Press—Side Lateral Raise—Side Plank—Mountain Climber
	Week 3	Push-up—Plank—Shoulder Press—Squat—Side Plank
	Week 4	Side Lateral Raise—Lunge—Bridge—Bird Dog—Mountain Climber
Yoga	Week 5	Seated Forward Bend Pose—Half Moon Pose—Triangle Pose—Warrior 2—Cat Cow Pose
	Week 6	Half Moon Pose—Pelvic Opening—Warrior 1—Warrior 2—Triangle Pose—Chair Pose—Side Plank—Cobra Pose
	Week 7	Half Moon Pose—Triangle Pose—Chair Pose—Warrior 2—Cobra Pose
	Week 8	Seated Forward Bend Pose—Cat Cow Pose—Pelvic Opening—Warrior 1—Side Plank

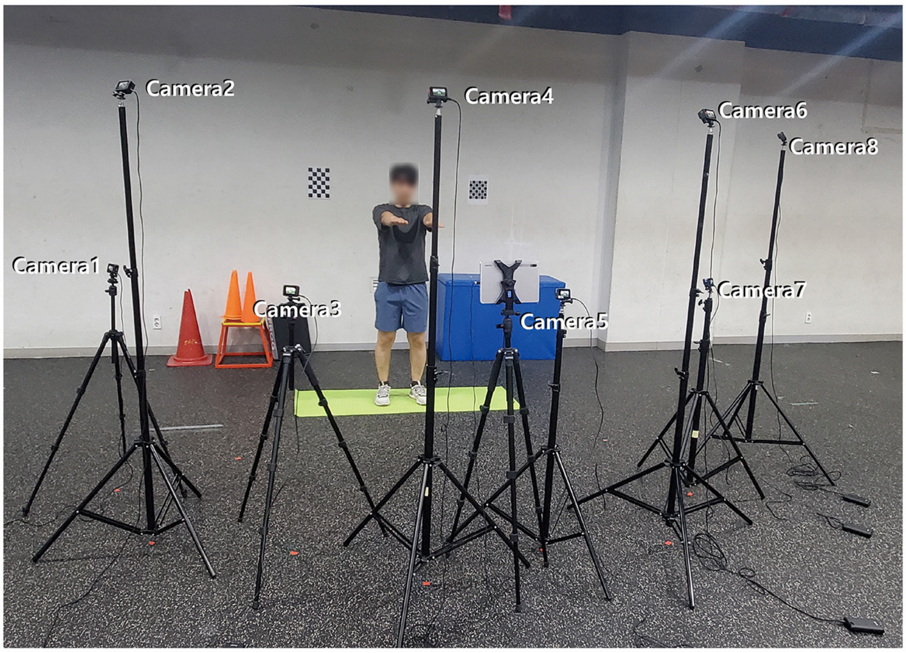


FIGURE 2
Experimental setup with eight installed cameras that are situated at distinct angles and oriented toward the subject.

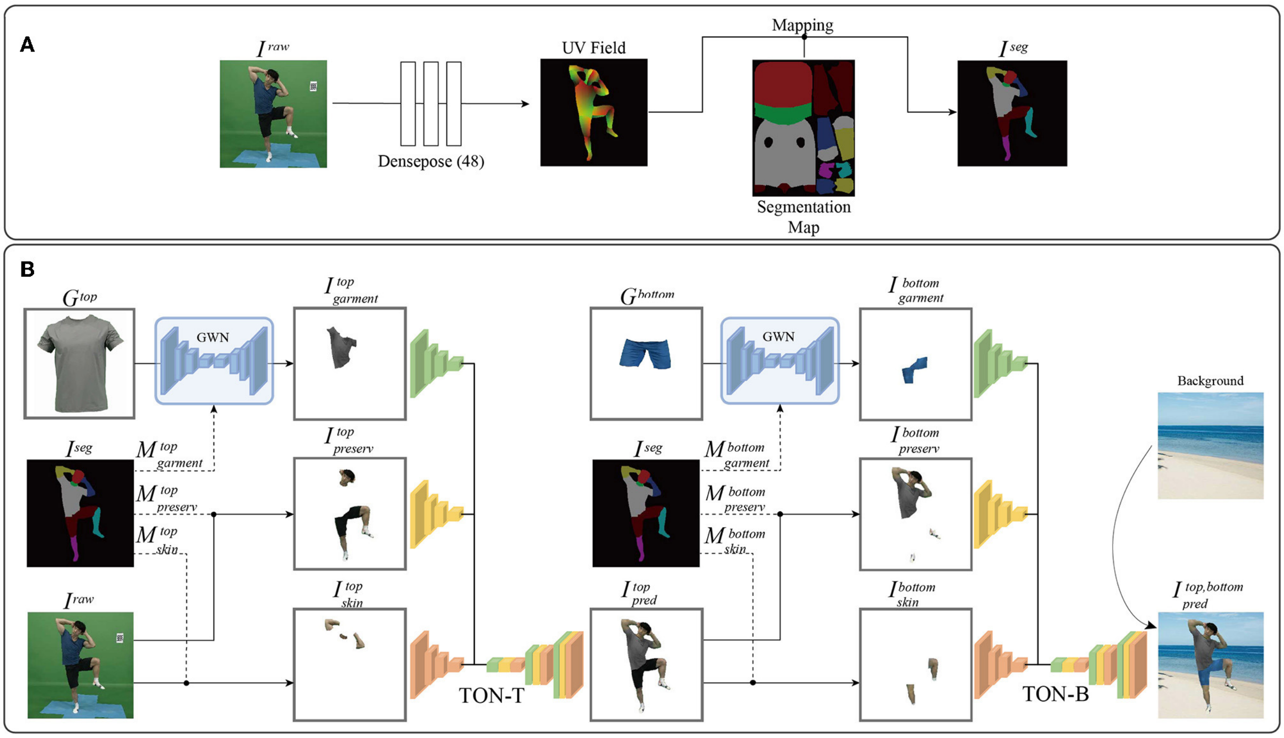


FIGURE 3
(A) Segmentation predictor: It aims to generate a semantic segmentation image (I^{seg}) in real time from a raw image (I^{raw}) by applying the segmentation map to the UV field predicted by Densepose (48). **(B)** Overview of the proposed network: The semantic segmentation image (I^{seg}) is used to try G^{top} on I^{raw}, where the garment warping network (GWN) warps G^{top} to $I^{top}_{garment}$ and inputs it to TON-T. Then, I^{raw} is divided into the parts for preservation ($I^{top}_{preserv}$) and skin parts for regeneration (I^{top}_{skin}), which are used as inputs to TON-T to generate I^{top}_{pred}. Similarly, I^{seg} is employed to try G^{bottom} on I^{top}_{pred}, where the GWN warps G^{bottom} to $I^{bottom}_{garment}$ and inputs it to TON-T. Subsequently, I^{top}_{pred} is divided into the parts for preservation ($I^{bottom}_{preserv}$) and skin parts for regeneration (I^{bottom}_{skin}), which are used as inputs to TON-B to generate $I^{top,\ bottom}_{pred}$.

preservation mask ($M^{bottom}_{preserv}$) consisting of the top, head, both arms, both hands, and both feet. The GWN was then used to warp G^{bottom} to match the shape of $M^{bottom}_{garment}$, resulting in $I^{top,\ bottom}_{garment}$. This I^{top}_{pred} was then divided into preserved parts ($I^{bottom}_{preserv}$) determined by $M^{bottom}_{preserv}$ and regenerated skin parts (I^{bottom}_{skin}) of the lower body defined by M^{bottom}_{skin}. Consequently, $I^{bottom}_{preserv}$, I^{bottom}_{skin}, and $I^{bottom}_{garment}$ were combined as inputs to TON-B to predict I^{bottom}_{pred}.

To train the proposed network, an objective loss function was constructed by combining four loss components: L2 loss, VGG19-based perceptual loss, second-order smooth constraint loss (49), and adversarial loss computed using the pix2pixHD discriminator (49).

4 Experiments

To assess the effectiveness of UVI-VTON and VR, participants were recruited to perform and evaluate home-based exercise training programs using the developed application. Test exercises consisted of weight training and yoga guided by a coach such that participants had the opportunity to try out proposed methodologies instructed by the coach. After completing the exercise session, participants responded to a usability survey.

4.1 Participants

Experiments to test UVI-VTON and VR comprised a total of 35 participants, including 5 males and 30 females. These participants aged from 18 to 48 years, with a mean age of 30 (±8) years. Participants were chosen based on self-selection through recruitment advertisements. All participants were enrolled after this study obtained Institutional Review Board (IRB) approval (Sungkyunkwan University IRB approval number: SKKU 2021-12-014).

4.2 Procedures

Participants were assigned to five different teams, each having distinct time schedules. Participants joined teams based on their preferred time slots for the experiment. Simultaneously, efforts were made to minimize disparities in team sizes, aiming to achieve a balanced distribution of members among teams. Each team had six to eight members. Experiments were conducted once a week for 8 weeks. During each session, participants engaged in a total of eight exercise sessions, alternating between four sessions each of weight training and yoga, each lasting 30 min. Two coaches, one specializing in yoga and the other in weight training, participated in the experiment. To enhance experiences of participants in home-based exercise training, types and order of exercises were varied over a 4-week period as depicted in Table 1.

The coach, situated in a prepared studio, shared their own video feed with participants through a camera, while participants shared their video feeds using their mobile devices from the comfort of their homes. The coach provided instructions for prescribed exercises in a sequential manner. Upon demonstrating each exercise, participants repeated movements accordingly. Simultaneously, participants adjusted their exercise postures in accordance with instructions from the coach, allowing for personalized training modifications. To assess the impact of utilizing UVI-VTON and VR in home-based exercise training compared to not using any methodology, participants engaged in training sessions employing UVI-VTON and VR as well as sessions where no specific methodology was employed. As shown in Figures 4, 5, when utilizing methodologies, participants activated either UVI-VTON or VR upon request from the coach, thereby sharing their real-time modified visual representations on the screen.

4.3 Usability survey

Participants were asked to respond to a survey based on their own experiences with home-based exercise training after an 8-week period of participation. As depicted in Table 2, the usability survey was categorized under three evaluation factors: privacy protection scale (PPS), self-confidence and motivation scale (SMS), and coaching satisfaction scale (CSS). The PPS questionnaire comprehensively assessed participants' perceived burden of camera recording and the extent to which their privacy was protected. This survey comprised three items rated on a scale of 1 (strongly disagree) to 5 (strongly agree) by participants. The SMS questionnaire measures self-confidence and motivation effects derived from home-based exercise training. This questionnaire consisted of two items rated on a scale of 1 (strongly disagree) to 5 (strongly agree) by participants. The CSS questionnaire measures participants' perceptions of whether their exercise performances were accurately conveyed to the instructor and whether they received appropriate feedback based on their experiences in home-based exercise training. This questionnaire comprised two items rated on a scale of 1 (strongly disagree) to 5 (strongly agree) by participants. The raw data of usability survey is depicted in Table 3. Usability survey data were analyzed using one-way analysis of variance (ANOVA). Significant interactions and main effects were investigated through Bonferroni-corrected pairwise comparisons. A significance level of $p < 0.05$ was considered to be statistically significant. Precise values of p are reported unless $p < 0.001$. Descriptive statistics for all data are displayed as mean ± standard deviation (SD) with 95% confidence intervals (CIs).

5 Results

Results of the ANOVA conducted for PPS, SMS, and CSS of home-based exercise training with VR or UVI-VTON indicated no significant effects, suggesting that both methodologies elicited similar emotional responses. However, when home-based exercise training with VR or UVI-VTON were compared to home-based exercise training without VR or UVI-VTON, significant differences in PPS, SMS, and CSS were observed.

When conducting home-based exercise training, significant differences were observed for question 1 of PPS among the use of UVI-VTON, VR, and no intervention [$F_{(2,68)} = 15.880; p < 0.001$]. *Post-hoc* analysis revealed no significant difference ($p = 0.095$) between the use of UVI-VTON and VR for question 1. However, the use of UVI-VTON showed a significantly higher rating of 0.514 ($p = 0.003$) than the use of nothing for question 1. The use of VR also exhibited a significantly higher rating of 0.686 ($p < 0.001$) than the use of nothing for question 1.

When conducting home-based exercise training, significant differences were observed in question 2 of PPS among the use of UVI-VTON, VR, and nothing [$F_{(2,68)} = 20.035; p < 0.001$]. *Post-hoc* analysis revealed no significant difference ($p = 0.069$) between the use of UVI-VTON and VR for question 2. However, the use of UVI-VTON showed a significantly higher rating of 0.600 ($p < 0.001$) than the use of noting for question 2. The use of VR also exhibited a significantly higher rating of 0.743 ($p < 0.001$) than the use of nothing for question 2.

When conducting home-based exercise training, significant differences were observed in question 3 of PPS among the use of UVI-VTON, VR, and nothing [$F_{(2,68)} = 16.302; p < 0.001$]. *Post-hoc* analysis revealed no significant difference ($p = 1.000$) between the use of UVI-VTON and VR for question 3. However, the use of UVI-VTON showed a significantly higher rating of 0.971 ($p < 0.001$) than the use of nothing for question 3. The use of VR also exhibited a significantly higher rating of 1.000 ($p < 0.001$) than the use of nothing for question 3.

When conducting home-based exercise training, significant differences were observed in question 4 of SMS among the use of UVI-VTON, VR, and nothing [$F_{(2,68)} = 7.299; p = 0.006$]. *Post-hoc* analysis revealed no significant difference ($p = 1.000$) between the use of UVI-VTON and VR for question 4. However, the use of UVI-VTON showed a significantly higher rating of 0.686 ($p = 0.0133$) than the use of nothing for question 4. The use of VR also exhibited a significantly higher rating of 0.629 ($p = 0.042$) than the use of nothing for question 4.

When conducting home-based exercise training, significant differences were observed in question 5 of SMS among the use of UVI-VTON, VR, and nothing [$F_{(2,68)} = 8.279; p = 0.003$). *Post-hoc* analysis revealed no significant difference ($p = 0.506$) between the use of UVI-VTON and VR for question 5. However, the use of UVI-VTON showed a significantly higher rating of 0.743 ($p = 0.004$) than the use of nothing for question 5. The use of VR also exhibited a significantly higher rating of 0.600 ($p < 0.039$) than the use of nothing for question 5. When conducting home-based exercise training, there were no significant differences in question 6 of CSS among groups using UVI-VTON, VR, and no intervention [$F_{(2,68)} = 2.121; p = 0.136$].

The present study aimed to investigate emotional responses of participants regarding home-based exercise training with or without the use of UVI-VTON or VR. This study is a pioneering effort to develop and apply VTON and VR in home-based exercise training. Understanding impacts of these new methodologies on

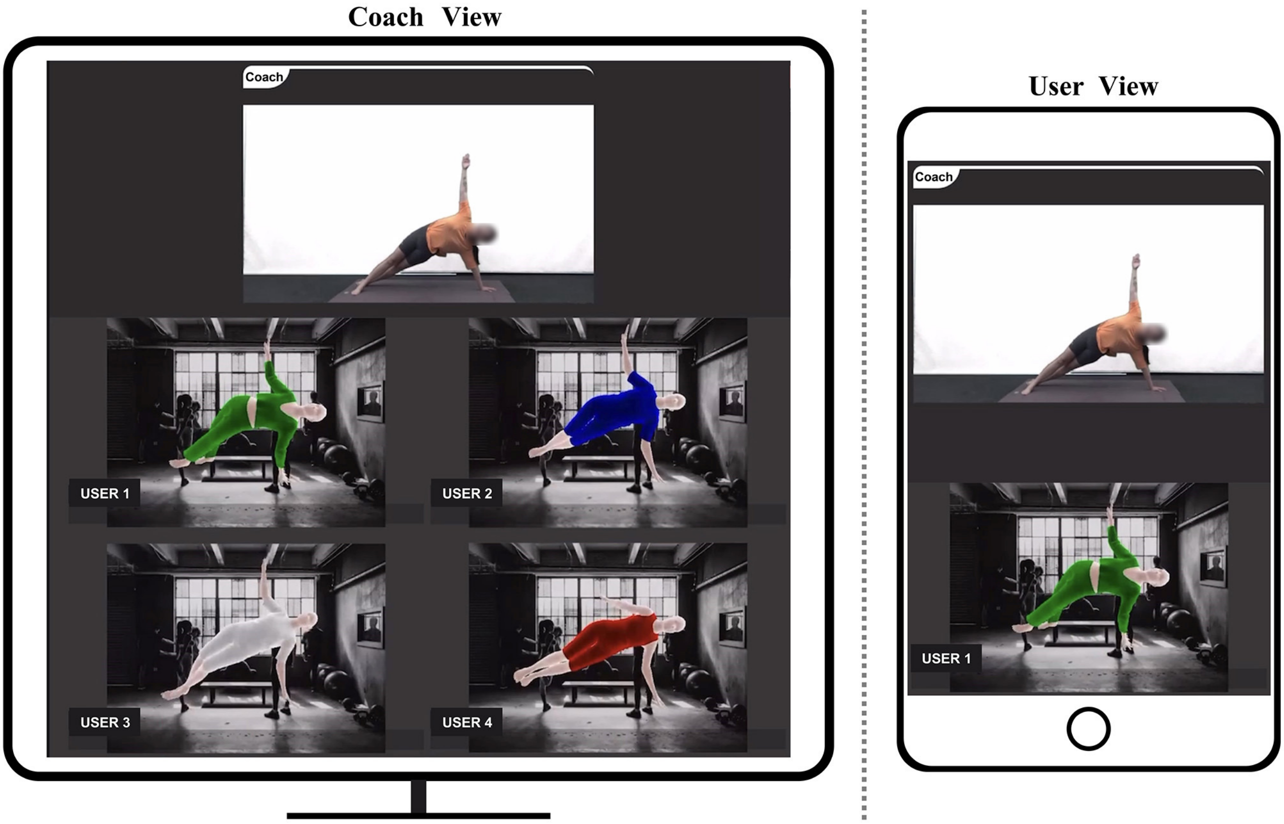


FIGURE 4
Home-based exercise training with VR (3D human-model rendering based on 3D key-points tracking).

participants can help us develop and provide more effective exercise programs.

In this study, three measurement criteria were used to evaluate emotional responses of participants, namely PPS, SMS, and CSS. PPS was used to assess perceptions regarding privacy protection. SMS was used to measure confidence and motivation levels of users and CSS was used to evaluate user satisfaction with exercise posture and feedback.

As seen in Figure 6, significant differences in PPS were observed when comparing the use of UVI-VTON vs. non-use, indicating that UVI-VTON had a positive impact on privacy protection. Similarly, significant differences in PPS were observed between the use of VR vs. non-use, suggesting that VR could also have a positive impact on privacy protection. These research findings demonstrate that the implementation of UVI-VTON or VR can help alleviate concerns regarding user privacy while enhancing trust in exercise programs. Moreover, results indicate the potential of UVI-VTON or VR to improve user exercise experiences and contribution to personal data protection, thus creating a safer and more comfortable exercise environment. Analysis results also revealed no significant differences in PPS between the use of UVI-VTON and VR. This suggests that both methodologies could elicit similar emotional responses in relation to measured criteria for PPS. While UVI-VTON covers all parts except the face, VR obscures all information captured by the camera. One might expect that the PPS would be significantly higher in VR than UVI-VTON owing to its ability to conceal more than just the face. However, when comparing the use of UVI-VTON vs. the use of VR, the absence of a significant difference in questions 4 and 5, as depicted in Figure 7, implies that concerns regarding personal data exposure primarily stem from the exposure of one's body and background while wearing clothing, excluding the face.

From Figure 7, when comparing the use of UVI-VTON vs. non-use, significant differences were observed for SMS comprising questions 4 and 5. This indicates that UVI-VTON has a positive impact on confidence enhancement and motivation. Similarly, significant differences were observed for SMS when comparing the use of VR vs. non-use. This suggests that VR also has a positive impact on confidence enhancement and motivation. These research findings highlight that engaging in exercise while utilizing UVI-VTON or VR technology provides a means to conceal personal appearance, enabling users to perform their exercises with enhanced self-assurance. This aligns with the goal of improving the exercise environment in the proposed home-based exercise training program, which aims to encourage and motivate users to participate in physical activities. Furthermore, analysis results revealed no significant differences in SMS between the use of UVI-VTON and VR. This suggests that both methodologies can elicit similar emotional responses when measured against the criteria for SMS. Similar to findings from the analysis of PPS responses, these results indicate that methodologies capable of obscuring users' bodies except their face and background

10.3389/fpubh.2023.1333776

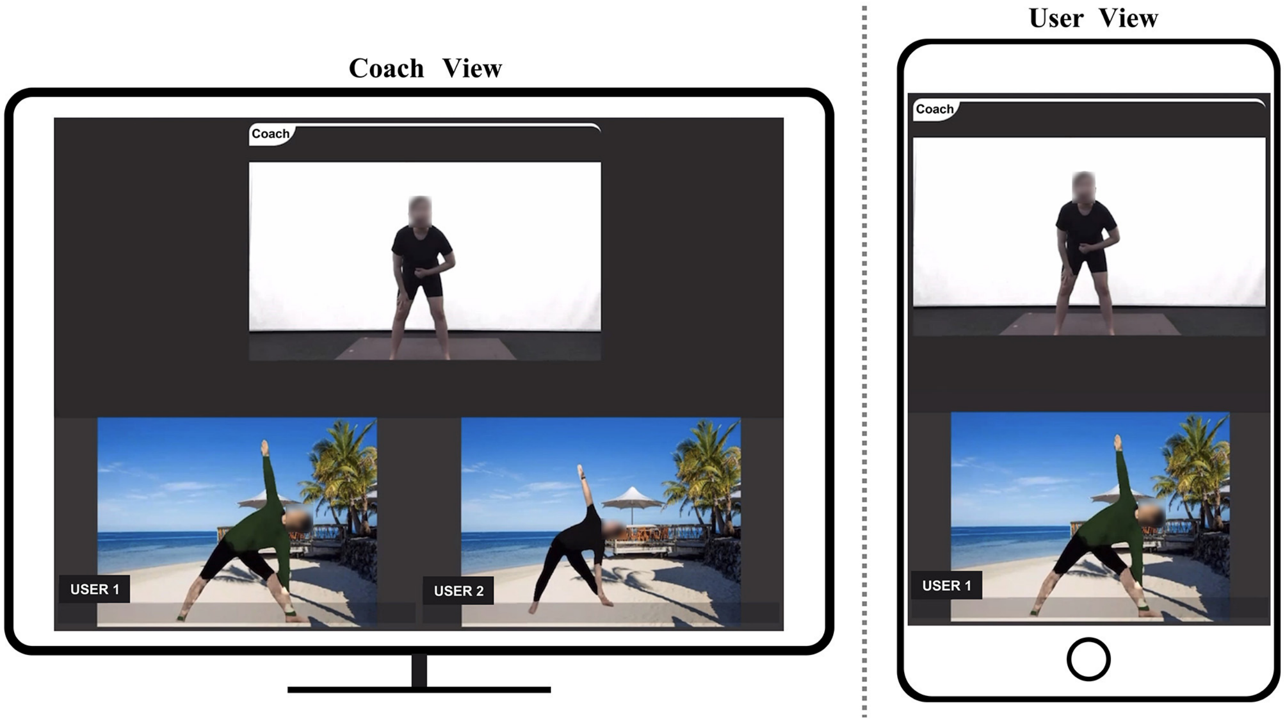


FIGURE 5
Home-based exercise training with UVI-VTON (UV mapping-enhanced instance-segmentation-based virtual try-on network).

TABLE 2 Survey questions administered upon completion of home-based exercise training.

	Question
Privacy Protection Scale (PPS)	1. Is the user privacy protected effectively during home-based exercise training?
	2. Is the user privacy adequately ensured during home-based exercise training?
	3. Is there no burden associated with camera recording for home training purposes?
Self-confidence and Motivation Scale (SMS)	4. Did you feel confident during the exercise?
	5. Did you feel motivated to exercise?
Coaching Satisfaction Scale (CSS)	6. Are you satisfied with the quality of the instructions from the coach for your exercise posture and feedback?

have potential to enhance self-confidence and promote exercise participation. Additionally, Figure 7 shows that there are no significant differences among the three cases for CSS comprising question 6. This indicates that there is no qualitative difference in guidance by an exercise coach when compared with delivery of postprocessed exercise videos in home-based exercise training using UVI-VTON and VR against the delivery of raw exercise videos without using these methodologies. This finding suggests the potential of UVI-VTON and VR to replace existing camera-based home-based exercise training programs. These results further demonstrate the potential of UVI-VTON and VR to provide positive experiences for individuals participating in home-based exercise training. UVI-VTON and VR can enhance self-confidence, increase motivation, and strengthen privacy protection, which can in turn improve the quality of exercise experiences and help participants maintain healthy lifestyles by fostering a willingness to continue exercising.

Findings of this study show that applying new methodologies to home-based exercise training can improve emotional responses of participants. Specifically, UVI-VTON and VR have positive impacts on privacy protection, confidence enhancement, and exercise motivation. These findings can help promote exercise training participation in the home environment. Advancements made in the home-based exercise training environment through this research are expected to facilitate greater accessibility for a wider population. Furthermore, this study serves as a crucial indicator for developing methodologies aimed for at enhancing home-based exercise training environment.

6 Conclusions

This research aims to help individuals who rely on home-based exercise, encompassing patients, older adults, and

TABLE 3 Raw data of usability survey (mean ± SD).

	PPS			SMS		CSS
	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6
Use of nothing	4.14 ± 0.97	4.11 ± 0.96	3.63 ± 1.40	3.86 ± 1.22	3.89 ± 1.32	4.71 ± 0.53
Use of UVI-VTON	4.66 ± 0.64	4.71 ± 0.57	4.60 ± 0.74	4.54 ± 0.78	4.63 ± 0.73	4.63 ± 0.84
Use of VR	4.83 ± 0.45	4.86 ± 0.43	4.63 ± 0.73	4.49 ± 0.85	4.49 ± 0.82	4.51 ± 0.85

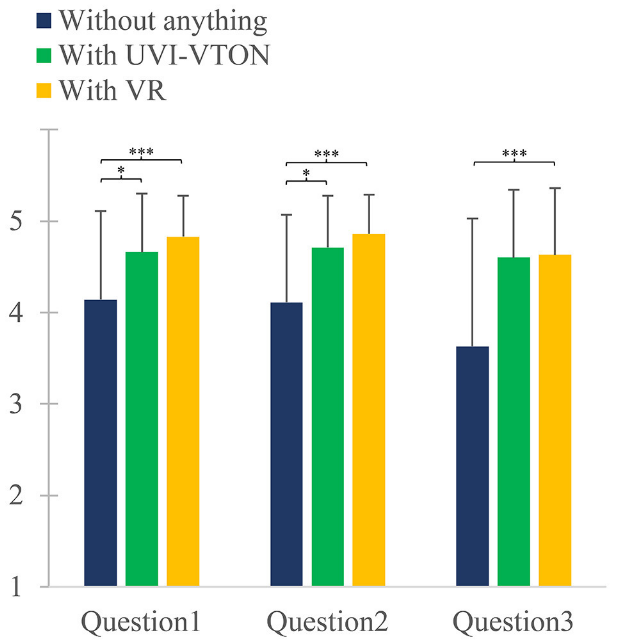


FIGURE 6
Privacy Protection Scale (PPS) consisting of three questions (Question 1, Question 2, Question 3) for all participants. PPS serves as a tool to assess participants' privacy protection levels based on the presence or absence of UVI-VTON and VR. Vertical error bars represent standard deviation for response for each question. *$p < 0.05$, ***$p < 0.001$.

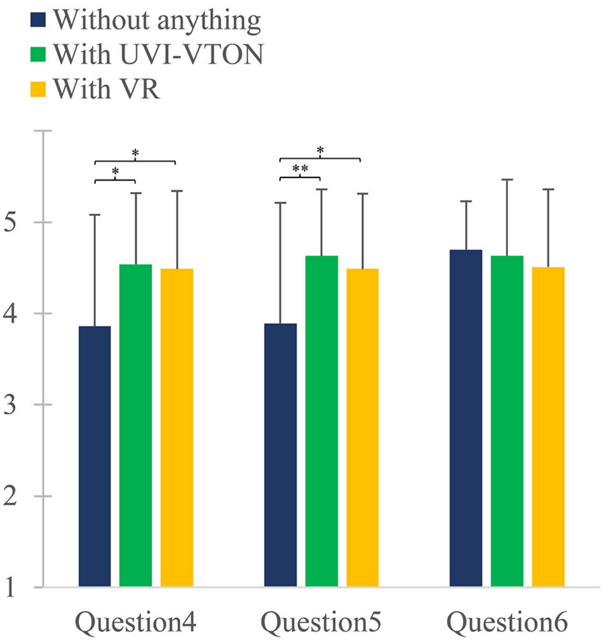


FIGURE 7
Self-confidence and Motivation Scale (SMS) comprising questions 4 and 5 is employed to assess participants' self-confidence and motivation levels with consideration of the presence or absence of UVI-VTON and VR. Additionally, the Coaching Satisfaction Scale (CSS), comprising a single question (Question 6), is used to gauge participants' coaching satisfaction level, taking into account the presence or absence of UVI-VTON and VR. Vertical error bar represents standard deviation for response for each question. *$p < 0.05$; **$p < 0.005$.

individuals in need of courage and motivation to engage in physical activities post COVID-19. This study seeks to enhance home-based exercise training experience with the objective of facilitating participant engagement while alleviating their concerns regarding privacy and emotional burden associated with home-based training. To achieve this, we proposed the use of VR and UVI-VTON. Participants were recruited to engage in home-based exercise training using the proposed approach. A usability survey was administered upon training completion to assess the effectiveness and user experiences. Results demonstrated the potential of the proposed method to improve home-based exercise training and enhance the overall exercise participation experience in terms of privacy protection, enhanced self-confidence, motivation, and real-time coaching instructions. These findings suggest that VR or UVI-VTON can be effectively employed to enhance user experiences in home-based exercise training.

7 Limitations

This study has a limitation. Since this study was conducted with a demand group for home training, which was dominated by women, it resulted in an unbalanced sample, which might limit the generalizability of results. Therefore, future studies over an extended period with a larger and more diverse participant base while considering factors such as gender, age, and physical attributes need to be conducted to enhance the reliability and potential applicability of this study's results.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Sungkyunkwan University IRB approval number: SKKU 2021-12-014. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

K-IY: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft, Writing – review & editing. T-SJ: Data curation, Methodology, Software, Writing – review & editing. S-CK: Conceptualization, Project administration, Supervision, Writing – review & editing. S-CL: Conceptualization, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Technology Innovation Program (Grant No. 20015188, Development of AI-Based Motion Analysis, Coaching and Visualization Technology During 1:N Non-face-to-face Sports Training) funded by the Ministry of Trade, Industry & Energy (MOTIE), Republic of Korea.

Acknowledgments

The authors would like to express our gratitude and appreciation to all experimental participants who participated in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet.* (2012) 380:247–57. doi: 10.1016/S0140-6736(12)60646-1

2. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet.* (2012) 380:219–29. doi: 10.1016/S0140-6736(12)61031-9

3. Ráthonyi G, Kósa K, Bács Z, Ráthonyi-Ódor K, Füzesi I, Lengyel P, et al. Changes in workers' physical activity and sedentary behavior during the COVID-19 pandemic. *Sustainability.* (2021) 13:9524. doi: 10.3390/su13179524

4. Neville RD, Lakes KD, Hopkins WG, Tarantino G, Draper CE, Beck R, et al. Global changes in child and adolescent physical activity during the COVID-19 pandemic: a systematic review and meta-analysis. *JAMA Pediatr.* (2022) 176:886–984. doi: 10.1001/jamapediatrics.2022.2313

5. Wunsch K, Kienberger K, Niessner C. Changes in physical activity patterns due to the COVID-19 pandemic: a systematic review and meta-analysis. *Int J Environ Res Public Health.* (2022) 19:2250. doi: 10.3390/ijerph19042250

6. Bu F, Bone JK, Mitchell JJ, Steptoe A, Fancourt D. Longitudinal changes in physical activity during and after the first national lockdown due to the COVID-19 pandemic in England. *Sci Rep.* (2021) 11:17723. doi: 10.1038/s41598-021-97065-1

7. Janssen X, Fleming L, Kirk A, Rollins L, Young D, Grealy M, et al. Changes in physical activity, sitting and sleep across the COVID-19 national lockdown period in Scotland. *Int. J. Environ Res Public Health.* (2020) 17:9362. doi: 10.3390/ijerph17249362

8. Madruga M, Prieto J, Rohlfs P, Gusi N, editors. Cost-effectiveness effects of a home-based exercise intervention for female caregivers of relatives with dementia: study protocol for a randomized controlled trial. *Healthcare.* (2020) 8:54. doi: 10.3390/healthcare8010054

9. Suikkanen SA, Soukkio PK, Aartolahti EM, Kautiainen H, Kääriä SM, Hupli MT, et al. Effects of home-based physical exercise on days at home and cost-effectiveness in pre-frail and frail persons: randomized controlled trial. *J Am Med Direct Assoc.* (2021) 22:773–9. doi: 10.1016/j.jamda.2020.06.005

10. Eckert T, Wronski P, Bongartz M, Ullrich P, Abel B, Kiss R, et al. Cost-effectiveness and cost-utility of a home-based exercise program in geriatric patients with cognitive impairment. *Gerontology.* (2021) 67:220–32. doi: 10.1159/000512748

11. Do Amaral VT, Viana AA, Heubel AD, Linares SN, Martinelli B, Witzler PHC, et al. Cardiovascular, respiratory, and functional effects of home-based exercise training after COVID-19 hospitalization. *Med Sci Sports Exerc.* (2022) 54:1795–1803. doi: 10.1101/2022.01.24.22269745

12. Longobardi I, Prado DMLd, Goessler KF, Oliveira Júnior GN, de Andrade DCO, Gualano B, et al. Benefits of home-based exercise training following critical SARS-CoV-2 infection: a case report. *Front Sports Act living.* (2022) 3:791703. doi: 10.3389/fspor.2021.791703

13. Kagan D, Alpert GF, Fire M. Zooming into video conferencing privacy. *IEEE Trans Comput Soc Syst.* (2023) 1–12. doi: 10.1109/TCSS.2022.3231987

14. Sandhu RK, Vasconcelos-Gomes J, Thomas MA, Oliveira T. Unfolding the popularity of video conferencing apps-A privacy calculus perspective. *Int J Inform Manage.* (2023) 68:102569. doi: 10.1016/j.ijinfomgt.2022.102569

15. Gui F, Tsai C-H, Vajda A, Carroll JM. Workout connections: investigating social interactions in online group exercise classes. *Int J Hum Comput Stud.* (2022) 166:102870. doi: 10.1016/j.ijhcs.2022.102870

16. Guo J, Fussell SR. "It's great to exercise together on zoom!": understanding the practices and challenges of live stream group fitness classes. In: *Proceedings of the ACM on Human-Computer Interaction.* (2022). p. 1–28. doi: 10.1145/3512918

17. Brunet J, Sabiston CM. Social physique anxiety and physical activity: a self-determination theory perspective. *Psychol Sport Exerc.* (2009) 10:329–35. doi: 10.1016/j.psychsport.2008.11.002

18. Sabiston CM, Pila E, Pinsonnault-Bilodeau G, Cox AE. Social physique anxiety experiences in physical activity: a comprehensive synthesis of research studies focused on measurement, theory, and predictors and outcomes. *Int Rev Sport Exerc Psychol.* (2014) 7:158–83. doi: 10.1080/1750984X.2014.904392

19. Gammage KL, Hall CR, Ginis KAM. Self-presentation in exercise contexts: differences between high and low frequency exercisers. *J Appl Soc Psychol.* (2004) 34:1638–51. doi: 10.1111/j.1559-1816.2004.tb02791.x

20. Gammage KL, Lamarche L, Drouin B. Self-presentational efficacy: does it moderate the relationship between social physique anxiety and physical activity in university students? *Int J Sport Exerc Psychol.* (2014) 12:357–67. doi: 10.1080/1612197X.2014.932824

21. Park I, Sah YJ, Lee S, Lee D. Avatar-mediated communication in video conferencing: effect of self-affirmation on debating participation focusing on moderation effect of avatar. International *J Hum Comput Interact*. (2022) 39:464–475. doi: 10.1080/10447318.2022.2041897

22. Li BJ, Lwin MO, Jung Y. Wii, myself, and size: the influence of proteus effect and stereotype threat on overweight children's exercise motivation and behavior in exergames. *Games Health Res Dev Clin Appl.* (2014) 3:40–8. doi: 10.1089/g4h.2013.0081

23. Peña J, Kim E. Increasing exergame physical activity through self and opponent avatar appearance. *Comput Hum Behav.* (2014) 41:262–7. doi: 10.1016/j.chb.2014.09.038

24. Peña J, Khan S, Alexopoulos C. I am what I see: How avatar and opponent agent body size affects physical activity among men playing exergames. *J Comput Mediat Commun.* (2016) 21:195–209. doi: 10.1111/jcc4.12151

25. Qian J, McDonough DJ, Gao Z. The effectiveness of virtual reality exercise on individual's physiological, psychological and rehabilitative outcomes: a systematic review. *Int J Environ Res Public Health.* (2020) 17:4133. doi: 10.3390/ijerph17114133

26. Park E-C, Kim S-G, Lee C-W. The effects of virtual reality game exercise on balance and gait of the elderly. *J Phys Therapy Sci.* (2015) 27:1157–9. doi: 10.1589/jpts.27.1157

27. Arlati S, Colombo V, Spoladore D, Greci L, Pedroli E, Serino S, et al. A social virtual reality-based application for the physical and cognitive training of the elderly at home. *Sensors.* (2019) 19:261. doi: 10.3390/s19020261

28. da Silva Vieira AS, de Melo MCDA, Noites SPARS, Machado JP, Gabriel MMJ. The effect of virtual reality on a home-based cardiac rehabilitation program on body composition, lipid profile and eating patterns: a randomized controlled trial. *Eur J Integrat Med.* (2017) 9:69–78. doi: 10.1016/j.eujim.2016.11.008

29. Menhas R, Qin L, Saqib ZA, Younas M. The association between COVID-19 preventive strategies, virtual reality exercise, use of fitness apps, physical, and psychological health: testing a structural equation moderation model. *Front Public Health.* (2023) 11:1170645. doi: 10.3389/fpubh.2023.1170645

30. Liu R, Menhas R, Dai J, Saqib ZA, Peng X. Fitness apps, live streaming workout classes, and virtual reality fitness for physical activity during the COVID-19 lockdown: an empirical study. *Front Public Health.* (2022) 10:852311. doi: 10.3389/fpubh.2022.852311

31. Ali SG, Wang X, Li P, Jung Y, Bi L, Kim J, et al. A systematic review: Virtual-reality-based techniques for human exercises and health improvement. *Front Public Health.* (2023) 11:1143947. doi: 10.3389/fpubh.2023.1143947

32. Gao Z, Lee JE, McDonough DJ, Albers C. Virtual reality exercise as a coping strategy for health and wellness promotion in older adults during the COVID-19 pandemic. *J Clin Med.* (2020) 9:1986. doi: 10.3390/jcm9061986

33. Groenveld T, Achttien R, Smits M, de Vries M, van Heerde R, Staal B, et al. Feasibility of virtual reality exercises at home for post-COVID-19 condition: cohort study. *JMIR Rehabil Assist Technol.* (2022) 9:e36836. doi: 10.2196/36836

34. Betka S, Kannape OA, Fasola J, Lance F, Cardin S, Schmit A, et al. Virtual reality intervention alleviates dyspnoea in patients recovering from COVID-19 pneumonia. *ERJ Open Res*. (2023) 9:570–2022. doi: 10.1183/23120541.00570-2022

35. Langer A, Gassner L, Hasenauer S, Gruber J, Wizany L, Pokan R, et al. How COVID-19 will boost remote exercise-based treatment in Parkinson's disease: a narrative review. *NPJ Parkinsons Dis.* (2021) 7:25. doi: 10.1038/s41531-021-00160-3

36. Rutkowski S, Kiper P, Cacciante L, Mazurek J, Turolla A. Use of virtual reality-based training in different fields of rehabilitation: a systematic review and meta-analysis. *J Rehabil Med.* (2020) 52:1–16. doi: 10.2340/16501977-2755

37. Ng Y-L, Ma F, Ho FK, Ip P, Fu K-w. Effectiveness of virtual and augmented reality-enhanced exercise on physical activity, psychological outcomes, and physical performance: a systematic review and meta-analysis of randomized controlled trials. *Comput Hum Behav*. (2019). 99:278–91. doi: 10.1016/j.chb.2019.05.026

38. Fele B, Lampe A, Peer P, Struc V, editors. C-VTON: context-driven image-based virtual try-on network. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, HI (2022). doi: 10.1109/WACV51458.2022.00226

39. Fenocchi E, Morelli D, Cornia M, Baraldi L, Cesari F, Cucchiara R, editors. Dual-branch collaborative transformer for virtual try-on. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2022). doi: 10.1109/CVPRW56347.2022.00246

40. He S, Song Y-Z, Xiang T, editors. Style-based global appearance flow for virtual try-on. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA (2022). doi: 10.1109/CVPR52688.2022.00346

41. Morelli D, Fincato M, Cornia M, Landi F, Cesari F, Cucchiara R, editors. Dress code: high-resolution multi-category virtual try-on. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA (2022). doi: 10.1109/CVPRW56347.2022.00243

42. Zhang T, Wang WYC, Cao L, Wang Y. The role of virtual try-on technology in online purchase decision from consumers' aspect. *Internet Res.* (2019) 29:529–51. doi: 10.1108/IntR-12-2017-0540

43. Riar M, Xi N, Korbel JJ, Zarnekow R, Hamari J. Using augmented reality for shopping: a framework for AR induced consumer behavior, literature review and future agenda. *Internet Res.* (2022). 33:242–279. doi: 10.1108/INTR-08-2021-0611

44. Sun K, Xiao B, Liu D, Wang J, editors. Deep high-resolution representation learning for human pose estimation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA (2019). doi: 10.1109/CVPR.2019.00584

45. Waltemate T, Gall D, Roth D, Botsch M, Latoschik ME. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans Visual Comput Graph.* (2018) 24:1643–52. doi: 10.1109/TVCG.2018.2794629

46. Santesteban I, Otaduy MA, Casas D, editors. Snug: self-supervised neural dynamic garments. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA (2022). doi: 10.1109/CVPR52688.2022.00797

47. Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ. SMPL: a skinned multi-person linear model. *ACM Trans Graph.* (2015) 34:1–16. doi: 10.1145/2816795.2818013

48. Güler RA, Neverova N, Kokkinos I, editors. Densepose: dense human pose estimation in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT (2018). doi: 10.1109/CVPR.2018.00762

49. Ge Y, Song Y, Zhang R, Ge C, Liu W, Luo P, editors. Parser-free virtual try-on via distilling appearance flows. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, TN (2021). doi: 10.1109/CVPR46437.2021.00838

frontiers | Frontiers in Public Health

TYPE Perspective
PUBLISHED 17 January 2024
DOI 10.3389/fpubh.2023.1337151

Check for updates

OPEN ACCESS

EDITED BY
Saman Ghaffarian,
University College London, United Kingdom

REVIEWED BY
Victoria Ramos Gonzalez,
Carlos III Health Institute (ISCIII), Spain

*CORRESPONDENCE
Jason Tucker
✉ Jason.tucker@mau.se

RECEIVED 12 November 2023
ACCEPTED 26 December 2023
PUBLISHED 17 January 2024

CITATION
Tucker J and Lorig F (2024) Agent-based social simulations for health crises response: utilising the everyday digital health perspective.
Front. Public Health 11:1337151.
doi: 10.3389/fpubh.2023.1337151

Agent-based social simulations for health crises response: utilising the everyday digital health perspective

Jason Tucker[1]* and Fabian Lorig[2,3]

[1]Department of Global Political Studies, Faculty of Culture and Society, Malmö University, Malmö, Sweden, [2]Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden, [3]Internet of Things and People Research Center, Malmö University, Malmö, Sweden

There is increasing recognition of the role that artificial intelligence (AI) systems can play in managing health crises. One such approach, which allows for analysing the potential consequences of different policy interventions is agent-based social simulations (ABSS). Here, the actions and interactions of autonomous agents are modelled to generate virtual societies that can serve as a "testbed" for investigating and comparing different interventions and scenarios. This piece focuses on two key challenges of ABSS in collaborative policy interventions during the COVID-19 pandemic. These were defining valuable scenarios to simulate and the availability of appropriate data. This paper posits that drawing on the research on the "everyday" digital health perspective in designing ABSS before or during health crises, can overcome aspects of these challenges. The focus on digital health interventions reflects a rapid shift in the adoption of such technologies during and after the COVID-19 pandemic, and the new challenges this poses for policy makers. It is argued that by accounting for the everyday digital health in modelling, ABSS would be a more powerful tool in future health crisis management.

KEYWORDS

agent-based social simulations, health policy, crisis, COVID-19, everyday digital health, artificial intelligence

1 Introduction

During the COVID-19 pandemic, there was a momentous increase in interest in how artificial intelligence (AI) systems could be used to manage the crisis (1, 2). Policy makers faced a huge challenge of having to try to analyse the potential consequences of different policy interventions, doing do with very limited time, recourses and tools to assess the impact or potential outcomes. Compounding these challenges, was the issue that, while pandemics are not new, the COVID-19 pandemic posed unique challenges for policy makers (3). As such, there was a limit to the usefulness of drawing on past lessons on pandemic management. At the same time the pandemic saw a surge in the adoption of digital health technology to overcome the pressure on health services and lack of mobility many people faced, with this adoption having varying levels of success (4–6). The use of digital health such as websites, social media, telemedicine, patient self-care devices, health apps, wearable tracking devices,

 10.3389/fpubh.2023.1337151

persuasive computing and diagnostics, to name but a few, fundamentally altered health and healthcare for many people, doing so in ways that we still do not understand.[1]

Within this context, agent-based social simulations (ABSS) proved to be very well suited to informing policy decision making (7, 8). ABSS aim to model and simulate the actions and interactions of intelligent agents[2], creating virtual populations or social systems composed of autonomous (artificial) individuals (11). These virtual societies can serve as a "testbed" for investigating and comparing different policy interventions and scenarios prior to their implementation. While ABSS cannot predict the future, it is a powerful tool to help inform policy makers of potential outcomes or consequences of interventions. ABSS enables decision makers to "play" with policy, and variations of policies, and investigate their impact under different circumstances and scenarios (12). Doing so in a virtual population allows for the conducting of experiments in a time and cost-efficient manner, removes the risk of harming real-world individuals and can facilitate greater levels of preparedness and response to health crises.

However, while ABSS is not a new technology its application in the policy making space is still in its early days, with its implementation during COVID-19 not being without a certain amount of friction. This piece focuses on two prominent points of friction that were identified, these being defining valuable scenarios to simulate and a lack of appropriate data (13). This paper explores how we can reduce these frictions if an "everyday" digital health perspective is adopted in the development of ABSS for health crisis management. It is claimed that by recognising the adoption and normalisation of digital health technologies (and crucially how and why this is experienced differently by various groups) can strengthen ABSS, positioning it as a powerful tool in manging health crises.

2 ABSS and COVID-19

> *"Simulation is the imitation of the operation of a real-world process or system over time"* (14).

Computer simulations provide a virtual environment for conducting experiments with a target system and can be used to analyse the behaviour or dynamics of that system under different circumstances (15). Similar to real-world experiments, this allows us to conduct *what-if* analyses and to explore different scenarios to investigate how changes to the system's configuration or exogenous factors might affect the system's overall behaviour. A major advantage of using a virtual environment is that the experiments do not risk jeopardising the system under investigation. Simulation experiments allow for more time and cost-efficient analyses and enable the generation and investigation of situations or conditions that might rarely occur, e.g., crisis or disaster situations. This type of experiment is also referred to as *in silico experiment* and is increasingly used in, for instance, biology, social sciences, and engineering (16, 17).

There exists a variety of applications where computer simulations are used in healthcare (18, 19). Particularly during the COVID-19 pandemic, a great number of simulation models were been developed to investigate infection dynamics and the effects of non-pharmaceutical interventions (12, 13, 20). Traditionally, epidemiological models make use of differential equations and transition probabilities to describe the dynamics of an entire population of individuals, neglecting individual infection statuses and infection chains. For simulating the consequences of specific non-pharmaceutical interventions or other policy measures, however, individual behaviour needs to be taken into account as the success of these intervention greatly depends on how and to what extent each individual adapts to and comply with instructions and restrictions.

A simulation paradigm that is particularly well suited to simulate individual behaviour is ABSS. ABSS make use of an artificial population of autonomous individuals, so called *Agents*, each of which is characterised by a set of attributes, e.g., age, gender, and health status. Based on these personal attributes, its environment, and its individual needs and goals, each agent individually plans its actions by imitating human-like behaviour using AI (10). The goal of ABSS is to imitate the relevant aspects of the real-world population, i.e., composition and behaviour, as closely as possible to allow for drawing sound conclusions regarding the target system.

Traditional simulation approaches in healthcare and epidemiology aim to directly model the dynamics of the phenomenon of interest. In ABSS, however, the system's behaviour and the corresponding macro-scale phenomena emerge from micro-scale agent behaviour. This allows not only for analysing *what* potential consequences a given scenario or intervention might result in but also provides a better understanding *why* certain effects can be observed. Gilbert and Troitzsch (21, p. 1) argue that individual-based simulations imply a "new way of thinking about social and economic processes," due to the emergence of complex behaviour from simple actions and interactions. ABSS can be applied to review theories, to verify assumptions, and to generate data and can therefore, according to the authors, be considered a new method of theory development.

In the early phase of the COVID-19 pandemic, policy makers were supposed to make critical decisions facing rapid developments and incomplete data. Identifying appropriate policy interventions under such a high degree of uncertainty is challenging and sound decisions require considering different sources of information and evidence. To address these challenges and to assess uncertainty, ABSS of the COVID-19 pandemic were developed and applied (22). By modelling an artificial population and presumed transmission processes, the possible effects of different interventions (e.g., lockdowns, social distancing, or facemasks) under different possible scenarios (e.g., in different countries or at different stages of the pandemic) could be simulated to inform policy makers of their potential outcomes. The use of ABSS to inform policy making has gained popularity, and other recent application areas include disaster management (23), healthcare (24), agriculture (25) and transportation (26).

1 We use digital health as a broad and inclusive category where ICT and digital technology are involved in a range of health(care) arenas.

2 Intelligent agents can be defined as autonomous entities that are characterised by individual attributes and that autonomously make decisions and perform actions based on their environment, individual situations, and attributes (9). In the context of ABSS, the goal of the agents is to imitate realistic human behaviour in order to create an artificial population that can be used for investigating complex phenomena within social systems (10).

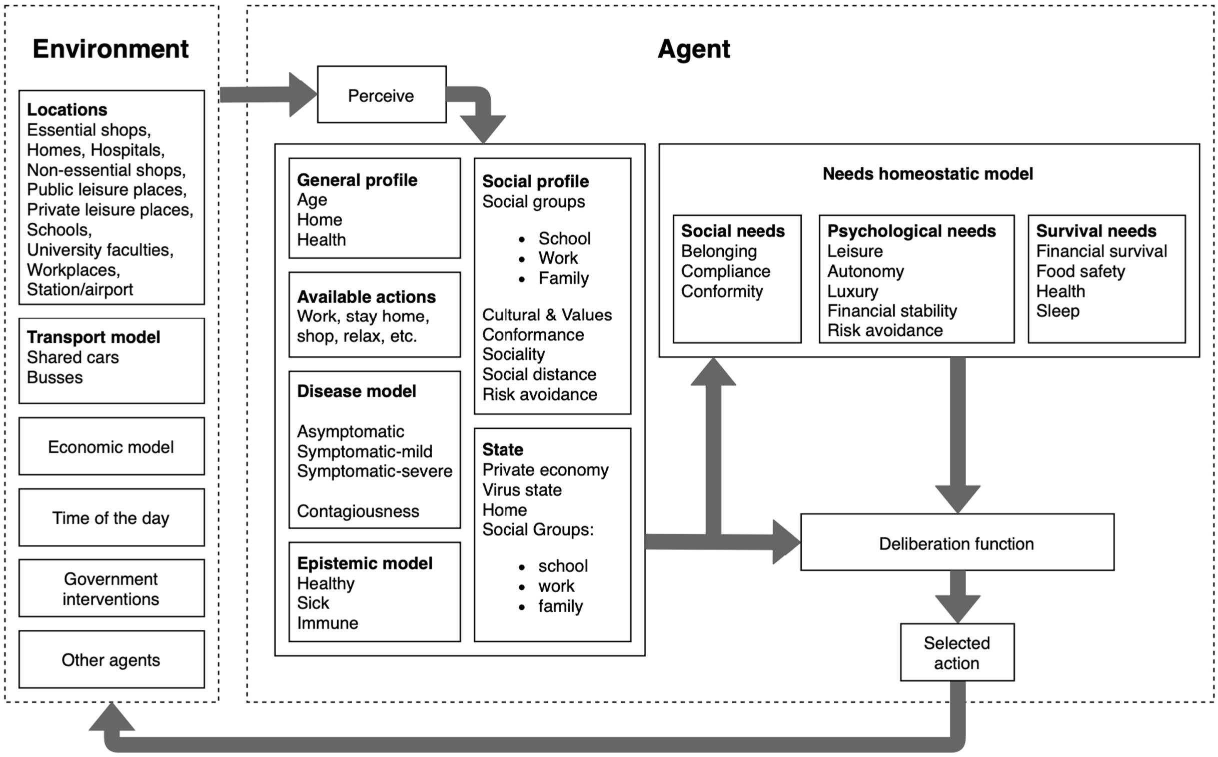


FIGURE 1
Architecture of the ASSOCC simulation model (29).

The ASSOCC model (27) is one example of an ABSS that was developed in the early phase of the COVID-19 pandemic for analysing the effects of different non-pharmaceutical interventions and to inform policy making. The agents' behaviour is based on individual needs, which are inspired by Maslow's hierarchy of needs, and the satisfaction of these needs through specific actions promotes different values, according to Schwarz theory of basic human values (28). In the ASSOCC model, the satisfaction of needs is implemented using a water tank approach, where each need is represented by a water tank depleting over time and the agent will try to keep their tanks filled up by pursuing appropriate actions. The model further consists of a representation of different locations (e.g., homes, hospitals, schools, and workplaces), an economic model, a transport model, and time progress is modelled by dividing the day into four phases (i.e., morning, afternoon, evening, and night). The architecture of the ASSOCC simulation model is shown in Figure 1.

One of the interventions that can be simulated using the ASSOCC model are tracking and tracing apps (30). The implementation of these apps was been discussed by different countries at an early stage of the pandemic (31). By tracking contacts with other persons using mobile phones, infected individuals can, once they are diagnosed with COVID-19, inform those they have met during the last days such that they can isolate themselves. The simulation results indicated that higher numbers of app users can indeed decrease the peak of infections, however, the effects are marginal even in moderate app usage scenarios, which is due to contacts with other persons shifting from public to private places (see Figure 2).

3 The "everyday" digital health perspective

This everyday perspective on technology[3] highlights that technology does not exist independently of humans. It warns against seeing humans as without agency and society as merely a site upon which technology is dropped from above (33). The everyday pushes back against the tendency to reduce people to data points (itself very problematic), or ignoring individuals' agency, approaches which are prone to inaccuracy (*ibid*). In adopting an everyday perspective on digital health, we are reminded to focus on "weaker" actors who may not normally be considered as being important players in transforming global contexts through their local actions (34). This everyday approach calls for the recognition that a broad range of actors are considered as interacting with digital health, not just the technology developers, but the patients (consumers,) carers, practitioners etc. (*ibid*). The everyday perspective centralises and embraces the reality that digital health exists within complex, fluid, and co-constitutive everyday spaces where people and technology interact.

3 The everyday perspective has its foundations in sociology but has had far reaching applications across a range of disciplines. By focusing on the everyday, scholars "not only give importance to the ordinary, and take the ordinary seriously as a category of analysis, but they also evidence how everyday life, social relations, experiences and practices are always more than simply or straightforwardly mundane, ordinary and routine" [(32): 811].

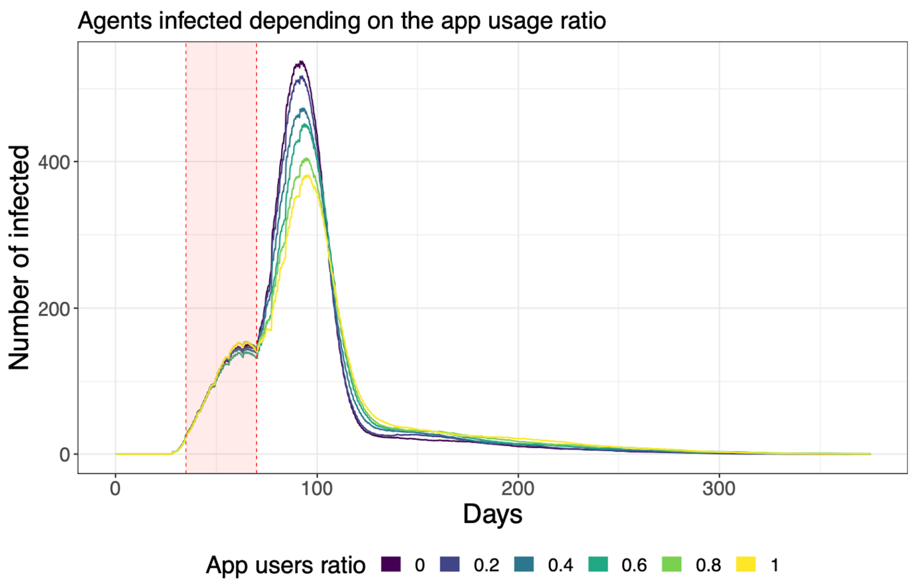


FIGURE 2
Number of infected agents for different levels of app usage in the simulation (30).

Foregrounding the everyday digital health perspective in policy development is crucial if we are to be able to best develop policy for the implementation of digital health applications and improve health outcomes. However, in crises contexts, with severe limitations on time and resources, doing so is a challenge. The richness of the everyday perspective of digital health, with its focus on the highly individual, fluid, deeply complex and context dependent co-constitutive relations means there is a vast but disjointed body of research. The need to connect this data is essential: fragmented data was one of the main limitations with policy making during COVID-19 (35). ABSS with its focus on individual agent autonomy is well suited to using the fragmented micro (everyday) level to inform the macro level. It is argued here that doing so would not only elevate the existing research on the everyday, but would address two significant challenges of using ABSS in future health crises.

4 ABSS and the everyday digital health perspective

Even though a great number of ABSS models were developed during the COVID-19 pandemic, policy actors still seemed hesitant to use them to inform their decision-making processes (36). There was also a mismatch between the developed models and the information decision makers actually needed for facilitating and supporting their work (13). Developing appropriate and useful models is a joint effort and requires a close collaboration between modellers, policy actors, and other stakeholders. Based on a survey of collaborations between ABSS modellers and policy makers, Belfrage et al. (36) identify five challenges that commonly occur including disagreement regarding what scenarios to model, unrealistic expectations in terms of the contribution of simulations, lack of stakeholder engagement, lack of technical understanding, and general scepticism regarding the generated results. We focus on two key challenges faced when implementing ABSS during the COVID-19 pandemic, defining valuable scenarios to simulate and the lack of appropriate data, and how the everyday digital health perspective can, in part, address these.

4.1 Defining valuable scenarios

In ABSS, a scenario is a specific configuration of the model that can be used to investigate a particular situation or phenomenon. It consists of the input data that is used to initialise the model, the parameter values to adapt the model to the specific circumstances, the synthetic population, and different assumptions regarding the underlying mechanisms. For instance, when analysing a COVID-19 intervention, such as lockdowns, a scenario must be developed that specifies the length and extent of the lockdown but also the composition of the population and the time horizon of the simulation. In many ABSS of the COVID-19 crisis, lockdowns were modelled as the entire population not leaving their homes, resulting in a rapid decline in infections. In most countries, however, lockdowns were associated with certain exceptions and easings, e.g., they were only applicable after certain hours, grocery shopping was permitted, and essential workers could go to their jobs. This resulted in a discrepancy between the scenarios that would be valuable for policy makers and what was implemented by the modelers. This discrepancy could be attributed to a lack of collaboration during model development. Modellers have limited domain expertise and, thus, rely on policy actors and stakeholders to actively engage in participatory modelling, where they contribute with their expertise to define valuable scenarios and models (37).

These discrepancies are further complicated when there is a reliance on emerging digital health technology as a policy response to a health crisis. Here it is crucial not only to bridge the gap between the policy makers and the modellers, but to actively include domain experts in the everyday perspective of digital health technology for a range of technologies and groups within society. In so doing, the knowledge of the everyday experiences of past and current digital health technologies can support the scenario definition. By foregrounding the everyday in developing digital health interventions, the scenario development is enriched by the incorporation of the fluidity of the use, misuse, rejection or adaption of digital health technologies in the everyday. Centralising the everyday, and collaborating with a broad range of domain experts not only grounds scenarios in "real life," but also serves to challenge the narrowing of scenarios of digital health in society by elite actors [see (38, 39)].

4.2 More appropriate data for model development

To be able to draw sound conclusions about the target system based on simulation results, it is necessary that the model correctly and realistically reflects the relevant parts of the target system. In simulations, verification and validation are applied to ensure the appropriateness and correctness of the model for a certain purpose (40). While verification ensures the correctness of the model's implementation, validation assesses the applicability of the model and ensures it is an accurate representation of the target system. Both for the development but also for the validation of a simulation model, suitable data on the system and phenomenon that shall be studied is required, a lack of data might result in unrealistic models and results (41).

In ABSS, this need for appropriate data mainly concerns population data from the original population that is required for the generation of a synthetic population. This includes data on socio-demographic attributes of the population (e.g., age, gender, and household composition) but also data on the needs, desires, and goals of these individuals, which is required for appropriate modelling human decision making under different circumstances. Two common issues that modellers face include that data on specific attributes might be difficult to acquire and, if existing and accessible, this data often needs to be expanded and harmonized (42).

This demand of ABBS modellers can be seen as sitting between the social determinants of health and the everyday in digital health interventions in health crises. As Lupton (43) notes, numerous factors, such as income, education, location, age, disability, etc. impact the social structuring of digital health use. While these social determinants of health are vital to consider when developing and implementing public health crisis policy, everyday perspective of digital health can provide a wealth of data upon which to build models. Everyday digital health is largely a qualitative field, but there are examples of surveys and mixed methods approaches upon which to draw. This is facilitated by the self-collection and monitoring of data by "digitally engaged patients" [(44), p.256]. In addition, Stanley (45) describes data on the everyday as almost "naturally occurring" if we are to broaden our gaze to include already existing data sources. Given the time and resource restraints present during health crises, the ability to draw on already existing data is a significant advantage. One key aspect of this, which is related to the above opportunity, would be engagement with domain experts to help navigate this breadth of data, thus, mitigating the challenges which Chapuis and Taillandier (42) describe.

5 Summary

The rapid adoption of digital health technologies during and after the COVID-19 pandemic posed new challenges for policy makers. This paper argues that by accounting for the everyday digital health in modelling, ABSS can better support policy makers, and be a more powerful tool in future health crisis management. This claim in grounded in how the everyday digital health perspective can address two challenges of using ABSS identified during the COVID-19 pandemic, defining valuable scenarios and lack of appropriate data.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JT: Writing – original draft. FL: Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partly supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation. Funding to make this open access was provided by Malmö University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Chen J, Li K, Zhang Z, Li K, Yu PS. A survey on applications of artificial intelligence in fighting against COVID-19. *ACM Comput. Surv.* (2021) 54:1–32. doi: 10.1145/3465398

2. Rasheed J, Jamil A, Hameed AA, Aftab U, Aftab J, Shah SA, et al. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos, Solitons Fractals*. (2020) 141:110337. doi: 10.1016/j.chaos.2020.110337

3. Patterson GE, McIntyre KM, Clough HE, Rushton J. Societal impacts of pandemics: comparing COVID-19 with history to focus our response. *Front Public Health*. (2021) 9:630449. doi: 10.3389/fpubh.2021.630449

4. Dron L, Kalatharan V, Gupta A, Haggstrom J, Zariffa N, Morris AD, et al. Data capture and sharing in the COVID-19 pandemic: a cause for concern. *Lancet Digit Health*. (2022) 4:e748–56. doi: 10.1016/S2589-7500(22)00147-9

5. Iyamu I, Gómez-Ramírez O, Xu AX, Chang HJ, Watt S, Mckee G, et al. Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review. *Digit. Health*. (2022) 8:205520762211022. doi: 10.1177/20552076221102255

6. Scott BK, Miller GT, Fonda SJ, Yeaw RE, Gaudaen JC, Pavliscsak HH, et al. Advanced digital health technologies for COVID-19 and future emergencies. *Telemed e-Health*. (2020) 26:1226–33. doi: 10.1089/tmj.2020.0140

7. Gilbert N, Ahrweiler P, Barbrook-Johnson P, Narasimhan KP, Wilkinson H. Computational modelling of public policy: reflections on practice. *J Artif Soc Soc Simul*. (2018) 21:14. doi: 10.18564/jasss.3669

8. Hoad K, Watts C. Are we there yet? Simulation modellers on what needs to be done to involve agent-based simulation in practical decision making. *J. Simul.* (2012) 6:67–70. doi: 10.1057/jos.2011.19

9. Wooldridge M, Jennings NR. Intelligent agents: theory and practice. *Knowl Eng Rev*. (1995) 10:115–52. doi: 10.1017/S0269888900008122

10. Gilbert N. Agent-based social simulation: dealing with complexity. *Complex Syst. Net. Excell.* (2004) 9:1–14.

11. Davidsson P. Agent based social simulation: a computer science view. *J Artif Soc Soc Simul*. (2002) 5.

12. Lorig F, Vanhée L, Dignum F. Agent-based social simulation for policy making. In: Chetouani M, Dignum V, Lukowicz P, Sierra C, editors. Human-centered artificial intelligence. ACAI 2021. Lecture notes in computer science. Cham: Springer (2023) 13500:391–414.

13. Lorig F, Johansson E, Davidsson P. Agent-based social simulation of the COVID-19 pandemic: a systematic review. *J Artif Soc Soc Simul*. (2021) 24. doi: 10.18564/jasss.4601

14. Banks J. Principles of simulation. In: Banks J, editor. *Handbook of Simulation*. New York: John Wiley & Sons (1998). 3–30.

15. Winsberg E. Simulated experiments: methodology for a virtual world. *Philos Sci*. (2003) 70:105–25. doi: 10.1086/367872

16. Conte R, Paolucci M. On agent-based modeling and computational social science. *Front Psychol*. (2014) 5:668. doi: 10.3389/fpsyg.2014.00668

17. Kollmann M, Sourjik V. In silico biology: from simulation to understanding. *Curr Biol*. (2007) 17:R132–4. doi: 10.1016/j.cub.2006.12.034

18. Katsaliaki K, Mustafee N. Applications of simulation within the healthcare context. *J Oper Res Soc*. (2011) 62:1431–51. doi: 10.1057/jors.2010.20

19. Mielczarek B, Uziałko-Mydlikowska J. Application of computer simulation modeling in the health care sector: a survey. *SIMULATION*. (2012) 88:197–216. doi: 10.1177/0037549710387802

20. Tang J, Vinayavekhin S, Weeramongkolkul M, Suksanon C, Pattarapremcharoen K, Thiwathittayanuphap S, et al. Agent-based simulation and modeling of COVID-19 pandemic: a bibliometric analysis. *J Disaster Res*. (2022) 17:93–102. doi: 10.20965/jdr.2022.p0093

21. Gilbert N, Troitzsch K. *Simulation for the social scientist*. McGraw-Hill Education (2015).

22. Dignum F, Dignum V, Davidsson P, Ghorbani A, van der Hurk M, Jensen M, et al. Analysing the combined health, social and economic impacts of the coronavirus pandemic using agent-based social simulation. *Mind Mach*. (2020) 30:177–94. doi: 10.1007/s11023-020-09527-6

23. Fikar C, Hirsch P, Nolz PC. Agent-based simulation optimization for dynamic disaster relief distribution. *CEJOR*. (2018) 26:423–42. doi: 10.1007/s10100-017-0518-3

24. Comis M, Cleophas C, Büsing C. Patients, primary care, and policy: agent-based simulation modeling for health care decision support. *Health Care Manag Sci*. (2021) 24:799–826. doi: 10.1007/s10729-021-09556-2

25. Delmotte S, Barbier JM, Mouret JC, Le Page C, Wery J, Chauvelon P, et al. Participatory integrated assessment of scenarios for organic farming at different scales in Camargue, France. *Agric Syst*. (2016) 143:147–58. doi: 10.1016/j.agsy.2015.12.009

26. Holmgren J, Davidsson P, Persson JA, Ramstedt L. TAPAS: a multi-agent-based model for simulation of transport chains. *Simul Model Pract Theory*. (2012) 23:1–18. doi: 10.1016/j.simpat.2011.12.011

27. Dignum F. (ed.). *Social simulation for a crisis: Results and lessons from simulating the COVID-19 crisis computational social sciences* Springer International Publishing (2021).

28. Dignum F. Foundations of social simulations for crisis situations. In: *Social simulation for a crisis: Results and lessons from simulating the COVID-19 crisis*. Cham: Springer International Publishing (2021). 15–37.

29. Jensen M, Vanhée L, Kammler C. Social simulations for crises: from theories to implementation. In: *Social simulation for a crisis: Results and lessons from simulating the COVID-19 crisis*. Cham: Springer International Publishing (2021). 39–84.

30. Jensen M, Lorig F, Vanhée L, Dignum F. Deployment and effects of an app for tracking and tracing contacts during the COVID-19 crisis. In: Dignum F, editor. *Social simulation for a crisis: Results and lessons from simulating the COVID-19 crisis*. Cham: Springer International Publishing (2021). 167–88.

31. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the mobile application rating scale (MARS). *J Med Syst*. (2020) 44:1–15. doi: 10.1007/s10916-020-01633-3

32. Neal S, Murji K. Sociologies of everyday life: editors' introduction to the special issue. *Sociology*. (2015) 49:811–9. doi: 10.1177/0038038515602160

33. Pink S, Ruckenstein M, Berg M, Lupton D. Everyday automation setting a research agenda In: S Pink, M Berg, D Lupton and M Ruckenstein, editors. *Everyday automation: Experiencing and anticipating emerging technologies*. New York: Routledge (2022). 1–19.

34. Strange M, Tucker J. AI and the everyday political economy of Global Health In: S Lindgreen, editor. *The handbook of critical AI studies*. Cheltenham: Edward Elgar (2023). 367–77.

35. Wang Q, Su M, Zhang M, Li R. Integrating digital technologies and public health to fight Covid-19 pandemic: key technologies, applications, challenges and outlook of digital healthcare. *Int J Environ Res Public Health*. (2021) 18:6053. doi: 10.3390/ijerph18116053

36. Belfrage M., Lorig F., Davidsson P. (2022). *Making sense of collaborative challenges in agent-based modelling for policy-making*.

37. Ahrweiler P, Frank D, Gilbert N. Co-designing social simulation models for policy advise: lessons learned from the INFSO-SKIN study. In: D'Ambrogio A, Barros F, Hu X, editors. *2019 spring simulation conference (SpringSim)*: IEEE (2019). 1–12.

38. Hoff JL. Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services. *Futures*. (2023) 148:103131. doi: 10.1016/j.futures.2023.103131

39. Tucker J. The future vision (s) of AI health in the Nordics: comparing the national AI strategies. *Futures*. (2023) 149:103154. doi: 10.1016/j.futures.2023.103154

40. Balci . (2003). Verification, validation, and certification of modeling and simulation applications. In Proceedings of the 2003 winter simulation conference, 2003. (Vol. 1, pp. 150–158). IEEE.

41. Kleijnen J. P. (1999). Validation of models: statistical techniques and data availability. In Proceedings of the 31st conference on winter simulation: Simulation---a bridge to the future-volume 1 (pp. 647–654).

42. Chapuis K., Taillandier P. (2019). A brief review of synthetic population generation practices in agent-based social simulation. In submitted to SSC2019, Social Simulation Conference.

43. Lupton D. *Digital health: Critical and cross-disciplinary perspectives* Routledge (2017).

44. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. *Soc Theory Health*. (2013) 11:256–70. doi: 10.1057/sth.2013.10

45. Stanley L ed. *Documents of life revisited: Narrative and biographical methodology for a 21st century humanism*. Farnham: Ashgate (2013).

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 01 February 2024
DOI 10.3389/fpubh.2024.1330801

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Sunyang Fu,
University of Texas Health Science Center at Houston, United States
Lei Zhu,
Agency for Science, Technology and Research (A*STAR), Singapore

*CORRESPONDENCE
Dragos Paul Mihai
✉ dragos_mihai@umfcd.ro

RECEIVED 31 October 2023
ACCEPTED 15 January 2024
PUBLISHED 01 February 2024

CITATION
Valeanu A, Mihai DP, Andrei C, Puscasu C, Ionica AM, Hinoveanu MI, Predoi VP, Bulancea E, Chirita C, Negres S and Marineci CD (2024) Identification, analysis and prediction of valid and false information related to vaccines from Romanian tweets. *Front. Public Health* 12:1330801. 10.3389/fpubh.2024.1330801

Identification, analysis and prediction of valid and false information related to vaccines from Romanian tweets

Andrei Valeanu, Dragos Paul Mihai*, Corina Andrei, Ciprian Puscasu, Alexandra Mihaela Ionica, Miruna Ioana Hinoveanu, Valentina Patricia Predoi, Ema Bulancea, Cornel Chirita, Simona Negres and Cristian Daniel Marineci

Faculty of Pharmacy, "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania

Introduction: The online misinformation might undermine the vaccination efforts. Therefore, given the fact that no study specifically analyzed online vaccine related content written in Romanian, the main objective of the study was to detect and evaluate tweets related to vaccines and written in Romanian language.

Methods: 1,400 Romanian vaccine related tweets were manually classified in true, neutral and fake information and analyzed based on wordcloud representations, a correlation analysis between the three classes and specific tweet characteristics and the validation of several predictive machine learning algorithms.

Results and discussion: The tweets annotated as misinformation showed specific word patterns and were liked and reshared more often as compared to the true and neutral ones. The validation of the machine learning algorithms yielded enhanced results in terms of Area Under the Receiver Operating Characteristic Curve Score (0.744–0.843) when evaluating the Support Vector Classifier. The predictive model estimates in a well calibrated manner the probability that a specific Twitter post is true, neutral or fake. The current study offers important insights regarding vaccine related online content written in an Eastern European language. Future studies must aim at building an online platform for rapid identification of vaccine misinformation and raising awareness for the general population.

KEYWORDS

vaccines, public health, misinformation, wordcloud, machine learning, support vectors

1 Introduction

Vaccines are among the most important medications worldwide. It is estimated that they have saved millions of lives and that they will continue to do so (1). Vaccines had a crucial role in the eradication of smallpox in 1980 and in bringing poliomyelitis very close to eradication (2, 3). In addition, a report found that as of November 2021, the Covid-19 vaccines saved nearly half a million lives in less than a year in the over 60 years old group across the WHO European Region (4).

However, despite their essential therapeutic effect and good safety profile, various disinformation articles, news and social media posts have emerged in the last decades, leading to the anti-vaccine movement. Even though the facts behind such information were proven to be false, the vaccine fake news phenomenon has led in many countries to a reduction of the vaccination rates, both in the adult and the pediatric population (5). Low vaccination rates pose the risk of diseases that currently have a low impact in the population to return with a higher impact, with an additional burden on the healthcare system (1).

Numerous fake news related to the Covid-19 vaccines have also emerged and spread during the pandemic (5, 6). The online disinformation, in combination with other social and economic factors (such as media usage, educational background, health literacy, public trust in the government and health system) have been hypothesized to influence a person's decision of getting the Covid-19 vaccine (7–9).

Social media platforms (such as Facebook and Twitter) are among the most important tools for spreading information about vaccines, whether it is valid information or fake news (5). Therefore, the analysis of the content distributed through such platforms might be of utmost importance in order to inform the general population and the health policy makers.

With regards to Twitter content, several studies have evaluated vaccine related posts (whether Covid or non-Covid), with regards to identifying and predicting disinformation, analyzing vaccine hesitancy, performing sentiment classification or other relevant analyses (10–19). The majority of the studies were based on tweets written in English. Other analyzed languages were Dutch, Moroccan and Turkish, while one study involved a multi-language approach for detection and classification of tweets related to Covid-19 (12, 14, 15, 18). However, to our knowledge, no such study specifically analyzed vaccine related content based on Romanian tweets.

Therefore, the objective of the present study was to analyze vaccine related content, with the main goal of developing specific machine learning models for predicting disinformation from tweets written in Romanian.

2 Materials and methods

2.1 Data collection

The vaccine related tweets were automatically extracted by using snscrape package developed in Python programming language (20, 21). The Twitter API was queried by using all the Romanian forms of the noun "vaccine," the verb "to vaccinate," as well as other ironical related terms (such as "vax," "vaxin" or "vaxxin") (20). All tweets (both original posts and replies, both Covid and non-Covid vaccine information) from 4 relevant 4-week periods during the Covid-19 pandemic were initially collected (First period: March 16, 2020 – April 12, 2020; second period: December 27, 2020 – January 23, 2021; third period: May 3, 2021 – May 30, 2021; fourth period: October 18, 2021 – November 14, 2021). Each period was considered suggestive for the aim of extracting relevant batches of tweets. March 16, 2020 was the date in which the Emergency State was declared in Romania due to the Covid-19 pandemic; December 27, 2020 was the first day of the Covid-19 vaccination campaign in Romania; May 2021 was the month in which the highest average number of Covid-19 vaccine doses were administered and October–November 2021 was the period with the highest number of deaths due to Covid-19 in Romania (22).

After the initial collection, for each of the 4 periods, the tweets from the 7-day period with the highest number of tweets were considered and represented the internal dataset (1,300 tweets). The final collection stage also included selecting the tweets with at least one retweet. The two filters were applied in order to obtain a relevant batch of posts related to vaccines, feasible for manual annotation (20, 21). In addition, a random batch of 100 tweets from April 2021 were collected, which represented the external validation dataset.

In order to obtain relevant information for the data analysis phase, the following parameters were extracted for each tweet: date and time, tweet ID, tweet content, number of likes, number of retweets, number of replies. All the information was anonymously collected through snscrape package, which is based on the Twitter API (20).

2.2 Manual annotation

In order to analyze the collected Twitter posts in a relevant manner, all the tweets had to be manually annotated. The tweets were classified in true, neutral or fake based on their text content. The true classification (class 0) meant valid scientific information related to vaccines (whether it was about Covid-19 vaccines or other vaccine types) or true general information regarding the Romanian vaccination campaign. The neutral classification (class 1) regarded irrelevant, ironical or other vaccine related comments, without manipulative or misleading intent. The fake classification (class 2) referred to false or misleading information related to vaccines (both Covid-19 and other types) or the Romanian vaccination campaign. The scientific validity of the posts, when appropriate, was assessed in relation to the official sources of health information (such as the European Center for Disease Prevention and Control, the Summaries of Product Characteristics of the vaccines approved in Romania or trusted health fact check websites) (23–27). It should also be noted that the majority of the tweets were related to Covid vaccines. However, the Twitter posts related to other types of vaccines were not eliminated, in order to increase the variability and complexity of the obtained dataset.

A total number of 9 annotators participated in the task. The external validation dataset was assessed by all 9 annotators and the final classification of each tweet was obtained by a majority vote. The internal dataset was annotated in a similar manner; however, due to the larger number of posts, the internal data was split into 3 parts of similar number of tweets and each part was annotated by 3 different annotators and the final classification was established by a majority vote. Hence, all 9 annotators took part both in the external validation data and in the internal data annotation. In addition, it should be mentioned that when a majority vote could not be applied (due to an equal distribution of votes among the three classes), the tweet was annotated as neutral, in order to ensure an unbiased data analysis. No tweet was eliminated when a majority vote could not be applied, in order to enhance the variability of the processed vaccine data. The agreement between annotators was established on the internal and external data by using Krippendorff's alpha coefficient. The computation of the metric was considered relevant since it provides an ordinal option when assessing the agreement (28). Therefore, the differences between true and neutral annotations are not penalized as hard as the differences between true and fake annotations.

2.3 Text preprocessing

The text content had to be preprocessed to accurately analyze the annotated tweets. The text preprocessing and machine learning development and validation were performed by using Python Programming Language, version 3.9.2 (21).

In order to curate the text and obtain a simplified version, all special characters and stop words were removed from the tweets and all letters were converted to lowercase. The standard stop word list for Romanian provided by spacy was used. In addition, with the aim of providing a bias reduction for the development of the machine learning algorithm, all hyperlinks and words starting with the "@" symbol (with which the content of tweet replies begins) were also eliminated. However, it should be noted that no lemmatization was performed on the selected tweets, since, taking into consideration practical reasons, it was considered that different word forms might provide different meaning and intent to specific phrases; moreover, as an example, as opposed to English language, the Romanian language has a higher number of forms for the noun "vaccine" and the verb "to vaccinate" (29, 30).

2.4 Preliminary analysis

In order to characterize and extract relevant characteristics from the obtained dataset, a preliminary analysis was performed, based on two important methods. The first one implied extracting the most frequent single words and word combinations based on a wordcloud technique, in order to offer a simplified and relevant visualization of the dataset The words were obtained for each of the 3 classes (true, neutral and fake) from the 1,300 vaccine tweets. The second method was applied in order to evaluate the relationships between the manual classification of the Twitter posts and other characteristics. Hence the Spearman's correlation coefficient, along with the p value for statistical significance were computed between the manual classification (true – class 0, neutral – class 1, fake – class 2) and each of the following characteristics of the 1,300 tweets: number of replies, number of retweets, number of likes and the length of the post, quantified by the number of words (31).

2.5 Training and validation of machine learning algorithms

The machine learning algorithm was developed by using Python's scikit-learn package (four classical machine learning models: Support Vector Machines Classifier (SVM), Multilayer Perceptron (MLP, a type of neural networks), Random Forest Classifier (RFC) and an ensemble model (scikit-learn Voting Classifier), developed by averaging the probabilities which were predicted by the SVM and the MLP), as well as Tensorflow (for two specific deep learning models: recurrent convolutional neural networks (RCNN) – Tensorflow implementation and BERT – based on a model which was pretrained on a Romanian 15 GB uncased text corpus, downloaded from Huggingface (dumitrescustefan/bert-base-romanian-uncased-v1 model) and then executed through Tensorflow) (32–34). With regards to the classical machine learning models since scikit-learn does not accept string data as input, the text content had to be converted to numerical data, by using the TfidfVectorizer function. No words were eliminated from the text corpus when performing the string-to-float conversion (34). On the other hand, the deep learning models which were implemented required specific word tokenizers. The RCNN model was built after using the specific Tensorflow tokenizer, while the BERT model implemented the specific Romanian based AutoTokenizer downloaded from the huggingface website (32, 33). All six machine learning algorithms were trained with a constant random seed (35), in order to ensure the reproducibility of the results.

The six machine learning algorithms were validated and compared on the obtained data. They were tested based on their ability of estimating the probability that a specific tweet is true, neutral or fake, as well as of correctly classifying a tweet as being true, neutral or fake. The Area Under the Receiver Operating Characteristic Curve Score (ROC AUC Score, both a One-Versus-One (OVO) strategy and a One-Versus-Rest (OVR) strategy) was used for testing the probability prediction ability of the algorithms and was the most important overall measure for evaluating the machine learning models: the higher the ROC AUC Score is, the better are the probabilities calibrated. In addition, the accuracy, precision, recall, F1 Score and Matthews Correlation Coefficient were used to test the classification ability of the developed models. The Matthews Correlation Coefficient was considered the most important global classification measure, since it provides a relevant bias reduction approach and takes into consideration class imbalance (34, 36, 37).

Both an internal and an external validation were performed for the machine learning algorithms. The internal validation was performed on the 1,300 tweets (internal dataset) and aimed at evaluating the internal consistency of the model combined with the ability of perform on unseen data. Hence, the dataset was split into 4 parts based on the 4 pandemic periods for which the posts were collected (internal period validation). The predictive algorithms were validated 4 times: each time, the training set included the tweets from 3 of the periods; the model was trained on the 3 periods and was evaluated based on the unseen data from the 4th period. Therefore, the model was trained and validated until all the 4 periods represented in turn the test set. In addition, a repeated 5-fold cross-validation (with 10 iterations) was also performed (32–34). However, the internal period validation strategy was considered much more relevant than the cross-validation, since all tweets from a specific period were either in the training set or the test set and the risk that the model was evaluated on similar tweets was significantly reduced.

The external validation of the algorithms implied training the models on the internal data and evaluating their performance on the external dataset represented by the 100 tweets from April 2021 (32–34).

Figure 1 briefly presents the three main strategies within the validation process of the machine learning algorithms.

The final model (SVM) for future identification of specific vaccine tweets was chosen based on the best results obtained in terms of OVO and OVR ROC AUC Scores and was built by taking into consideration the internal dataset (1,300 tweets). The model was implemented based on a probabilistic approach (useful for reliable probability estimation), a radial basis function (RBF) kernel, a penalty parameter of the error term (C value) set to 1, while reducing bias caused by class imbalance and breaking ties according to the confidence values of the RBF. In addition, a detailed analysis was undertaken based on the probability predictions of the final model on 3 tweets from the external data (one true post, one neutral post and one fake post) (34, 38).

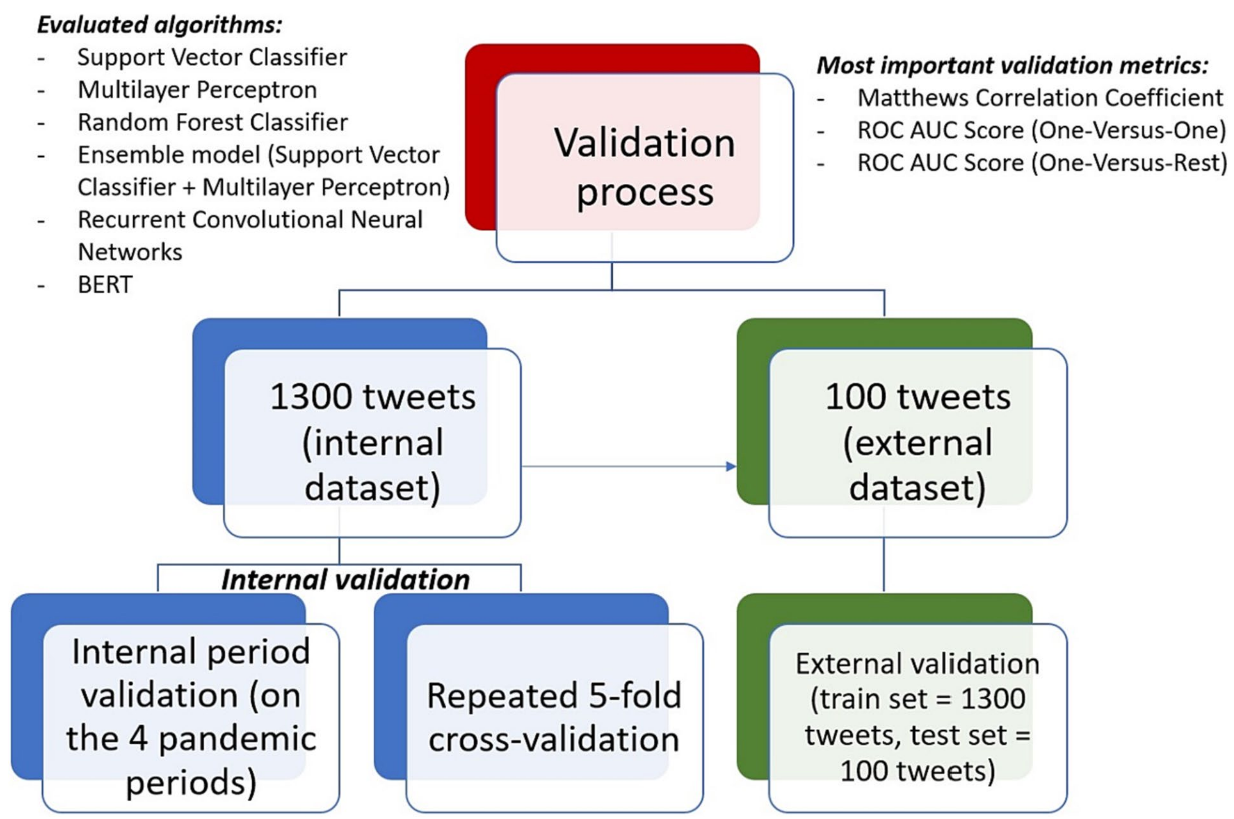


FIGURE 1
The validation process of the machine learning algorithms.

3 Results

3.1 Data collection

A total number of 1,344 tweets were obtained, of which 44 were eliminated due to content in other languages. An additional 100 tweets were randomly selected from another period (April 2021, from the 7-day period with the highest number of tweets, independent on the number of retweets) and represented the dataset for external validation. Table 1 presents the final 7-day periods from which the posts were collected, as well as the number of tweets and the distribution of each of the three classes (true, neutral, and fake) for each weekly time interval.

3.2 Manual annotation

The manual annotation yielded an average inter-agreement Krippendorff's alpha of 0.64 for the internal dataset (0.69 for Team 1, 0.58 for Team 2 and 0.64 for Team 3) and of 0.7 for the external dataset. After applying the majority vote rule, from the 1,300 tweets (in-ternal dataset), a total number of 488 (37.5%) were classified as true, 373 (28.7%) as neutral and 439 (33.8%) as fake. From the 100 tweets representing the external dataset, 53 (53%) were classified as true, 24 (24%) as neutral and 23 (23%) as fake.

In terms of overall inter-annotator agreement, from the 1,300 tweets, 686 (52.8%) reached perfect agreement between the 3 annotators. Of the 100 tweets from the external dataset, 15 (15%) reached perfect agreement between all 9 annotators.

3.3 Preliminary analysis

Table 2 presents the most relevant words and word combinations for the true, neutral and fake tweets within the internal dataset. The most relevant 7 words and word combinations (as considered by the annotators) of the most frequent 30 are presented. The words were translated from Romanian to English and the original Romanian version is also presented in parenthesis, when appropriate. Table 3 summarizes the results obtained by computing the Spearman's correlation coefficient. The p values are not given, since all pairs yielded statistically significant results ($p < 0.05$). Wordcloud representations of the most relevant words written in Romanian and graphical illustration of correlation analysis are shown in Figure 2.

3.4 Validation of machine learning algorithms

The average performance metrics obtained after validating the 6 predictive algorithms (SVM, MLP, RF, the ensemble model – SVM + MLP, RCNN and BERT) are shown in Table 4, along with the standard deviations obtained during multiple runs. All 3 validation types are presented – repeated cross-validation, internal period

TABLE 1 The time periods corresponding to the collected vaccine related tweets.

Period	Number of tweets	Number of true tweets (%)	Number of neutral tweets (%)	Number of fake tweets (%)
Internal data				
March 20, 2020 – March 26, 2020	48	26 (54.2%)	14 (29.2%)	8 (16.6%)
January 15, 2021 – January 21, 2021	491	212 (43.2%)	152 (31%)	127 (25.8%)
May 4, 2021 – May 10, 2021	322	125 (38.8%)	67 (20.8%)	130 (40.4%)
October 19, 2021 – October 25, 2021	439	125 (28.5%)	140 (31.9%)	174 (39.6%)
	Total: 1300 tweets	488 (37.5%)	373 (28.7%)	439 (33.8%)
External validation data				
April 10, 2021 – April 16, 2021	100	53 (53%)	24 (24%)	23 (23%)

validation and external validation. The standard deviations are only shown for the cross-validation, since the internal period validation and external validation yielded null standard deviations values for all algorithms.

The distribution of predicted probabilities generated with all 6 predictive models for the external dataset is illustrated in Figure 3.

3.5 Implementation example of the SVM algorithm

The final algorithm chosen for implementation was the Support Vector Classifier, due to its enhanced predictions quantified through the ROC AUC Score (Table 4). Table 5 presents the probabilities returned by the algorithm for 3 tweets from the external dataset (one true tweet, one neutral tweet and one fake tweet). In order to comply with the General Data Protection Regulation, the exemplified tweets were translated and partially reformulated. In addition, in order to allow a better understanding and exemplification of tweet structure and machine learning predictive abilities, the probabilities for nine extra tweets are presented in Supplementary Table S1 (three true tweets, three neutral tweets and three fake tweets).

The validation of the machine learning predictive algorithms yielded modest results in terms of classification evaluation (Table 4). The Matthews Correlation Coefficient, considered the most important metric used to assess the discriminative power of the implemented models, yielded values ranging from 0.4 to 0.535 for the cross-validation technique, from 0.308 to 0.416 for the internal period validation, as well as from 0.231 to 0.53 for the external validation of the developed algorithms. Overall, by averaging the 2 types of internal validation, BERT resulted in the highest Matthews Correlation Coefficient, of 0.535 for the cross-validation and 0.416 for the internal period validation, with a 5.5% increase for cross-validation, as well as a 6.2% increase for internal period validation as compared to SVM (which yielded a 0.48 Matthews Correlation Coefficient for the cross-validation and a 0.416 value for the internal period validation). However, it should be noted that on the external validation, the SVM algorithm outperformed BERT in terms of raw classification ability, with a 0.53 Matthews Correlation Coefficient, while BERT yielded a value of 0.49 for this validation metric. Since the Matthews Correlation Coefficient is a particular case of the Pearson product–moment correlation coefficient, its values have the same interpretation and hence it can be stated that in most cases, the validation of the BERT and SVM algorithms (the best models with regards to the internal and external validation respectively) yielded moderate to moderately high positive correlations between the true and predicted labels (36).

Nevertheless, the most important evaluation of the machine learning models was represented by the probability prediction evaluation, which tested the ability of the algorithms of estimating well calibrated probabilities, as quantified though the ROC AUC Score (using both an OVO and an OVR approach). As for the Matthews Correlation Coefficient, the ROC AUC Score yielded the lowest results for the internal period validation (OVO ROC AUC Score ranged from 0.702 to 0.787; OVR ROC AUC Score ranged from 0.71 to 0.797), followed by the external validation (OVO ROC AUC Score ranged from 0.718 to 0.818; OVR ROC AUC Score ranged from 0.727 to 0.843), while the cross-validation resulted in the highest ROC AUC values (OVO ROC AUC Score ranged from 0.77 to 0.849; OVR ROC AUC Score ranged from 0.779 to 0.858). Similar to the results obtained in terms of Matthews Correlation Coefficient, the highest ROC AUC Scores were obtained in case of BERT for both internal validation strategies (for cross-validation: OVO ROC AUC = 0.849, OVR ROC AUC = 0.858; for internal period validation: OVO ROC AUC = 0.787, OVR ROC AUC = 0.797), followed by the SVM algorithm. On the other hand, similar to the raw classification validation, SVM resulted in improved results for the ROC AUC Scores when taking into consideration the external validation (OVO ROC AUC = 0.818, OVR ROC AUC = 0.843, as opposed to a 0.806 value in case of OVO ROC AUC and a 0.829 OVR ROC AUC for the BERT model). The enhanced results which were obtained for internal validation in case of BERT might be explained that the current study implemented a pre-trained BERT model based on a large Romanian text corpus of 15 GB. Nevertheless, BERT validation yielded less accurate results than the SVM when taking the external tweets dataset into consideration, which might have been caused by a moderate amount of overfitting on the internal data (1,300 tweets), as well as by the low level of complexity of the processed tweets. In addition, the RCNN model implemented through the Tensorflow library provided poor results both in terms of raw classification and probability estimation (0.702–0.710 ROC AUC Scores for internal period validation and 0.718–0.727 for external validation), which were in most cases the lowest of all 6 implemented machine learning models. These results were obtained despite the high complexity of RCNN and its ability to memorize both temporal and spatial relationships from texts. One reason for the poor results might be related to the relatively short posts which are usually distributed through the Twitter platform and to the fact that the RCNN, in contrast to the implemented BERT model, lacked a specific Romanian based text corpus and did not include any pretrained algorithm. Moreover, we argue that a complex model architecture (with both recurrent and convolutional layers), without any predefined recommendations, is difficult to model so that it reaches optimal results on a text corpus which contains posts in a narrowly spoken language, such as Romanian (32, 33).

In terms of analysis of predicted probabilities (for the external dataset) quantified through the boxplot representations (Figure 3), both SVM (Figure 3A) and BERT (Figure 3F) offered good discrimination when comparing the estimated probabilities with the true (annotated) class. However, the main difference in the performance of the two models can be seen in the probability estimation for the tweets labelled as neutral. More specifically, the SVM offered a more accurate discrimination when predicting the probabilities that the neutral tweets from the external dataset are true, neutral or fake, the probability of being neutral being higher on average than the probability that the tweet was true or fake, which was also reflected in the lower ROC AUC Scores for BERT, when compared to SVM. By contrast, the BERT model returned on average a higher probability that the neutral tweets are fake, as compared to neutral. However, the BERT model discriminated more accurately between the true tweets, as well as the fake tweets and the rest (neutral/fake and true/neutral, respectively), while the SVM offered a more close, but still valid discrimination.

TABLE 2 Most relevant words and word combinations identified through a wordcloud model for each of the 3 classes.

Class	Most relevant words/word combinations
True	Covid-19 "News" AND "Romania" Against Covid (împotriva Covid) Vaccinated persons (personae vaccinate) Vaccine dose (doze de vaccin) Vaccination campaign (campania de vaccinare) Vaccination centers (center de vaccinare)
Neutral	Vaccinated (vaccinat) Against Covid (împotriva Covid) Covid-19 "News" AND "Romania" Vaccination campaign (campania de vaccinare) Klaus Iohannis Vaccination certificate (certificat de vaccinare)
Fake	Vaccinated (vaccinat) Against Covid (împotriva Covid) To give birth (adus pe lume) (aggressive connotation) "Persons" AND "died" ("persoane" ŞI "murit") Adverse reactions (reacții adverse) Experimental vaccine (vaccin experimental) Mandatory vaccination (vaccinarea obligatorie)

4 Discussion

A detailed analysis of a batch of relevant vaccine tweets from several periods within the Covid-19 pandemic was undertaken.

The preliminary analysis implied the manual annotation of a total number of 1,400 tweets, as well as a preliminary analysis for establishing specific word patterns within the posts and the correlations between the manual annotation and other tweet characteristics. The supervised analysis consisted of building and validating several machine learning prediction models based on their ability of estimating the probabilities that a specific Twitter post related to vaccines is true, neutral or fake.

The manual annotation of the collected Twitter posts yielded good results in terms of inter-agreement evaluation based on Krippendorff's alpha (39). The inter-agreement was better for the external dataset (100 tweets, Krippendorff = 0.7) than for the internal one (1,300 tweets, average Krippendorff = 0.64), partly due to the fact that the tweets from the external dataset were annotated by all 9 annotators. Moreover, the Krippendorff obtained for each of the 3 subsets within the internal data showed a certain degree of variability, with its values ranging from 0.58 to 0.69. Indeed, as with other social media posts, the ones from Twitter, even when relating to health issues, are written in a free, subjective manner, since they are mostly written by individual persons which are granted the freedom of expression (40). Therefore, there is a high probability that the annotators ran into several ambiguous tweets and hence the interpretation of such content could have been made different depending on the content and the annotator's subjective interpretation.

In addition, the subjective and diverse ways in which the vaccines posts were written are emphasized in Table 2, where the most relevant word patterns within the 3 classes (true, neutral and fake) are given, as well as in Figures 2A–C. Interestingly, the true posts contained most often different forms of the noun "vaccine" and the verb "to vaccinate," which could be explained by the fact that the true posts, when compared to the neutral and fake ones, contained the most references to news articles and to official data related to the Romanian vaccination campaign, such as the number of persons which were partially and fully vaccinated within a specific time period, the number of administered vaccine doses or the updated vaccine supply. In contrast, the tweets which were labelled as fake (false or

TABLE 3 Correlation analysis results (Spearman's correlation coefficient).

Parameter	Class	Number of replies	Number of retweets	Number of likes	Number of words
Class	1.000	0.198	0.190	0.282	0.222
Number of replies	0.198	1.000	0.453	0.654	0.117
Number of retweets	0.190	0.453	1.000	0.601	0.099
Number of likes	0.282	0.654	0.601	1.000	0.150
Number of words	0.222	0.117	0.099	0.150	1.000

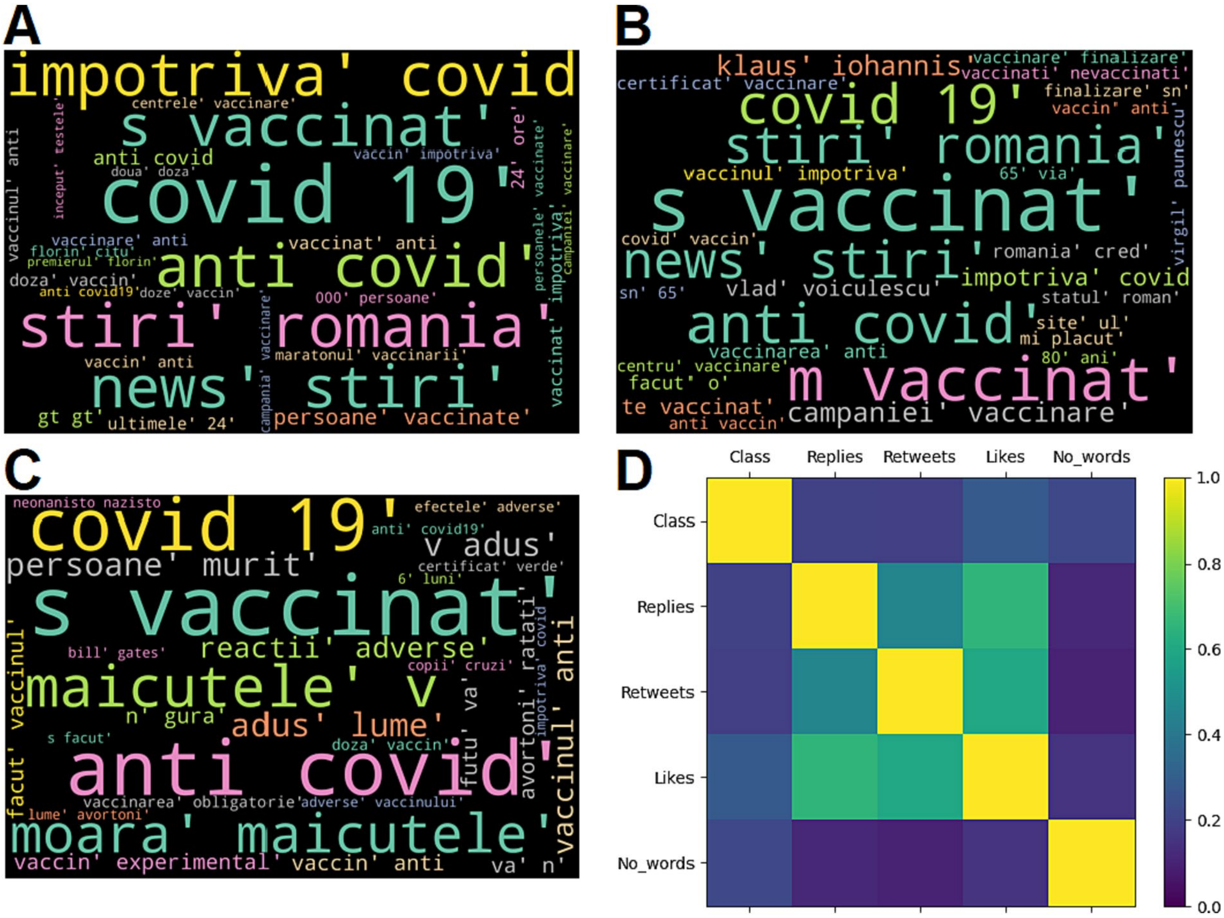


FIGURE 2
Wordcloud representation (the most relevant 30 words written in Romanian) for the tweets labelled as true **(A)**, neutral **(B)** and fake **(C)**; Correlation analysis results **(D)**.

manipulative information) also contained many references to the various forms of "vaccine" and "to vaccinate," but quite often they were referenced in a subjective and manipulative manner. As an example, the word combination "experimental vaccine" was identified as one of the most relevant patterns of the fake tweets, suggesting that the Covid vaccines were not tested enough before being administered in the general population, information which is false, according to various health authorities and fact check websites. Another frequent word pattern identified in the fake tweets was the word "to die" in various forms, combined with the noun "person," suggesting a high degree of mortality for the Covid vaccines. It is true that one of the main reasons of propagating such misinformation would be to make the population believe that the vaccines are dangerous for health and cause many severe adverse reactions (23, 24). In addition, words with aggressive connotation and with negative political references were more frequent within the tweets labelled as fake, as compared to the true and neutral ones. Indeed, it was already emphasized that the spread of vaccine misinformation is partly correlated with hate speech and aggressive language (41). Moreover, the fact that a significant part of pandemic and vaccine misinformation was politically driven (6) warrants better communication campaigns and public health strategies which must address vaccine hesitancy and national and global action plans during health crises (42). This need maybe even more pronounced in countries such as Romania, where the level of Covid vaccination was significantly lower than the level reached in Western European countries (22).

In terms of correlation analysis results, the majority of obtained Spearman's coefficients showed moderate, but statistically significant correlations (Table 3; Figure 2D). The manual classification (considered as class 0 – true, class 1 – neutral, class 2 – fake) was positively correlated with all of the three tweet characteristics: number of replies ($r = 0.198$), number of retweets ($r = 0.190$) and number of likes ($r = 0.282$). These results, even though suggesting a modest positive correlation, imply that the fake vaccine tweets have a higher impact on social media, tending to be retweeted and liked more often than the true and neutral ones (this might in turn prioritize vaccine false information even more because of the Twitter algorithm) (35). The results are similar to the ones reflected in other studies by analyzing the online spread of misinformation (43–45). For example, the work conducted by Vosoughi et al. found among 126,000 stories related to various topics that the ones labelled as false misinformation had a more pronounced spread on Twitter as compared to the valid ones (44). Even though the effect was more pronounced for information about politics, the study raises important awareness, especially considering the fact that several studies show that the online spread of health misinformation may be, at least partially, politically

TABLE 4 Validation results for the machine learning algorithms (mean ± SD*).

Metric	SVM	MLP	RF	Ensemble (SVM + MLP)	RCNN	BERT
Cross-validation						
Accuracy	0.657 ± 0.006	0.603 ± 0.007	0.595 ± 0.004	0.614 ± 0.008	0.623 ± 0.013	0.689 ± 0.008
Precision	0.641 ± 0.007	0.587 ± 0.008	0.600 ± (0.004)	0.598 ± 0.008	0.609 ± 0.015	0.693 ± 0.01
Recall	0.634 ± 0.007	0.588 ± 0.008	0.593 ± 0.004	0.599 ± 0.008	0.602 ± 0.014	0.667 ± 0.006
F1 Score	0.632 ± 0.007	0.585 ± 0.008	0.583 ± 0.004	0.596 ± 0.009	0.599 ± 0.014	0.658 ± 0.013
MCC	0.480 ± 0.01	0.400 ± 0.011	0.401 ± 0.006	0.416 ± 0.012	0.425 ± 0.02	0.535 ± 0.003
ROC AUC (OVO)	0.813 ± 0.004	0.782 ± 0.004	0.779 0.003	0.797 ± 0.005	0.770 ± 0.006	0.849 ± 0.005
ROC AUC (OVR)	0.825 ± 0.004	0.788 ± 0.004	0.783 0.003	0.803 ± 0.005	0.779 ± 0.006	0.858 ± 0.005
Internal period validation*						
Accuracy	0.567	0.546	0.551	0.576	0.537	0.601
Precision	0.572	0.539	0.554	0.562	0.525	0.632
Recall	0.552	0.536	0.526	0.557	0.519	0.573
F1 Score	0.532	0.520	0.515	0.542	0.512	0.539
MCC	0.352	0.317	0.308	0.354	0.291	0.416
ROC AUC (OVO)	0.744	0.738	0.727	0.745	0.702	0.787
ROC AUC (OVR)	0.756	0.743	0.732	0.754	0.710	0.797
External validation*						
Accuracy	0.680	0.648	0.490	0.688	0.480	0.670
Precision	0.661	0.581	0.480	0.636	0.529	0.622
Recall	0.691	0.595	0.465	0.649	0.499	0.630
F1 Score	0.655	0.583	0.456	0.638	0.470	0.606
MCC	0.530	0.435	0.231	0.504	0.271	0.490
ROC AUC (OVO)	0.818	0.772	0.736	0.800	0.718	0.806
ROC AUC (OVR)	0.843	0.796	0.756	0.826	0.727	0.829

MCC, Matthews Correlation Coefficient; SD, standard deviation; * for internal period validation and external validation, a standard deviation of zero was obtained for all algorithms.

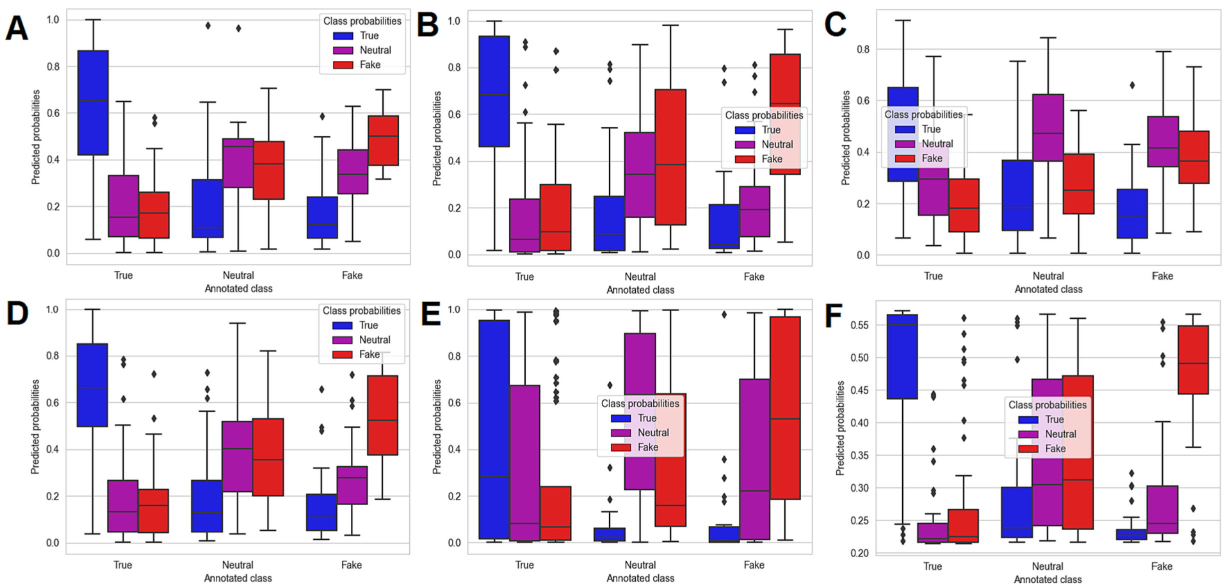


FIGURE 3
Boxplot representation of the predicted probabilities for the machine learning algorithms obtained on the external dataset: SVM algorithm **(A)**, MLP algorithm **(B)**, RF algorithm **(C)**, Ensemble algorithm (SVM + MLP) **(D)**, RCNN algorithm **(E)**, and BERT algorithm **(F)**.

 10.3389/fpubh.2024.1330801

TABLE 5 Detailed example of implementation of SVM algorithm on 3 tweets from the external dataset.

Reformulated tweet content	Predicted probability (true)	Predicted probability (neutral)	Predicted probability (fake)	Predicted class	Annotated class
Tweet A: At 10 days after the second Covid vaccine shot, the risk of getting infected is very low.	72.56%	6.79%	20.66%	Class 0 (true)	Class 0 (true)
Tweet B: I got the vaccine, mind your own business.	9.94%	43.38%	46.68%	Class 2 (fake) – erroneous prediction	Class 1 (neutral)
Tweet C: Mass vaccination would be catastrophic for humankind.	11.60%	23.02%	65.37%	Class 2 (fake)	Class 2 (fake)

driven (45). In addition, a recent longitudinal study conducted by Pierri et al., which analyzed the Covid vaccine misinformation on Twitter throughout 2021, found that the most popular misinformation sources were retweeted more often than the official health information sources, such as The World Health Organization or the Center for Disease Control and Prevention (46). These observations, combined with the fact that online anti-vaccine groups and accounts are more strongly connected and more likely to influence those with neutral views, make enforcing policies on limiting the spread of vaccine misinformation of an utmost importance (5, 6).

Based on the OVO and OVR ROC AUC Score values which were computed for the external validation of the machine learning models (Table 4; Figure 3), the SVM algorithm was chosen for building the final predictive model (34). In addition, all ROC AUC Scores which were obtained when validating the SVM algorithm were above 0.74, which proved the well calibrated probabilities returned by the model. Hence, from a practical point of view, the final SVM model could be used for future identification of the most relevant vaccine related Twitter posts, by sorting the automatically collected large tweet lists based on the predicted probabilities that the specific posts represent true, neutral or fake content. The obtained information, after manually analyzed, if presented through a web platform, could further aid in raising awareness regarding valid information, fake news content, as well as irrelevant information related to vaccines and shared through Twitter platform (20, 34). With regards to the machine learning validation results obtained in other studies, a relevant comparison with the ones from the current studies would be difficult, since the majority of the studies which used vaccine related Twitter content reported the F1 Score as the most important classification evaluation metric (12–17, 37, 47, 48). The F1 Score was computed in the current study as well and can be regarded as an acceptable balanced measure between precision and recall. However, as highlighted by Chicco and Jurman, F1 Score can provide overoptimistic results when evaluating the performance of a predictive model (37). That is the main reason for which the focus in the current study, when evaluating the classification performance of the 6 algorithms, was put on the Matthews Correlation Coefficient (36, 37). Moreover, the reported F1 Scores showed a high degree of variability, some studies reported F1 Scores of under 0.6, while others reported enhanced values, of 0.7–0.8 and others obtained almost perfect values, of over 0.95, while the implemented machine learning algorithms included both classical (such as Random Forest and SVM) and newer model types (such as deep learning and BERT) applied on various languages, such as English, French, Dutch or Moroccan (12–17, 47, 48). As a comparison, the maximum F1 Score which was obtained in the current study SVM ranged from 0.542 (internal period validation – SVM + MLP Ensemble) to 0.658 (cross-validation – BERT) and 0.655 (external validation – SVM). The obtained F1 Score is therefore smaller than that reported by most of the studies; however, as was already mentioned, the most important validation metric in our study, the ROC AUC Score, yielded maximum values of over 0.8, which translates into well calibrated probabilities (34).

Another relevant example is a study which implemented an algorithm based on recurrent convolutional neural networks, with BERT as a word embedding model (49). Even though it did not use the F1 Score as evaluation metric, it achieved superior accuracy, of 0.989, when tested on a real-world dataset, which contains real and fake news propagated during the US General Presidential Election from 2016, with over 20,000 instances, both Twitter and Facebook being widely used for disinformation purpose (49, 50). Other research which aimed at detecting social media non-vaccine related disinformation implemented a hybrid deep learning model (based on recurrent neural networks) and achieved a F1 Score of 0.894, lower than in the study which specifically used BERT (49, 51). As a comparison, in our study, we obtained an accuracy of 0.601–0.689 when evaluating the BERT model; however, our dataset was much smaller and was specifically related to vaccine information distributed through Twitter.

In terms of studies which performed unsupervised analysis on specific disinformation propagated through Twitter, the research performed by Kobayashi et al. is worth mentioning. It included 100 million vaccine related Japanese tweets, on which a topic, as well as a time series analysis were performed (52). In addition, with respect to other studies which specifically evaluated social media disinformation, a study, conducted by De Clerck analyzed the general spread of disinformation through Twitter platform by taking into consideration numerous countries included in the Twitter information operations report. It proposed maximum entropy networks for identifying and quantifying specific patterns in the interactions between numerous Twitter users which might have had an important impact on spread of disinformation (whether or not health related). The analysis had the advantage of applying various algorithms and including a large number of tweets from different countries (e.g., Armenia, China, Russia, Serbia, Turkey) (53). While our study did not implement any form of unsupervised analysis, we argue that the wordcloud representation and correlation analysis which were undertaken give context to the implemented and publicly available machine learning model.

Regarding the practical implementation of the SVM model (Table 5), the given examples provide relevant insights regarding the Romanian tweets structure, as well as the predictive algorithm use case. The first tweet (Tweet A) refers to a valid scientific information – indeed, especially considering the fact that the post was written in April 2021, when the highly contagious Omicron variant and its subvariants were not circulating, two doses of either the RNA or the viral vector vaccine (the ones which were available within the Romanian Vaccination Campaign) significantly reduced the risk of symptomatic Covid-19 (24, 25). The predictive model accurately estimated a 72.56% probability that the content is true, with only 20.66% chance of being misinformation and 6.79% of being neutral. The second tweet (Tweet B) was manually labelled as neutral, being an irrelevant statement regarding someone who got the Covid vaccine. However, the SVM algorithm erroneously classified the tweet as being fake, possibly due to the fact that it was written in a slightly aggressive manner. Nonetheless, when analyzing the predicted probabilities, the model returned a 43.38% risk that the content is neutral and a 46.68% risk that the tweet refers to false information, with only 3.3% higher than the probability of containing neutral information. The third tweet given as a practical example (Tweet C) was manually labelled as false information (fake). The algorithm returned the same classification, with a 65.37% probability that the content is fake, a 23.02% probability that it is neutral and a 11.60% chance of being true. The content of the tweet is a classical conspiracy theory, which tries to suggest that mass vaccination is not only unnecessary, but detrimental. The information is obviously false: the essential role of vaccines in leading to herd immunity and controlling infectious diseases is well established (24).

The current study has a few important advantages. First of all, to our knowledge, this is the first study analyzing vaccine fake news written in Romanian from social media posts. While Romanian is a narrowly spoken language, limited to Romania and Republic of Moldova, we argue that by providing the detailed Python code which includes the specified analyses and the developed predictive machine learning algorithm, as well as the processed (annotated, vectorized and anonymized) internal and external data, our work could be used by other researchers in future studies, with easy translation to other languages (21, 54).

Secondly, as a difference from other similar studies, which used two classes (such as general information and misinformation) during the data labelling process, the current work used for manual annotation three classes (true, neutral and fake) (12, 13, 15–17). It can be argued that this approach enhances the complexity of machine learning models and provides context to the social media analysis. In addition, besides the raw classification, the machine learning models which were developed provide probability estimates, a relevant feature which may aid in future selection of relevant vaccine tweets based on approaches which imply sorting the predicted probabilities, such as the ones presented in Table 5 and Supplementary Table S1. The predictive algorithms were validated in a consistent manner, both for classification and probability estimation. Relevant validation strategies were implemented: the internal period validation ensured the internal consistency of the models with regards to performing on tweets from different pandemic periods, while the external validation ensured the evaluation of the algorithms on unseen data (Table 4; Figure 3) (32–34).

Nevertheless, the current work has a series of limitations. First of all, the number of collected and annotated tweets (1,300 – internal dataset, 100 – external dataset) can be regarded as very low when compared to other studies (therefore, the variability and complexity of the developed SVM algorithm could have been negatively impacted) (12–17, 38). For example, Kunneman et al. conducted a study for measuring the stance towards vaccination (non-Covid vaccines: the messages were extracted prior to the pandemic period), based on a total number of 8,259 annotated tweets written in Dutch; however, the study only achieved a Krippendorff's alpha between 0.27 and 0.35, significantly lower than that from the current study (14). However, Hayawi et al. undertook a vaccine misinformation analysis based on 15,073 annotated English tweets; the annotation process had the advantage of being further validated by health experts and also lead to very good machine learning validation metrics (0.97 precision, 0.98 recall, 0.98 F1 Score) (17). Other studies focused on Covid-19 vaccine hesitancy; while they initially automatically collected large numbers of vaccine related tweets (for example, written in English, Turkish or French), the manual analysis of the content implied, as in our study, a small number of tweets (approximately 1,000–2000) (18, 19, 47, 48). Therefore, it should be noted that while our study comprised indeed in a small dataset chosen for annotation, the fact that the tweets were chosen and annotated following a standardized methodology (selecting 4 relevant pandemic periods and eliminating the tweets with no retweets, as well as the fact that each Twitter post was classified by at least 3 annotators) could ensure reproducibility, especially considering the fact that the Python code for data preprocessing, wordcloud representation, correlation analysis and the development and validation of the machine learning predictive models, as well as the Tfidf vectorized dataset and the final SVM algorithm are publicly available at https://github.com/valeanuandrei/vaccine-tweets-ro-research (54).

Secondly, even though the results of the probability validation were satisfactory, the evaluation of the classification ability of the machine learning algorithms, especially for the internal period validation (a maximum Matthews Correlation Coefficient of under 0.42 and a maximum F1 Score of under 0.55), yielded modest results (36).

Therefore, the implemented natural language processing and data mining techniques, combined with the 12 practical examples of tweet classification and probability prediction, provide relevant insights regarding vaccine general information and misinformation spread through Twitter platform and written in Romanian. Future studies must aim at collecting a large number of tweets and classifying them based on a semi-supervised approach, in order to enhance the variability, complexity and predictive ability of the machine learning algorithm. After these steps are undertaken, an online platform might be developed, based on identifying new vaccine related Twitter content, to aid in raising awareness regarding the vaccine misinformation shared through social media and consequently reduce vaccine hesitancy (55, 56).

5 Conclusion

A study aiming at analyzing and automatically classifying relevant vaccine related posts from Twitter content was undertaken. A total number of 1,400 tweets from relevant pandemic periods were collected and manually classified as true information, neutral in-formation or fake information related to vaccines. Both an unsupervised analysis (consisting of a wordcloud evaluation and a correlation analysis) and a supervised analysis (based on building several predictive machine learning algorithms – SVM, MLP, RF, an ensemble voting classifier: SVM + MLP, as well as complex deep learning models: RCNN and BERT) were implemented. The correlation analysis yielded moderate, but significant positive correlations between the tweets labelled as misinformation and the tweet engagement metrics, quantified through the number of replies, retweets and likes. The ma-chine learning algorithms were mainly validated based on their ability of estimating the probability that a specific tweet is true, neutral or fake. The optimal results were obtained for the Support Vector Classifier, with a ROC AUC Score ranging from 0.744 to 0.843 and BERT, with a ROC AUC Score ranging from 0.787 to 0.858. Future studies must aim in en-larging the vaccine tweets database and optimizing the machine learning predictive abilities, in order to automatically identify and classify new vaccine related valid, neutral and false information distributed through Twitter platform.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

AV: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. DM: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. CA: Conceptualization, Data curation, Methodology, Writing – review & editing. CP: Conceptualization, Data curation, Methodology, Writing – review & editing. AI: Data curation, Methodology, Writing – review & editing. MH: Data curation, Methodology, Writing – review & editing. VP: Data curation, Methodology, Writing – review & editing. EB: Data curation, Methodology, Writing – review & editing. CC: Conceptualization, Writing – review & editing. SN: Conceptualization, Writing – review & editing. CM: Writing – review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by “Carol Davila” University of Medicine and Pharmacy, Bucharest, Romania, through *Publish not Perish* institutional program.

Acknowledgments

The current paper is based on a preprint version, available on a preprint server at https://www.medrxiv.org/content/10.1101/2023.08.19.23294319v1 (57).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1330801/full#supplementary-material

References

1. Li X, Mukandavire C, Cucunubá ZM, Echeverria Londono S, Abbas K, Clapham HE, et al. Estimating the health impact of vaccination against ten pathogens in 98 low-income and middle-income countries from 2000 to 2030: a modelling study. *Lancet*. (2021) 397:398–408. doi: 10.1016/S0140-6736(20)32657-X

2. Garon JR, Cochi SL, Orenstein WA. The challenge of global poliomyelitis eradication. *Infect Dis Clin N Am*. (2015) 29:651–65. doi: 10.1016/j.idc.2015.07.003

3. Centers for disease control and prevention. Smallpox. (2022). Available at: https://www.cdc.gov/smallpox/index.html.

4. Meslé MM, Brown J, Mook P, Hagan J, Pastore R, Bundle N, et al. Estimated number of deaths directly averted in people 60 years and older as a result of COVID-19 vaccination in the WHO European region, December 2020 to November 2021. *Eur Secur*. (2021) 26:2101021. doi: 10.2807/1560-7917.ES.2021.26.47.2101021

5. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro-and anti-vaccination views. *Nature*. (2020) 582:230–3. doi: 10.1038/s41586-020-2281-1

6. Center for Countering Digital Hate. The Disinformation Dozen. (2021). Available at: https://counterhate.com/research/the-disinformation-dozen/.

7. Allington D, McAndrew S, Moxham-Hall VL, Duffy B. Media usage predicts intention to be vaccinated against SARS-CoV-2 in the US and the UK. *Vaccine*. (2021) 39:2595–603. doi: 10.1016/j.vaccine.2021.02.054

8. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav*. (2021) 5:337–48. doi: 10.1038/s41562-021-01056-1

9. Popa AD, Enache AI, Popa IV, Antoniu SA, Dragomir RA, Burlacu A. Determinants of the hesitancy toward COVID-19 vaccination in eastern European countries and the relationship with health and vaccine literacy: a literature review. *Vaccines (Basel)*. (2022) 10:672. doi: 10.3390/vaccines10050672

10. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: identifying negative sentiment about HPV vaccines on twitter. *Stud Health Technol Inform*. (2015) 216:761–5. doi: 10.3233/978-1-61499-564-7-761

11. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on twitter: a cross-sectional study of users in Australia, Canada and the UK. *BMJ Open*. (2017) 7:e016869. doi: 10.1136/bmjopen-2017-016869

12. Abdul-Mageed M, Elmandany AR, Pabbi D, Verma K, Lin R. Mega-COV: a billion-scale dataset of 100+ languages for COVID-19. (2020). Available at: https://arxiv.org/abs/200506012

13. Cui L, Lee D. CoAID: COVID-19 healthcare misinformation dataset. (2020). Available at: https://arxiv.org/abs/200600885.

14. Kunneman F, Lambooij M, Wong A, Bosch AVD, Mollema L. Monitoring stance towards vaccination in twitter messages. *BMC Med Inform Decis Mak*. (2020) 20:33. doi: 10.1186/s12911-020-1046-y

15. Madani Y, Erritali M, Bouikhalene B. Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets. *Results Phys*. (2021) 25:104266. doi: 10.1016/j.rinp.2021.104266

16. To QG, To KG, Huynh VAN, Nguyen NTQ, Ngo DTN, Alley SJ, et al. Applying machine learning to identify anti-vaccination tweets during the COVID-19 pandemic. *Int J Environ Res Public Health*. (2021) 18:4069. doi: 10.3390/ijerph18084069

17. Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS. ANTi-vax: a novel twitter dataset for COVID-19 vaccine misinformation detection. *Public Health*. (2022) 203:23–30. doi: 10.1016/j.puhe.2021.11.022

18. Küçükali H, Ataç Ö, Palteki AS, Tokaç AZ, Hayran O. Vaccine hesitancy and anti-vaccination attitudes during the start of COVID-19 vaccination program: a content analysis on twitter data. *Vaccines (Basel)*. (2022) 10:161. doi: 10.3390/vaccines10020161

19. Lanyi K, Green R, Craig D, Marshall C. COVID-19 vaccine hesitancy: analysing twitter to identify barriers to vaccination in a low uptake region of the UK. *Front Digit Health*. (2022) 3:804855. doi: 10.3389/fdgth.2021.804855

20. JustAnotherArchivist (GitHub Repository). Snscrape. (2022). Available at: https://github.com/JustAnotherArchivist/snscrape.

21. Python Software Foundation. Python language reference, version 3.9.2. (2021). Available at: http://www.python.org.

22. Ritchie H. Our world in data. 2020. Coronavirus pandemic (COVID-19). Available at: https://ourworldindata.org/coronavirus.

23. European Centre for Disease Prevention and Control. Covid-19. (2022). Available at: https://www.ecdc.europa.eu/en/covid-19.

24. The Annenberg Public Policy Center. COVID-19 misconceptions. (2022). Available at: https://www.factcheck.org/covid-misconceptions/.

25. European Medicines Agency. Treatments and vaccines for COVID-19. (2022). Available at: https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines-covid-19.

26. Elhadad MK, Li KF, Gebali F. Detecting misleading information on COVID-19. *IEEE Access*. (2020) 8:165201–15. doi: 10.1109/ACCESS.2020.3022867

27. Shu K, Mahudeswaran D, Wang S, Lee D, Liu H. FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*. (2020) 8:171–88. doi: 10.1089/big.2020.0062

28. Python Software Foundation. krippendorff 0.5.1. (2022). Available at: https://pypi.org/project/krippendorff/.

29. Romanian Academy ("Iorgu Iordan" Linguistic Institute). *The Romanian explanatory dictionary*. Ed. Univers Enciclopedic gold. (2016).

30. Merriam-Webster Inc. Merriam-Webster online dictionary. (2022). Available at: https://www.merriam-webster.com/.

31. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. (2020) 17:261–72. doi: 10.1038/s41592-019-0686-2

32. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *TensorFlow: large-scale machine learning on heterogeneous distributed systems*. arXiv (2016).

33. Dumitrescu SD, Avram AM, Pyysalo S. *The birth of Romanian BERT*. Association for Computational Linguistics (2020).

34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. (2011). 12:2825–30.

35. Twitter Inc. About your home timeline on twitter. (2022). Available at: https://help.twitter.com/en/using-twitter/twitter-timeline.

36. Powers DMW. *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. arXiv (2020).

37. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. (2020) 21:6. doi: 10.1186/s12864-019-6413-7

38. Scikit-learn developers. Sklearn.svm.SVC. (2022). Available at: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVM.html.

39. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol*. (2016) 16:93. doi: 10.1186/s12874-016-0200-9

40. Twitter Inc. Defending and respecting the rights of people using our service. (2022). Available at: https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice.

41. Nguyen T. Merging public health and automated approaches to address online hate speech. *AI Ethics*. (2023). doi: 10.1007/s43681-023-00281-w

42. Khatri RB, Endalamaw A, Erku D, Wolka E, Nigatu F, Zewdie A, et al. Preparedness, impacts, and responses of public health emergencies towards health security: qualitative synthesis of evidence. *Arch Public Health*. (2023) 81:208. doi: 10.1186/s13690-023-01223-y

43. del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. *Proc Natl Acad Sci*. (2016) 113:554–9. doi: 10.1073/pnas.1517441113

44. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*. (2018) 359:1146–51. doi: 10.1126/science.aap9559

45. Wang X, Zhang M, Fan W, Zhao K. Understanding the spread of COVID-19 misinformation on social media: the effects of topics and a political leader's nudge. *J Assoc Inf Sci Technol*. (2022) 73:726–37. doi: 10.1002/asi.24576

46. Pierri F, DeVerna MR, Yang KC, Axelrod D, Bryden J, Menczer F. One year of COVID-19 vaccine misinformation on twitter: longitudinal study. *J Med Internet Res*. (2023) 25:e42227. doi: 10.2196/42227

47. Sauvayre R, Vernier J, Chauvière C. An analysis of French-language tweets about COVID-19 vaccines: supervised learning approach. *JMIR Med Inform*. (2022) 10:e37831. doi: 10.2196/37831

48. Qorib M, Oladunni T, Denis M, Ososanya E, Cotae P. Covid-19 vaccine hesitancy: text mining, sentiment analysis and machine learning on COVID-19 vaccination twitter dataset. *Expert Syst Appl*. (2023) 212:118715. doi: 10.1016/j.eswa.2022.118715

49. Kaliyar RK, Goswami A, Narang P. FakeBERT: fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl*. (2021) 80:11765–88. doi: 10.1007/s11042-020-10183-2

50. Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y. Combating fake news: a survey on identification and mitigation techniques. *ACM Trans Intell Syst Technol*. (2019) 10:1–42. doi: 10.1145/3305260

51. Ruchansky N, Seo S, Liu Y. CSI: a hybrid deep model for fake news detection. *CIKM '17: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. (2017) 797–806. doi: 10.1145/3132847.3132877

52. Kobayashi R, Takedomi Y, Nakayama Y, Suda T, Uno T, Hashimoto T, et al. Evolution of public opinion on COVID-19 vaccination in Japan: large-scale twitter data analysis. *J Med Internet Res*. (2022) 24:e41928. doi: 10.2196/41928

53. De Clerck B, Rocha LEC, Van Utterbeeck F. Maximum entropy networks for large scale social network node analysis. *Appl Netw Sci*. (2022) 7:68. doi: 10.1007/s41109-022-00506-7

54. Valeanu A GitHub R. Vaccine-tweets-ro-research. (2022). Available at: https://github.com/valeanuandrei/vaccine-tweets-ro-research.

55. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. *BMJ Glob Health*. (2020) 5:e004206. doi: 10.1136/bmjgh-2020-004206

56. Pierri F, Perry BL, DeVerna MR, Yang KC, Flammini A, Menczer F, et al. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Sci Rep*. (2022) 12:5966. doi: 10.1038/s41598-022-10070-w

57. Valeanu A, Mihai DP, Andrei C, Puscasu C, Ionica AM, Hinoveanu MI, et al. Identification, analysis and prediction of valid and false information related to vaccines from Romanian tweets. *medRxiv*. (2023). doi: 10.1101/2023.08.19.23294319

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 07 February 2024
DOI 10.3389/fpubh.2024.1334881

Measuring the performance of computer vision artificial intelligence to interpret images of HIV self-testing results

Stephanie D. Roche[1*†], Obinna I. Ekwunife[1†], Rouella Mendonca[2], Benn Kwach[3], Victor Omollo[3], Shengruo Zhang[4], Patricia Ongwen[5], David Hattery[2], Sam Smedinghoff[2], Sarah Morris[2], Daniel Were[5], Dino Rech[2], Elizabeth A. Bukusi[3,6,7] and Katrina F. Ortblad[1]

[1]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, United States, [2]Audere, Seattle, WA, United States, [3]Centre for Microbiology Research, Kenya Medical Research Institute, Kisumu, Kenya, [4]Department of Epidemiology, University of Washington, Seattle, WA, United States, [5]Jhpiego, Nairobi, Kenya, [6]Department of Global Health, University of Washington, Seattle, WA, United States, [7]Department of Obstetrics and Gynecology, University of Washington, Seattle, WA, United States

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Kathiravan Srinivasan,
Vellore Institute of Technology, India
Ricardo Valentim,
Federal University of Rio Grande do Norte, Brazil

*CORRESPONDENCE
Stephanie D. Roche
✉ sroche@fredhutch.org

†These authors share first authorship

RECEIVED 08 November 2023
ACCEPTED 22 January 2024
PUBLISHED 07 February 2024

CITATION
Roche SD, Ekwunife OI, Mendonca R, Kwach B, Omollo V, Zhang S, Ongwen P, Hattery D, Smedinghoff S, Morris S, Were D, Rech D, Bukusi EA and Ortblad KF (2024) Measuring the performance of computer vision artificial intelligence to interpret images of HIV self-testing results. *Front. Public Health* 12:1334881. doi: 10.3389/fpubh.2024.1334881

Introduction: HIV self-testing (HIVST) is highly sensitive and specific, addresses known barriers to HIV testing (such as stigma), and is recommended by the World Health Organization as a testing option for the delivery of HIV pre-exposure prophylaxis (PrEP). Nevertheless, HIVST remains underutilized as a diagnostic tool in community-based, differentiated HIV service delivery models, possibly due to concerns about result misinterpretation, which could lead to inadvertent onward transmission of HIV, delays in antiretroviral therapy (ART) initiation, and incorrect initiation on PrEP. Ensuring that HIVST results are accurately interpreted for correct clinical decisions will be critical to maximizing HIVST's potential. Early evidence from a few small pilot studies suggests that artificial intelligence (AI) computer vision and machine learning could potentially assist with this task. As part of a broader study that task-shifted HIV testing to a new setting and cadre of healthcare provider (pharmaceutical technologists at private pharmacies) in Kenya, we sought to understand how well AI technology performed at interpreting HIVST results.

Methods: At 20 private pharmacies in Kisumu, Kenya, we offered free blood-based HIVST to clients ≥18 years purchasing products indicative of sexual activity (e.g., condoms). Trained pharmacy providers assisted clients with HIVST (as needed), photographed the completed HIVST, and uploaded the photo to a web-based platform. In real time, each self-test was interpreted independently by the (1) client and (2) pharmacy provider, with the HIVST images subsequently interpreted by (3) an AI algorithm (trained on lab-captured images of HIVST results) and (4) an expert panel of three HIVST readers. Using the expert panel's determination as the ground truth, we calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for HIVST result interpretation for the AI algorithm as well as for pharmacy clients and providers, for comparison.

Results: From March to June 2022, we screened 1,691 pharmacy clients and enrolled 1,500 in the study. All clients completed HIVST. Among 854 clients whose HIVST images were of sufficient quality to be interpretable by the AI algorithm, 63% (540/854) were female, median age was 26 years (interquartile range: 22–31), and 39% (335/855) reported casual sexual partners. The expert panel identified 94.9% (808/854) of HIVST images as HIV-negative, 5.1% (44/854) as HIV-positive, and 0.2% (2/854) as indeterminant. The AI algorithm demonstrated perfect sensitivity (100%), perfect NPV (100%), and 98.8% specificity, and 81.5% PPV (81.5%) due to seven false-positive results. By comparison, pharmacy clients and providers demonstrated lower sensitivity (93.2% and 97.7% respectively) and NPV (99.6% and 99.9% respectively) but perfect specificity (100%) and perfect PPV (100%).

Conclusions: AI computer vision technology shows promise as a tool for providing additional quality assurance of HIV testing, particularly for catching Type II error (false-negative test interpretations) committed by human end-users. We discuss possible use cases for this technology to support differentiated HIV service delivery and identify areas for future research that is needed to assess the potential impacts—both positive and negative—of deploying this technology in real-world HIV service delivery settings.

KEYWORDS

artificial intelligence, HIV self-testing, HIV prevention, mHealth, sub-Saharan Africa, differentiated service delivery, Kenya, pharmacy

Introduction

Despite significant progress toward global HIV targets, the world is not currently on track to achieve the Sustainable Development Goal of ending AIDS as a public health threat by 2030 (1). As of 2023, only five countries have achieved the UNAIDS 2025 targets for testing, treatment, and viral suppression—known as "95–95–95"—which stand globally at 86% of people living with HIV (PLHIV) knowing their status; 89% of PLHIV on antiretroviral therapy (ART); and 93% of those on ART virally suppressed (2). Similarly, with 1.3 million new HIV infections in 2022 (2) and only 4.3 million people ever-initiated on daily oral HIV pre-exposure prophylaxis (PrEP) (3), significant work will be needed to achieve the UNAIDS 2025 prevention targets to reduce new infections to 370,000 and make PrEP available to 10 million people (4).

Closing these gaps will likely require intensified differentiated service delivery (DSD) strategies to mitigate barriers to accessing and delivering HIV services at clinics, such as HIV-associated stigma, distance, understaffing, and long wait times (5–9). DSD is a person-centered approach recommended by the WHO for both HIV treatment and prevention interventions that aims to simplify delivery; reduce burden on clients, providers, and healthcare systems; and make HIV services more accessible and acceptable to the individuals in need (5). In practice, DSD models often move service delivery outside of traditional health facilities, task-shift delivery to new cadres of providers, and/or incorporate new innovations and technologies, such as electronic adherence monitors, SMS reminders, and decision support tools (10, 11).

HIV self-testing is one innovation that has been underutilized in HIV DSD (12, 13). Despite HIV self-tests (HIVSTs) having high sensitivity (93.6%—100%) and specificity (99.1%—100%) (14–17); increasing recent and frequent HIV testing in diverse populations (18); being largely acceptable to clients (19); and featuring in the national policies of 98 countries (20), implementation has lagged considerably, with only 52 countries routinely implementing HIV self-testing (20). Additionally, the WHO only recently (in July 2023) endorsed HIV self-testing as an additional testing strategy that should be offered at health facilities and one that should be used for PrEP initiation, continuation, and re-starts (21). To date, HIV self-testing has primarily been used as a screening tool, rather than a diagnostic tool, in part, due to concerns that the quality of test administration and interpretation would be lower than standard-of-care rapid diagnostic testing performed by trained HIV testing services (HTS) providers. Of particular concern are false-negatives that could lead to delays in ART initiation and/or incorrect initiation on PrEP, the latter of which carries an increased risk of developing HIV drug resistance (22). Scale-up of DSD models that fully leverage HIVST's accuracy, privacy, and convenience will likely be contingent on assuaging concerns about testing quality.

Computer vision, a field of artificial intelligence (AI) that gleans meaningful information from visual data (e.g., digital images), could potentially assist with ensuring HIV self-testing quality. The peer-reviewed literature includes dozens of examples of AI algorithms that, in research studies, have performed as well as trained healthcare providers at diagnosing conditions based on medical imaging data (e.g., X-rays, CT scans) (23); and in real-world healthcare delivery settings, some proprietary AI algorithms are already being used to assist clinicians with diagnosis (24). A handful of studies have found AI computer vision technology to also perform well at interpreting rapid antigen/antibody diagnostic tests (RDTs) in the form of lateral flow devices. Such tests indicate positivity by producing one or more test lines. Using machine learning, an AI algorithm can be trained on a set of images of completed tests (all of the same brand) to recognize line patterns for positive and negative results; thereafter, when fed an image of a completed test of that same brand, the AI algorithm can make a determination about (i.e., interpret) the test result. Published examples of this use case largely focus on rapid tests for COVID-19 (25–30), with a handful of examples from malaria (31), influenza (32), Cryptococcosis fungal infection (33), and HIV (34). The HIV example comes from Turbé et al. (34), who trained an AI algorithm using 11,374 images of two brands of HIV RDTs: ABON HIV 1/2/O Tri-Line HIV RDT [ABON Biopharm (Hangzhou) Co., Ltd.,] and Advanced Quality One Step Anti-HIV (1&2) Tri-line Test (InTec PRODUCTS, INC.). The training set images had been collected as part of routine household surveillance in KwaZulu-Natal, South Africa, with the RDTs conducted and photographed by trained fieldworkers. To assess the algorithm's performance, 40 RDTs of the same brands were activated using human blood samples. Each completed test was interpreted independently by 5 healthcare providers (2 nurses and 3 community health workers) using traditional visual interpretation. After interpreting the test, each healthcare worker photographed it using a Samsung tablet running an mHealth application. The images were then independently interpreted by the AI algorithm and by a panel of expert HIV test readers. The final dataset included 190 images. Using the expert panel's interpretation as the "ground truth," the AI algorithm was found to have 97.8% sensitivity (due to 2 false-negatives) and 100% specificity (i.e., 0 false-positives) and outperformed the healthcare worker group, which had 95.6% sensitivity (due to 4 false-negatives) and 89.0% specificity (due to 11 false-positives). The authors acknowledge several limitations of their evaluation (e.g., small sample size) but note the potential of AI computer vision technology to reduce the risk of false-positive and false-negative HIV RDT results.

To date, no study has investigated how well an AI algorithm might perform at interpreting images of HIVSTs conducted in real-world (non-laboratory) settings by clients. Compared to images of HIV RDTs performed by trained healthcare providers or activated by trained lab technicians using human blood samples, images of HIVSTs conducted by real-world clients on themselves might vary in ways that could affect AI algorithm performance. For example, inexperienced self-testers might apply smaller amounts of the blood sample to the test strip, making test lines harder to detect (35). We, therefore, sought to evaluate the performance of an AI algorithm at interpreting images of HIVSTs collected during routine service delivery at private, community-based pharmacies in Kenya. Secondarily, we sought to understand how the AI algorithm performed compared to trained pharmacy providers and clients—two groups of individuals who are legally allowed to conduct HIV self-testing in Kenya—to inform discussions of whether and how the Kenya Ministry of Health (MOH) might incorporate this technology into models of differentiated HIV service delivery.

Materials and methods

This study is part of a larger, observational study (hereafter, "HIVST Performance Study") measuring the performance of blood-based HIVSTs, compared to standard-of-care RDTs delivered by certified HTS providers, at private community-based pharmacies—a venue to which the Kenya MOH is interested in expanding HIV services as part of its HIV DSD strategy (36). The methods of this larger study have been described elsewhere (37). Below, we summarize the methods that are pertinent to this present study on AI algorithm performance.

Study design and setting

This study uses cross-sectional data from the HIVST Performance Study, which was observational in design and conducted at 20 private pharmacies in Kisumu County in western Kenya from March to June 2022. Kisumu has a population-level HIV prevalence of ~18%—one of the highest in the country (38). Pharmacies were eligible to serve as study sites if they were privately owned (i.e., not supported with government funding), operating legally (i.e., currently registered with Kenya's drug regulatory authority), had a back room where HIV testing and counseling could occur in private, and had on staff at least one full-time licensed pharmacist and/or pharmaceutical technologist—two cadres of pharmacy professionals that the Kenya MOH has expressed interest in leveraging for HIV service delivery (36)—who was willing to participate in research activities. We partnered with the Kisumu County Ministry of Health to identify and purposively select licensed pharmacies to serve as study sites (39).

Participants

Eligible pharmacy providers were ≥18 years old, worked at one of the study pharmacies, and were willing to complete the required training to deliver the intervention. All pharmacy providers attended a one-day in-person training on offering and assisting (if desired by clients) free HIV self-testing to clients purchasing products or services related to sexual and reproductive health (SRH), such as condoms and emergency contraception; conducting and interpreting Mylan blood-based HIVSTs (Mylan Pharmaceuticals Private Limited, India, manufactured in South Africa by Atomo); and providing counseling.

Eligible pharmacy clients were ≥18 years old and self-reported being HIV-negative or not knowing their HIV status, not currently taking PrEP or ART, and engaging (in the past 6 months) in at least one behavior associated with risk of HIV acquisition (e.g., condomless sex) included in Kenya's eight-item Rapid Assessment Screening Tool (40) or having a potential exposure to HIV within the past 72 h. To recruit pharmacy clients, participating pharmacy providers asked those purchasing the above-described SRH products if they would be interested in participating in a research study that was offering free HIVST services to eligible individuals. This being an observational study intended to evaluate outcomes of interest in their natural context, pharmacy client participants self-selected into the study (i.e., the study did not use probability sampling to recruit study participants).

Training of HIVST expert readers

Twenty HIVST expert readers were contracted by Audere from IndiVillage (Bangalore, India), a B-Corp-certified company that is compliant with US Health Insurance Portability and

Accountability Act (HIPAA), the International Organization for Standardization (ISO), System and Organization Controls (SOC), and the General Data Protection Regulation (GDPR). IndiVillage offers dataset creation, annotation, and labeling services for natural language processing and computer vision (41). In addition to the training and experience obtained as an expert test reader for IndiVillage, all of the test readers contracted by Audere for this evaluation completed a two-week training during which they received in-depth instruction specifically on Mylan HIVST result interpretation. To pass this training, attendees were presented with 50 images of Mylan HIVST results and needed to correctly interpret at least 48 (96%) of them. All 20 attendees passed the training; their role is described in the next section.

Study procedures

Figure 1 illustrates the flow of key study procedures in nine steps. As described above, prospective study participants were initially engaged by pharmacy providers (Step 1). In a private back room of the pharmacy, these individuals were assessed for eligibility by a trained research assistant (RA); eligible and interested clients were administered informed consent and provided with pre-test counseling by the RA, who was also a certified HTS provider (Step 2). Pharmacy providers gave clients the option to conduct the HIVST on their own or with their assistance on any aspect of administration except for result interpretation (Step 3). Clients who opted to conduct the HIVST on their own were instructed to follow the directions included in the test kit package and let the pharmacy provider know if they had questions. All HIVSTs were third-generation Mylan blood-based HIVST kits supplied to the pharmacy by the research team.

After the HIVST was completed, the RA photographed it using a Samsung Galaxy A6 tablet, with the image automatically uploaded to the study's secure electronic data collection platform, CommCare (Dimagi, USA) (Step 4). The result was interpreted independently by the pharmacy client (Step 5), then separately by the pharmacy provider (Step 6). These interpretations—which could be "negative," "positive" or "indeterminant"—were entered into CommCare by the RA. As part of the larger HIVST Performance Study, the client then received standard-of-care HIV rapid diagnostic testing with the RA/HTS provider and, based on the RA/HTS provider's reading of the HIV RDT result, the client received post-testing counseling, was encouraged to consider initiating HIV prevention or treatment services, and was issued a referral slip to a nearby public clinic offering these services for free (Step 7). The RA surveyed each pharmacy provider and client to capture their demographic information.

Within 1 week, each HIVST image was interpreted by a subset of three of the 20 trained HIVST readers (hereafter, "expert panel") who interpreted each image independently and without access to any additional client data (Step 8). The study treated the expert panel's majority consensus (i.e., the interpretation that two or more of the three HIVST readers gave) as the "ground truth" against which to judge the performance of the AI algorithm (and, secondarily, the performance of the pharmacy client and pharmacy provider). Lastly, each HIVST image was interpreted by an AI algorithm (described in the next section), which also was not provided with any additional client data (Step 9).

Development of the AI algorithm

A non-profit organization specializing in digital health, Audere (Seattle, USA), developed a platform called HealthPulse AI that leverages AI algorithms for rapid test identification and interpretation. For this study, Audere developed an AI algorithm specifically for interpreting Mylan HIVSTs. Multiple computer vision and machine learning (ML) models comprise the AI algorithm, including (1) an object detector that locates the HIVST and its sub-parts within the image and identifies the HIVST type; (2) a second object detector that examines the HIVST result window (found by the prior object detector) and locates the test and control line regions; (3) a classifier that examines each line region of the result window (found by the second object detector) and outputs line presence probability; and (4) an Image Quality Assurance pipeline that flags adverse image conditions, like blur, low lighting, and overexposure (42). Google's MediaPipe (GMP) framework is used to route images through this sequence of models and return results (43). Customized versions of YOLOX Nano—themselves pretrained on the COCO dataset—are used for both object detectors. A smaller, customized version of MixNet is used as a classifier for the final stage.

To determine how many images the AI algorithm should be trained on to minimize the risk of bias (i.e., the risk of creating an algorithm that only performs well on certain types of images), a comprehensive combinatorics approach was used, taking into consideration possible sources of variation in images when collected in real-world settings, such as environmental conditions (e.g., different lighting, test positioning, image backgrounds); camera quality (affecting, e.g., focal length, image resolution); the test result itself; and test activation (i.e., darkness of lines). Ultimately, it was determined that a reference set of 11,074 images−6,074 images on which to fit and train the ML models ('the training set") and 5,000 images on which to subsequently evaluate the model ("the hold-out set")—would suffice. The size of this reference set is similar to that used in other studies (34).

The 6,074 images in the training set—which was 82% (n = 4,965) negative tests and 18% (n = 1,109) positive tests—included 6,048 images of Mylan HIVSTs that had been activated by trained laboratory technicians using human blood samples and 26 images of Mylan HIVSTs featuring faint positive lines that had been conducted by clients on themselves in real-world pharmacy settings. By design, both the training and hold-out sets included images captured under various environmental conditions (as described above) and using the following iOS and Android smartphone devices commonly used in low- and middle-income country settings (44) that vary in camera quality: iPhone 12, SamsungA2Core, Samsung x20, Alcatel U3, Mobicel Geo Trendy Lite, Hurricane Link, Tecno Pop2 mini, Hisense U605, and Ulefone Note 8P. Each image in the reference set was labeled as "positive," "negative," or "indeterminant" by 3 human HIVST expert readers, as described above, with the majority consensus used as the "ground truth" result.

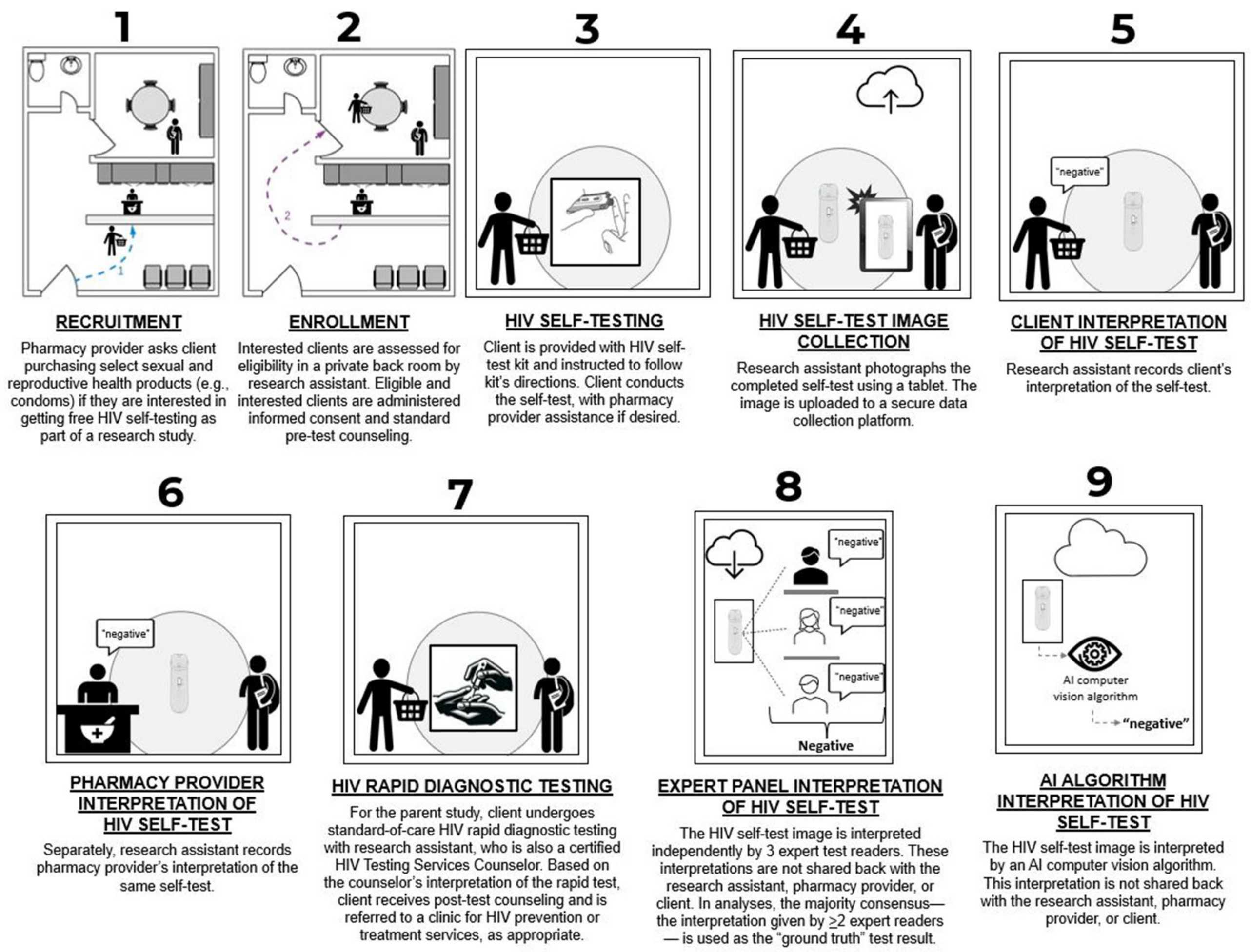


FIGURE 1
Flow diagram of study procedures.

After the model was developed, its performance was evaluated using the hold-out set. Based on the intended use case (clinical decision-making support), the a priori accuracy goal for the AI algorithm was a weighted F1 score of at least 95. During the first round of evaluation, the AI algorithm's weighted F1 score was 98.2. Based on this output and performance gaps identified in field deployments, the AI algorithm was further tuned to reduce false-positives and to improve its performance at interpreting images of tests with staining in the result window. The updated model was run against the hold-out set and found to have a weighted F1 score of 98.9. Because the ML models had been trained on 2 mega-pixel images, 2-megapixels—an amount that exceeds the default resolution of most smart devices—was the recommended minimum resolution for optimal performance of the AI algorithm.

Data analysis

We assessed the performance of the AI algorithm at interpreting the images collected in the HIVST Performance Study on four aspects. First, to understand how well the AI algorithm correctly identified true positives, we calculated sensitivity: the percent of HIVST images classified as "positive" by the expert panel that were interpreted as "positive" by the AI algorithm. Second, to understand how well the AI algorithm correctly identified true negatives, we calculated specificity: the percent of HIVST images classified as "negative" by the expert panel that were interpreted as "negative" by the AI algorithm. Third, to understand how likely it was that the image was truly a negative result if the AI algorithm interpreted it as such, we calculated negative predictive value (NPV): the percent of HIVST images interpreted as "negative" by the AI algorithm that were classified as "negative" by the expert panel. Lastly, to understand how likely it was that the image was truly of a positive result if the AI algorithm interpreted it as such, we calculated positive predictive value (PPV): the percent of HIVST images interpreted as "positive" by the AI algorithm that were classified as "positive" by the expert panel. To assess uncertainty around each of these estimates, we calculated binomial 95% confidence intervals (CI).

To achieve our secondary objective of understanding how well the AI algorithm performed compared to pharmacy providers and clients, we calculated the above-described performance

TABLE 1 Characteristics of the pharmacy clients and providers.

Pharmacy client characteristic	$N = 854$
Female	540 (63%)
Age, median (IQR)	27 (22, 31)
Average monthly income (±SD), Kenyan Shillings (USD)	15,274 ± 25,109 (131.4 ± 215.9)
Income source	
Trade/sales	303 (36%)
Laborer/semi-skilled	221 (26%)
Professional	118 (14%)
Student	108 (13%)
No income	17 (2%)
Other	87 (10%)
Relationship status	
Has one primary partner	466 (55%)
Has casual sex partners only	150 (18%)
Has one primary partner and casual partners	185 (22%)
Had more than one new sexual partner in the last 3 months	278 (33%)
Sexual partner is living with HIV	13 (2%)
Sexual partner has other partner(s)	103 (12%)
Pharmacy provider characteristic	$N = 40$
Female	16 (40%)
Age, median (IQR)	31 (27, 37)
Own the pharmacy	17 (42%)
Years in profession, median (IQR)	6 (4, 10)
Days worked at the pharmacy per week, median (IQR)	6 (6, 7)
Have not provided HIV tests prior to study implementation	5 (12%)
Counsel clients before and/or during HIV testing[a]	32 (91%)
Minutes typically spent counseling a client for an HIV test, median (IQR)	20 (15, 43)

IQR, interquartile range; SD, standard deviation; USD, United States dollar. [a]Denominator out of 35 providers who answered this question.

metrics for pharmacy providers and clients. For each pair-wise comparison of interest—the AI algorithm vs. pharmacy providers, and the AI algorithm vs. pharmacy clients—we calculated two performance indices: (1) a sensitivity index, and (2) a specificity index. The former is the ratio of the sensitivity achieved by the AI algorithm to that achieved by the human group of interest; the latter is calculated in the same way except using the values for specificity achieved by the groups of interest. For both performance indices, a value >1 would indicate that the AI algorithm outperformed the human group in question (34). For all analyses, we used Stata v17.0 (StataCorp LLC, USA).

Ethics

The Kenya Medical Research Institute's Scientific Ethics Review Unit and the Institutional Review Board of the Fred Hutchinson Cancer Center reviewed and approved all study procedures. All pharmacy clients and providers completed written informed consent, which was available in English, Dholuo, and Kiswahili, and were compensated 500 Kenyan Shillings (∼$5 US dollars [USD]) for their time completing study activities (e.g., research surveys). The owners of participating pharmacies received 15,000 KES (∼$129 USD) per month for use of their space and utilities. Both the expert panel and AI algorithm received only the HIVST image file and were blinded to (i.e., were not provided with) additional contextual information (e.g., client demographics; pharmacy location). To reduce the risk of harm related to HIVST misinterpretation by the AI algorithm, the algorithm's interpretation was not shared back with the client, pharmacy provider, or RA or used for clinical decision-making, with all post-test counseling and referrals to clinic-based HIV prevention or treatment services based on the standard-of-care HIV rapid diagnostic testing conducted by the certified HTS provider.

Results

Participants

From March to June 2022, we screened 1,691 pharmacy clients and enrolled 1,500, all of whom completed HIVST. Of 1500 HIVST images uploaded to CommCare, 854 (57%) were of sufficient quality to be interpreted by the AI algorithm (The remaining images had been collected early in the study, prior to adjusting an image resolution setting in CommCare, which compressed the images prior to saving them and reduced their resolution to below the 2-megapixel minimum required resolution to be interpretable by the AI algorithm). Among the pharmacy clients with interpretable HIVST images, 63% (540/854) were female, the median age was 26 years (IQR 22–31), and 39% (335/854) reported having casual sex partners in the past 6 months (Table 1). The majority (75%, 640/854) of clients opted to complete HIVST with some form of assistance from the pharmacy provider. Among 40 pharmacy providers who delivered HIVST and completed a survey, 40% (16/40) were female, the median age was 31 years (IQR 27–31), and 43% (17/40) owned the pharmacy they worked in. The median duration of practice among pharmacy providers was 6 years (IQR 4–10).

Breakdown of HIVST interpretations

Figure 2A shows the proportion of HIVSTs interpreted as negative, positive, and indeterminant by the expert panel, the AI algorithm, pharmacy providers, and pharmacy clients. The expert panel—the "ground truth" for this analysis—classified 95% (808/854) of the images as negative, 5% (44/854) as positive, and 0.2% (2/854) as indeterminant. Overall, this breakdown was similar for the AI algorithm, pharmacy clients, and pharmacy providers. The four performance measures we assessed for each group are listed in Table 2 and described in detail below.

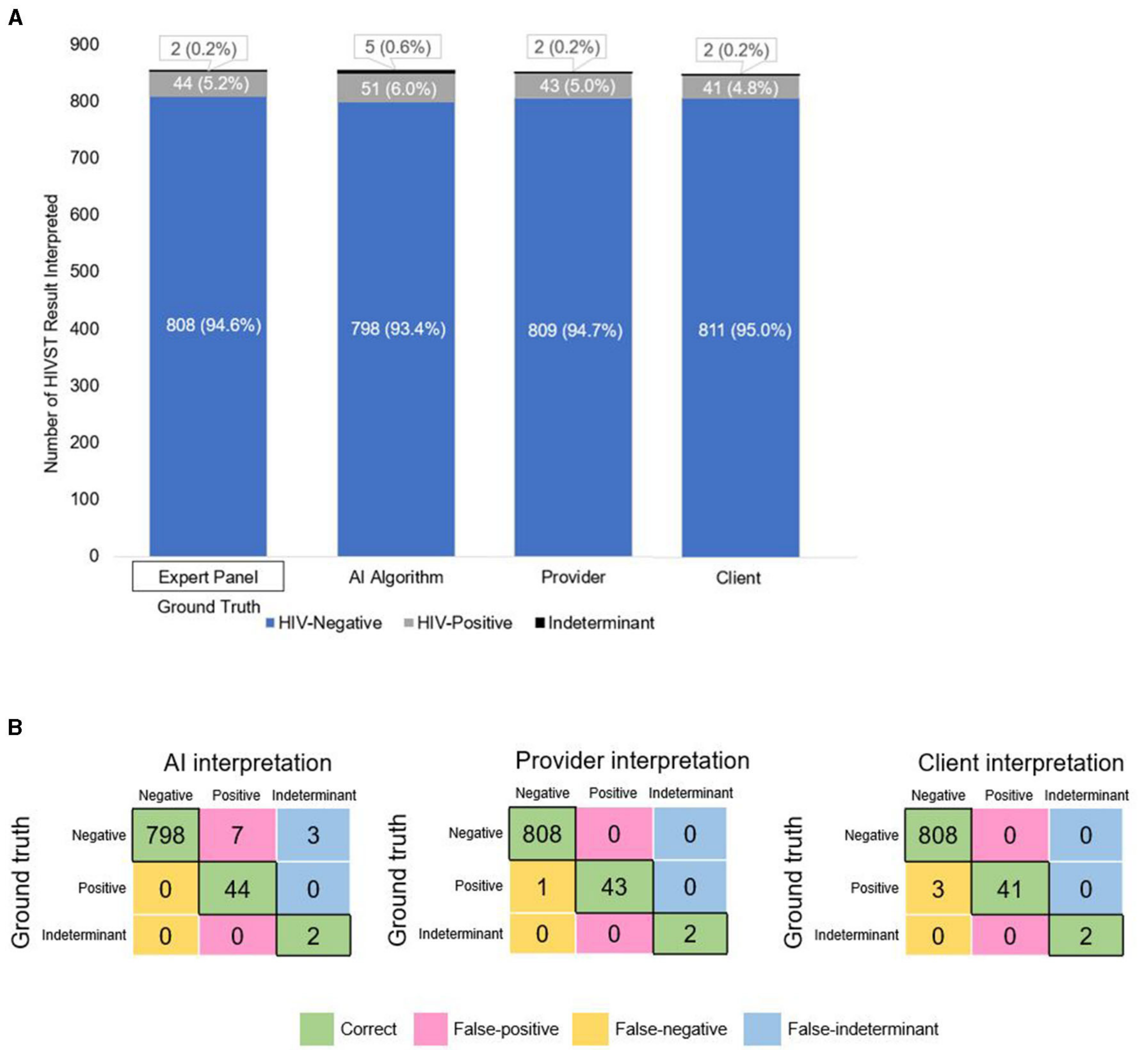


FIGURE 2
(A) Breakdown of HIVST interpretations by expert panel (ground truth), the AI algorithm, pharmacy providers, and pharmacy clients. **(B)** Confusion matrices showing the number of correct, false-positive, false-negative, and false-indeterminant interpretations given by the AI algorithm, pharmacy providers, and pharmacy clients.

Specificity and PPV

Correctly identifying HIV-negative individuals as HIV-negative—and minimizing false-positives—saves clients the unnecessary stress and burden of receiving a (false) positive test result and undergoing confirmatory HIV testing. In Figure 2B, the pink- shaded cells of each confusion matrix show the number of false-positive interpretations that each group gave. Focusing on each matrix's top row—which represents the images of HIV-negative tests—we see that the AI algorithm misclassified 10 negative tests-—7 as positive and 3 as indeterminant—and correctly classified 798 of the 808 negative tests as negative, giving it a specificity of 98.8% (95% CI: 98.0%, 99.5%). Because of the 7 false-positive interpretations, the AI algorithm's PPV was 81.5% (95% CI: 71.1%, 91.8%), meaning that if the AI algorithm interpreted a test as positive, there was an ~82% likelihood that the test was truly positive (and, by extension, an ~18% likelihood that the AI algorithm's positive interpretation was incorrect).

By comparison, pharmacy providers and clients correctly classified all 808 negative tests as negative (100% specificity). Neither of these human groups had any false-positives (100% PPV), meaning that if a pharmacy provider or client interpreted a test as positive, there was a 100% likelihood that the test was truly positive.

For both pairwise comparisons of interest—AI algorithm vs. pharmacy provider and AI algorithm vs. pharmacy client—the specificity indices are slightly <1 (Figure 3, orange bar), thus indicating that, compared to both human groups, the AI algorithm was slightly less effective at reading negative tests.

Sensitivity and NPV

Correctly identifying individuals living with HIV as HIV-positive—and minimizing false-negatives—is arguably even more important, as clients living with HIV who mistakenly believe themselves to be HIV-negative may inadvertently transmit HIV onward to others, experience delayed ART initiation, and/or be inappropriately initiated on PrEP (which has potential implications for developing HIV drug resistance). In Figure 2B, the yellow-shaded cells of each confusion matrix show the number of false-negative interpretations that each group gave. Focusing on each matrix's middle row—which represents the images of HIV-positive tests—we see that the AI algorithm correctly classified all 44 positive tests as positive (100% sensitivity). Because the AI algorithm did not misclassify any positive or indeterminant tests as negative (i.e., no false-negatives), its NPV was 100%, meaning that if the algorithm interpreted a test as negative, there was a 100% likelihood that the test was truly negative.

By comparison, pharmacy providers and clients missed a few positive results, correctly classifying 43 and 41, respectively, of the 44 positive tests as positive; thus, providers and clients had slightly lower sensitivity: 97.3% (95% CI: 96.7%, 98.7%) and 93.2% (95% CI: 91.5%, 94.9%), respectively. Because providers misclassified 1 positive test as negative (i.e., 1 false-negative) and clients misclassified 3 positive tests as negative (i.e., 3 false-negatives), their NPVs were 99.9% (95% CI: 99.6%, 100%) and 99.9% (95% CI 99.2%, 100%), respectively. In other words, if a pharmacy provider or client interpreted a test as negative, there was a 99.9% likelihood that the test was truly negative. Both sensitivity indices are slightly >1 (Figure 3, blue bar), thus indicating that the AI algorithm was slightly more effective at reading positive results than both the pharmacy provider and the pharmacy client groups.

Discussion

In this study, an AI algorithm trained to interpret images of HIVST results demonstrated perfect sensitivity and negative predictive value (each 100%), 99% specificity, and 86% positive predictive value. Our findings are notable for two main reasons. First, our evaluation found that the AI algorithm did not miss a single HIV infection and, importantly, outperformed humans at correctly identifying positive tests as positive. This finding is significant because false-negatives—i.e., missed diagnoses of

TABLE 2 Performance of an AI algorithm, pharmacy clients, and pharmacy providers at interpreting HIVST results, compared to an expert panel ($n = 854$ HIVST result images).

Metric[a]	AI algorithm interpretation, % (95% CI)	Pharmacy client interpretation, % (95% CI)	Pharmacy provider interpretation, % (95% CI)
Sensitivity	100%	93.2% (91.5%, 94.9%)	97.7% (96.7%, 98.7%)
Specificity	98.8% (98.0%, 99.5%)	100%	100%
PPV	81.5% (71.1%, 91.8%)	100%	100%
NPV	100%	99.6% (99.2%, 100%)	99.9% (99.6%, 100%)

AI, artificial intelligence; HIVST, HIV self-testing; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value. [a]The reference group for all metrics was the expert panel consensus. Sensitivity and PPV are out of a denominator of 44 HIVST images classified as "positive" by the expert panel. Specificity and NPV are out of a denominator of 808 HIVST images classified as "negative" by the expert panel.

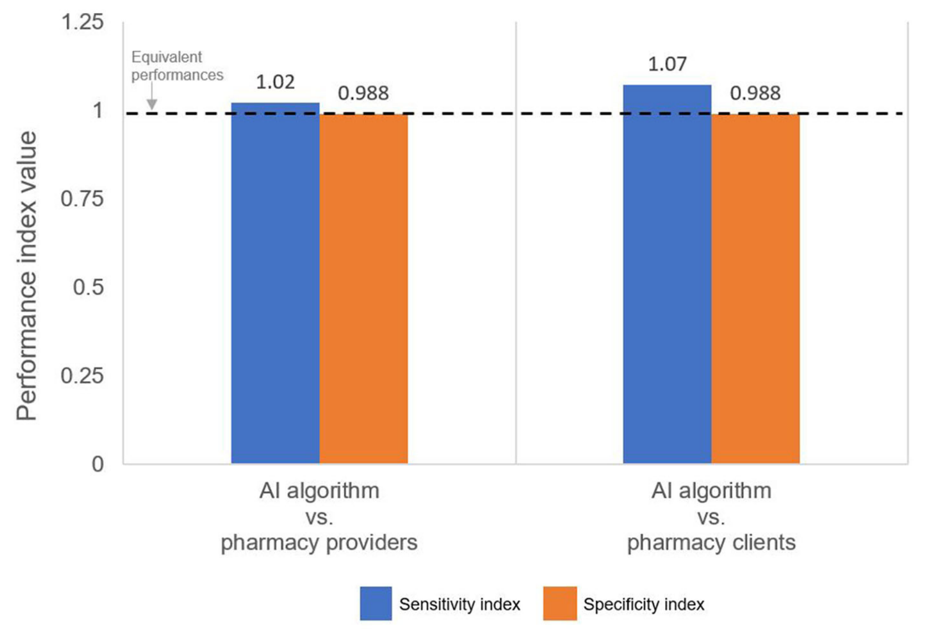


FIGURE 3
Performance indices comparing the sensitivity and specificity of the AI algorithm to the pharmacy provider and pharmacy client groups. An index >1 indicates that the AI algorithm performed better than the indicated human group.

HIV infection—are detrimental to the affected individuals and undermine efforts to control the HIV epidemic. Clients who receive false-negative results may unknowingly transmit their HIV infection onward to sex partners, and potentially lose precious time in initiating HIV treatment critical to slowing the progression of the HIV virus (45). Such clients may also be incorrectly initiated on antiretroviral-based HIV prevention drugs (e.g., PrEP)—a scenario that could lead to drug resistance (22) and make the client's HIV infection more difficult and more expensive to treat (46).

Many high HIV burden countries are striving to reach the UNAIDS target of 95% of PLHIV knowing their status (4); however, as countries approach this target, the more difficult and more labor-intensive it becomes to find the remaining, often hard-to-reach individuals (47). Because countries spend millions of dollars each year on HIV case-finding (48), misinterpretation of any such cases is an expensive failure (49, 50). Human error in interpreting positive HIV test results—particularly ones with weak positive lines—is a known and well-documented issue, and many of the proposed interventions to mitigate this issue (e.g., more specialized training for healthcare providers, use of a second test reader) require additional human resources (13, 51). In our study, the AI algorithm detected 4 HIV infections missed by pharmacy providers and pharmacy clients, possibly due to the AI's superior ability to detect faint lines compared to the unassisted human eye. While 4 additional cases of HIV infection detected may appear modest, HIV testing occurs at large scale globally, with over 100 million HIV tests conducted annually worldwide (52). Given this, our study findings give reason for cautious optimism that such AI computer vision technology could potentially help countries improve their HIV case-finding without requiring them to invest significant additional human resources, with the caveat that the exact value such technology provides will necessarily depend on numerous factors, including how commonly used the AI algorithm's specific brand of HIV test is; the existing rate of interpretation error among HIV test users; and the extent to which the test images being fed to the AI algorithm differ from the set on which it was trained (i.e., AI bias, discussed in more detail below).

Our second notable finding is that when the AI algorithm erred, it erred in the more conservative direction, producing some false-positive—but zero false-negative—interpretations. For any diagnostic test, perfect accuracy is virtually impossible to achieve due to factors such as variation in specimen quality; as such, there is a necessary trade-off between sensitivity and specificity (53). Recognizing this, global- and country-level clinical protocols for diagnosing HIV infection (12, 54) first utilize more sensitive HIV tests as first-line assays to weed out most HIV-negative individuals, followed by more specific HIV tests as second- and third-line assays to eliminate false-positives among those who tested positive (55). In our study, the AI algorithm gave seven false-positive interpretations, most likely due to image quality issues (e.g., blurriness), and was outperformed on this metric by the pharmacy client and provider groups, both of which gave zero false-positive interpretations. False-positive interpretations are, indeed, unfavorable, and can lead to adverse consequences for clients, such as anxiety and psychological distress (56); but from a public health perspective, these harms are considered less alarming than those of false-negatives, in part, because of the aforementioned three-test clinical protocols in place that nearly always catch false-positives on second- or third-line assays (57). In short, in the absence of a tool that can interpret HIV tests with perfect sensitivity and perfect specificity, the next-best scenario is, arguably, a tool with perfect sensitivity, deployed with contingency plans for handling its imperfect specificity (e.g., clear warnings to end-users that positive interpretations could be incorrect and should be confirmed via additional testing). For this reason, our study findings give reason for cautious optimism about AI computer vision technology and its potential to support HIV test interpretation. Below we discuss future areas of research; key ethical, financial, and legal considerations for this technology; and study limitations.

Future areas of research

Whether any HIV program chooses to incorporate AI computer vision technology into its service delivery and, if so, how will depend on a number of context-specific factors, such as how well the algorithm in question performs; how HIV service delivery is structured, staffed, and funded; internet connectivity; local laws regulating the use of AI; and the technology's acceptability to and usability by target end-users. With this caveat in mind, there are a number of potential use cases for AI computer vision technology within HIV service delivery that are worthy of further consideration—and additional research—by HIV stakeholders working in collaboration with AI experts and technology ethicists. These research areas—summarized in Table 3 as Examples A through F and discussed in further detail in the Supplementary material—touch upon potential uses of AI computer vision technology to support the following aspects of HIV service delivery: quality assurance and supported implementation; commodity accountability; and provider training and evaluation. The use cases focus specifically on models of differentiated HIV service delivery (DSD), which—as previously explained—often move service delivery outside of traditional health facilities, task-shift delivery to new cadres of providers, and/or incorporate new technologies or innovations (5).

Ethical, financial, and legal considerations

How to ethically implement AI computer vision technology and sustainably finance its development and long-term use at scale while still ensuring that it complies with local laws, are three fast-evolving areas that will determine whether, how, and the pace at which this technology is incorporated into HIV service delivery. Regulations and guidelines around the use of AI for health are still largely nascent (58, 59). Similar to other health services, AI-supported delivery of HIV services will require guardrails not only to ensure that the technology complies with patient privacy and data security laws, but also to handle situations in which the AI returns inaccurate information, especially if inaccuracies are more prevalent for certain groups of individuals due to bias in the data on which the AI was trained. The growing literature on bias in AI machine learning for medicine highlights the importance of and need for bias mitigation strategies and for any deployment of AI to be accompanied by routine evaluations

 10.3389/fpubh.2024.1334881

TABLE 3 Future areas of research related to AI computer vision in HIV differentiated service delivery (DSD).

Potential uses of AI		Relevance for HIV differentiated service delivery (DSD)
Area 1: Quality assurance and supported implementation		
Help verify test results	**Example A:** After conducting an HIV test (or self-test), providers (or clients) upload a photo of the result to a digital platform. On the back end, an AI algorithm also interprets the result and notifies parties responsible for overseeing quality of care of any discrepancies, creating an opportunity to change course of action, if needed.	Having a reliable way to verify that HIV tests are correctly interpreted might convince policymakers to allow delivery models that task-shift HIV testing to new cadres of healthcare providers, lay providers, or clients themselves as part of broader DSD efforts to make more efficient use of existing health resources, lower client wait times, reduce visit burden, and/or expand HIV services to new delivery venues.
Guide and reassure end-users	**Example B:** Healthcare providers are given the option to compare their interpretation of the HIV test result to that of the AI algorithm.	Because many DSD models engage individuals not previously involved in HIV service delivery, their success hinges on developing end-user knowledge, skills, and self-efficacy to engage in services as intended (e.g., to correctly and confidently conduct HIV tests). AI algorithms could support providers by offering a "second opinion" on result interpretation, giving providers the opportunity to double-check their interpretation of the test prior to making a final determination about the result.
Assess fidelity	**Example C:** An AI algorithm is trained to detect possible indications of HIV self-test (HIVST) misadministration—such as the blood sample being placed incorrectly (e.g., in the results window) and test end-users not waiting the recommended duration of time prior to interpreting the result—and flag such cases for further review.	Mechanisms for assuring the quality of HIVST administration and interpretation may make policymakers more willing to support HIV service delivery models that use HIVST in lieu of, or as an additional testing option to, HIV rapid diagnostic testing (RDT). This would benefit clients who prefer HIVST and create potential opportunities to move HIV service delivery, or select parts of it, outside of traditional healthcare settings (See telehealth PrEP model in next row).
Provide quality control	**Example D:** An AI algorithm checks the quality of an HIV test result photo uploaded to a digital platform. If a photo does not meet a prespecified quality standard, the user is prompted in real-time to re-take it (while the results are still valid).	Telehealth PrEP delivery models involve remote clinicians using digital data for clinical decision-making. Ensuring that user-submitted digital data is of sufficient quality to be usable could potentially help avoid incorrect clinical decisions and reduce inefficiencies (e.g., time delays) related to data re-collection.
Area 2: Provider training and evaluation		
Evaluate providers	**Example E:** After an individual completes HTS training, an AI algorithm is run on images from the first 100 HIV tests they conduct, with any discrepancies with the algorithm's interpretation flagged for review.	Compared to using human auditors, assessing HTS provider performance remotely and in an unannounced fashion may be more reliable, cost-saving, have better privacy for clients, and enable regulators to quickly identify providers in need of further training and support.
Area 3: Commodity accountability		
Mitigate fraud	**Example F**: A country's ministry of health or a donor agency agrees to support HIV service delivery in a private-sector setting by providing commodities (e.g., HIV test kits, PrEP drugs). For accountability purposes, providers are required to write unique client identifiers on used commodities and upload photos of them. An AI algorithm is trained to detect potential signs of fraud, such as upload of duplicate photos and erasure of the unique client identifiers (to re-use test kits).	To date, few countries with high HIV burden have partnered with the private sector to deliver HIV services at scale, in part, because the necessary legislation and systems for cross-sector service delivery (e.g., health information system, supply chain) are not yet in place. Used as a fraud mitigation strategy, AI computer vision could potentially provide an extra measure of accountability that might address policymaker and donor hesitation to support private sector-based HIV DSD models.

AI, artificial intelligence; DSD, differentiated service delivery; HIVST, HIV self-testing; RDT, HIV rapid diagnostic testing; HTS, HIV testing services; PrEP, pre-exposure prophylaxis; RDT, rapid diagnostic testing; HTS, HIV testing services.

of its performance and impact on patients and providers (60–63). Thinking through the potential ethical issues (59) (e.g., inequitable access to AI technologies exacerbating existing health disparities) and liability risks (e.g., the risk of client self-harm after receiving a false-positive result from an AI algorithm)—and deciding on risk mitigation measures (e.g., limiting client exposure to AI; deciding the content of user agreements for HIVST apps)—will influence this technology's incorporation into HIV service delivery, which will undoubtedly vary by setting and use case.

Similarly, financing AI computer vision algorithms is bound to take many forms. There are three primary costs to consider: (1) the development of the algorithm (a one-time cost); (2) the integration of the algorithm into a digital platform, such as an app or electronic medical record system (also a one-time cost); and (3) the operating costs of running and maintaining the algorithm (an ongoing cost). The development of the algorithm assessed in this study was funded by a private philanthropic organization as a global public good. As such, the governments of low- and middle-income countries can obtain the algorithm from the parent company free of charge. These governments, in turn, are responsible for covering the one-time cost (typically ~$10,000 USD) of integrating the algorithm into their digital health platform of choice. Lastly, governments and/or donors need to budget for the ongoing operating costs, the amount of which would depend on the scale at which the algorithm is used. For example, early data from pilot studies suggests that, when deployed at scale, the cost of running the algorithm could be kept as low as a few cents per test image assessed. Moreover, if an algorithm is deployed in multiple settings (e.g., multiple countries), then the cost for routine algorithm maintenance could be shared, with updates pushed to all end-users. Specific approaches to cost-sharing the development and maintenance of AI computer technology among governments, donors, third-party payers (e.g., private health insurers), and clients is an area for further investigation.

Study limitations

This study has limitations. First, because this study was conducted within a larger study on HIVST performance, pharmacy providers received comprehensive training on conducting and interpreting HIVST and pharmacy clients were given the option to receive provider assistance conducting HIVST. These factors may have increased the performance of pharmacy clients and providers, thus underestimating the degree to which an AI algorithm might outperform these human groups at HIVST interpretation. Second, due to the previously described error on behalf of the research team early on during data collection whereby an image resolution setting was not adjusted in the electronic data collection platform, the first 646 HIVST images collected during the study did not meet the 2-megapixel minimum required resolution prespecified by the algorithm developer; as such, 646 of the 1,500 total images collected (43%) were excluded from this evaluation. Although our remaining sample size ($n = 854$) was still robust, this may have given the provider group a slight advantage if their performance interpreting the tests whose images were discarded (collected early on in implementation when providers may have still been honing their test interpretation skills) was lower than it was for the tests included in our final analysis. Third, like all observational studies that do not employ probabilistic sampling, our pharmacy provider and client groups—and their respective performances at interpreting HIVSTs—may not be representative of all pharmacy providers and clients; as such, our findings about the performance of those two human groups are not generalizable to other pharmacy providers and clients in Kisumu County or, more broadly, Kenya and similar settings. Lastly, because our study assessed only one AI algorithm trained on a single brand of HIVST kits and only on images of at least 2-megapixel resolution, our findings are not necessarily generalizable to other AI algorithms, types of HIVSTs, or to images below 2-megapixel resolution. However, because the Mylan HIVSTs used in this study have the same general format of many other common biologic tests (e.g., positive results are indicated by two lines), our study findings may be a reasonable indicator of how this technology might perform on other similar tests.

Conclusions

AI computer vision technology shows promise as a quality assurance tool for HIV testing. Such technology may be especially useful for enabling HIV services to be delivered outside of traditional healthcare settings, by new cadres of providers, and/or at different cadences to better meet client needs and preferences and to use existing health resources more efficiently. Future research could measure the effect size of making AI algorithm result interpretations available to end-users in real-time compared to a control group unassisted by AI (e.g., effect on rate of false-negatives, on provider confidence); assess the feasibility, acceptability, and unintended consequences of using this technology (e.g., ethical issues related to AI interpretation errors, patient privacy, and healthcare worker job security)—among different end-user groups (e.g., community health workers) and in different settings; and conduct cost-effectiveness studies (e.g., quantify the cost per additional case of HIV identified). Future research should also explore AI biases, with an eye toward minimizing biases that may compromise care quality, fairness, and equity. Stakeholders of HIV service delivery should carefully consider leveraging AI computer vision technology as part of broader efforts to make services more client-centered and expedite progress toward HIV epidemic control.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the IRB of Fred Hutchinson Cancer Center and the Scientific Ethics Review Unit of the Kenya Medical Research Institute. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SR: Formal analysis, Writing—original draft, Writing—review & editing. OE: Formal analysis, Writing—review & editing, Writing—original draft. RM: Software, Project administration, Writing—review & editing. BK: Supervision, Investigation, Project administration, Writing—review & editing. VO: Investigation, Project administration, Supervision, Writing—review & editing. SZ: Data curation, Formal analysis, Writing—review & editing. PO: Resources, Supervision, Project administration, Writing—review & editing. DH: Software, Writing—review & editing. SS: Software, Writing—review & editing. SM: Conceptualization, Software, Project administration, Writing—review & editing. DW: Resources, Supervision, Project administration, Writing—review & editing. DR: Software, Conceptualization, Funding acquisition, Writing—review & editing. EB: Methodology, Supervision, Conceptualization, Funding acquisition, Writing—review & editing. KO: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Bill & Melinda Gates Foundation (INV-033052, INV-007492; INV-038498). KO also received funding from the National Institute of Mental Health (R00 MH121166).

Acknowledgments

We express our gratitude to the pharmacy clients and providers who participated in this study and took the time and effort to contribute to the research. We also acknowledge the research assistants who collected the data for the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1334881/full#supplementary-material

References

1. Godfrey-Faussett P, Frescura L, Abdool Karim Q, Clayton M, Ghys PD. HIV prevention for the next decade: Appropriate, person-centred, prioritised, effective, combination prevention. *PLoS Med.* (2022) 19:e1004102. doi: 10.1371/journal.pmed.1004102

2. UNAIDS. *Global HIV and AIDS Statistics Fact Sheet.* (2023). Available online at: https://www.unaids.org/en/resources/fact-sheet (accessed October 5, 2023).

3. AVAC. *The Global PrEP Tracker.* (2023). Available online at: https://data.prepwatch.org/ (accessed October 25, 2023).

4. UNAIDS. *Ending Inequalities and Getting on Tract to End AIDS by 2030.* (2022). Available online at: https://www.unaids.org/sites/default/files/media_asset/2021-political-declaration_summary-10-targets_en.pdf (accessed October 5, 2023).

5. IAS. *Differentiated Service Delivery.* (2023). Available online at: https://www.differentiatedservicedelivery.org/ (accessed October 5, 2023).

6. Baeten JM, Heffron R, Kidoguchi L, Mugo NR, Katabira E, Bukusi EA, et al. Integrated delivery of antiretroviral treatment and pre-exposure prophylaxis to HIV-1–serodiscordant couples: a prospective implementation study in Kenya and Uganda. *PLoS Med.* (2016) 13:e1002099. doi: 10.1371/journal.pmed.1002099

7. Mack N, Odhiambo J, Wong CM, Agot K. Barriers and facilitators to pre-exposure prophylaxis (PrEP) eligibility screening and ongoing HIV testing among target populations in Bondo and Rarieda, Kenya: results of a consultation with community stakeholders. *BMC Health Serv Res.* (2014) 14:231. doi: 10.1186/1472-6963-14-231

8. Roche SD, Wairimu N, Mogere P, Kamolloh K, Odoyo J, Kwena ZA, et al. Acceptability and feasibility of pharmacy-based delivery of pre-exposure prophylaxis in Kenya: a qualitative study of client and provider perspectives. *AIDS Behav.* (2021) 25:3871–82. doi: 10.1007/s10461-021-03229-5

9. Ortblad KF, Mogere P, Roche S, Kamolloh K, Odoyo J, Irungu E, et al. Design of a care pathway for pharmacy-based PrEP delivery in Kenya: results from a collaborative stakeholder consultation. *BMC Health Serv Res.* (2020) 20:1034. doi: 10.1186/s12913-020-05898-9

10. Celum C, Seidman D, Travill D, Dehlendorf C, Gumede S, Zewdie K, et al. A decision support tool has similar high PrEP uptake and increases early PrEP persistence in adolescent girls and young women in South Africa: results from a randomized controlled trial. *J Int AIDS Soc.* (2023) 26:e26154. doi: 10.1002/jia2.26154

11. Haberer JE, Musinguzi N, Tsai AC, Boum Y, Bwana BM, Muzoora C, et al. Real-time electronic adherence monitoring plus follow-up improves adherence compared with standard electronic adherence monitoring. *AIDS.* (2017) 31:169–71. doi: 10.1097/QAD.0000000000001310

12. WHO. *Consolidated Guidelines on HIV Prevention, Testing, Treatment, Service Delivery and Monitoring: Recommendations for a Public Health Approach.* (2022). Available online at: https://www.who.int/publications/i/item/9789240031593 (accessed August 25, 2023).

13. Johnson CC, Kennedy C, Fonner V, Siegfried N, Figueroa C, Dalal S, et al. Examining the effects of HIV self-testing compared to standard HIV testing services: a systematic review and meta-analysis. *J Int AIDS Soc.* (2017) 20:21594. doi: 10.7448/IAS.20.1.21594

14. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al. Uptake, accuracy, safety, and linkage into care over two years of promoting annual self-testing for HIV in Blantyre, Malawi: a community-based prospective study. *PLoS Med.* (2015) 12:e1001873. doi: 10.1371/journal.pmed.1001873

15. Choko AT, Desmond N, Webb EL, Chavula K, Napierala-Mavedzenge S, Gaydos CA, et al. The uptake and accuracy of oral kits for HIV self-testing in high HIV prevalence setting: a cross-sectional feasibility study in Blantyre, Malawi. *PLoS Med.* (2011) 8:e1001102. doi: 10.1371/journal.pmed.1001102

16. Martínez Pérez G, Steele SJ, Govender I, Arellano G, Mkwamba A, Hadebe M, et al. Supervised oral HIV self-testing is accurate in rural Kwa Zulu-Natal, South Africa. *Trop Med Int Health.* (2016) 21:759–67. doi: 10.1111/tmi.12703

17. Asiimwe S, Oloya J, Song X, Whalen CC. Accuracy of un-supervised versus provider-supervised self-administered HIV testing in Uganda: a randomized implementation trial. *AIDS Behav.* (2014) 18:2477–84. doi: 10.1007/s10461-014-0765-4

18. Kelvin EA, Akasreku B. The Evidence for HIV self-testing to increase HIV testing rates and the implementation challenges that remain. *Curr HIV/AIDS Rep.* (2020) 17:281–9. doi: 10.1007/s11904-020-00504-3

19. Dovel K, Shaba F, Offorjebe OA, Balakasi K, Nyirenda M, Phiri K, et al. Effect of facility-based HIV self-testing on uptake of testing among outpatients in Malawi: a cluster-randomised trial. *The Lancet Global Health.* (2020) 8:e276–e87. doi: 10.1016/S2214-109X(19)30534-0

20. WHO. *New US $1 Price for HIV Self-Tests.* (2022). Available online at: https://www.who.int/news/item/27-07-2022-new-1-dollar-price-for-hiv-self-tests (accessed February 3, 2023).

21. WHO. *Differentiated and Simplified Pre-Exposure Prophylaxis for HIV Prevention: Update to WHO Implementation Guidance.* (2022). Available online at: https://iris.who.int/bitstream/handle/10665/360861/9789240053694-eng.pdf?sequence=1 (accessed August 2, 2023).

22. Gibas KM, van den Berg P, Powell VE, Krakower DS. Drug resistance during HIV pre-exposure prophylaxis. *Drugs.* (2019) 79:609–19. doi: 10.1007/s40265-019-01108-x

23. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health.* (2019) 1:e271–e97. doi: 10.1016/S2589-7500(19)30123-2

24. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7

25. LFD AI Consortium. Machine learning for determining lateral flow device results for testing of SARS-CoV-2 infection in asymptomatic populations. *Cell Rep Med.* (2022) 3:100784. doi: 10.1016/j.xcrm.2022.100784

26. Tong H, Cao C, You M, Han S, Liu Z, Xiao Y, et al. Artificial intelligence-assisted colorimetric lateral flow immunoassay for sensitive and quantitative detection of COVID-19 neutralizing antibody. *Biosens Bioelectron.* (2022) 213:114449. doi: 10.1016/j.bios.2022.114449

27. Wong NCK, Meshkinfamfard S, Turbe V, Whitaker M, Moshe M, Bardanzellu A, et al. Machine learning to support visual auditing of home-based lateral flow immunoassay self-test results for SARS-CoV-2 antibodies. *Commun Med.* (2022) 2:78. doi: 10.1038/s43856-022-00146-z

28. Bermejo-Pelaez D, Marcos-Mencia D, Alamo E, Perez-Panizo N, Mousa A, Dacal E, et al. A smartphone-based platform assisted by artificial intelligence for reading and reporting rapid diagnostic tests: evaluation study in SARS-CoV-2 lateral flow immunoassays. *JMIR Public Health Surveill.* (2022) 8:e38533. doi: 10.2196/38533

29. Arumugam S, Ma J, Macar U, Han G, McAulay K, Ingram D, et al. Rapidly adaptable automated interpretation of point-of-care COVID-19 diagnostics. *Commun Med.* (2023) 3:91. doi: 10.1038/s43856-023-00312-x

30. Richardson S, Kohn MA, Bollyky J, Parsonnet J. Validity of at-home rapid antigen lateral flow assay and artificial intelligence read to detect SARS-CoV-2. *Diagn Microbiol Infect Dis.* (2022) 104:115763. doi: 10.1016/j.diagmicrobio.2022.115763

31. Guo X, Khalid MA, Domingos I, Michala AL, Adriko M, Rowel C. Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security. *Nat Electr.* (2021) 4:1–12 doi: 10.1038/s41928-021-00612-x

32. Lee S, Yoo YK, Wee KW, Kim C, Lee NE, Kim KH, et al. Machine-learning-assisted lateral flow assay for COVID-19 and influenza detection. *SSRN J.* (2022) 26:4073623. doi: 10.2139/ssrn.4073623

33. Bermejo-Pelaez D, Medina N, Alamo E, Soto-Debran JC, Bonilla O, Luengo-Oroz M, et al. Digital platform for automatic qualitative and quantitative reading of a cryptococcal antigen point-of-care assay leveraging smartphones and artificial intelligence. *J Fungi.* (2023) 9:217. doi: 10.3390/jof9020217

34. Turbé V, Herbst C, Mngomezulu T, Meshkinfamfard S, Dlamini N, Mhlongo T, et al. Deep learning of HIV field-based rapid tests. *Nat Med.* (2021) 27:1165–70. doi: 10.1038/s41591-021-01384-9

35. Omidfar K, Riahi F, Kashanian S. Lateral flow assay: a summary of recent progress for improving assay performance. *Biosensors.* (2023) 13:837. doi: 10.3390/bios13090837

36. National AIDS & STI Control Programme (NASCOP). *Framework for the Implementation of Pre-exposure Prophylaxis of HIV in Kenya.* Nairobi: National AIDS & STI Control Programme (NASCOP). (2017). Available online at: https://www.prepwatch.org/wp-content/uploads/2018/12/Kenya_PrEP_Implementation_Framework.pdf (accessed October 10, 2023).

37. Ortblad KF, Kwach B, Zhang S, Asewe M, Ongwen PA, Malen RC, et al. Measuring the performance of HIV self-testing at private pharmacies in Kenya: a cross-sectional study. *J Int AIDS Soc.* (2023) 26:e26177. doi: 10.1002/jia2.26177

38. NASCOP. *Preliminary Kenya Population-Based HIV Impact Assessment (KENPHIA) 2018 Report.* (2018). Available online at: https://phia.icap.columbia.edu/wp-content/uploads/2020/04/KENPHIA-2018_Preliminary-Report_final-web.pdf (Accessed August 30, 2023).

39. Audere. Audere (2023). Available online at: https://www.auderenow.org/ (accessed December 21, 2023).

40. Mugwanya KK, Pintye J, Kinuthia J, Abuna F, Lagat H, Begnel ER, et al. Integrating preexposure prophylaxis delivery in routine family planning clinics: a feasibility programmatic evaluation in Kenya. *PLoS Med.* (2019) 16:e1002885. doi: 10.1371/journal.pmed.1002885

41. Indivillage. Our Services. (2023). Available online at: https://indivillage.com/ (accessed December 22, 2023).

42. Gupta K RY, Ibrahim A, Mendonca R, Cooper S, Morris S, Hattery D. *Transforming Rapid Diagnostic Tests into Trusted Diagnostic Tools in LMIC using AI (abstract). 2023 IEEE Conference on Artificial Intelligence 2023. 5-6 June 2023; Santa Clara, California.* Available online at: https://www.computer.org/csdl/proceedings-article/cai/2023/398400a306/1PhCA3Pszew (accessed June 6, 2023).

43. Google. Google for Developers. *HealthPulse AI Leverages MediaPipe to Increase Health Equity.* (2024). Available online at: https://developers.googleblog.com/2024/01/healthpulse-ai-leverages-mediapipe-to-increase-health-equity.html (accessed January 16, 2024).

44. Audere. *HealthPulse.* (2022). Available online at: https://healthpulsenow.org/about-us (accessed December 14, 2022).

45. Johnson C, Fonner V, Sands A, Tsui S, Ford N, Wong V, et al. A Report on the Misdiagnosis of HIV Status. Consolidated Guidelines on HIV Testing Services: 5Cs: Consent, Confidentiality, Counselling, Correct Results and Connection 2015. Geneva: World Health Organization (2015).

46. Hayes R. *Second-Line Treatment Nearly Nine Times More Expensive Than First-Line in Upper-Middle Income Countries.* (2022). Available online at: https://www.aidsmap.com/news/aug-2022/second-line-treatment-nearly-nine-times-more-expensive-first-line-upper-middle-income (accessed January 10, 2024).

47. USAID. *Data and Advanced Analytics in HIV Service Delivery: Use Cases to Help Reach 95-95-95.* (2022). Available online at: https://www.usaid.gov/sites/default/files/2022-05/Data_and_advanced_analytics_in_HIV_service_delivery_Final.pdf (accessed December 21, 2023).

48. *The Foundation for AIDS Research. PEPFAR Country/Regional Operational Plans (COPs/ROPs) Database.* (2023). Available online at: https://copsdata.amfar.org/ (accessed January 10, 2024).

49. Gill M, Powell M, Vu Q, Krentz H. Economic impact on direct healthcare costs of missing opportunities for diagnosing HIV within healthcare settings. *HIV Med.* (2021) 22:723–31. doi: 10.1111/hiv.13121

50. Popping S, Versteegh L, Nichols BE, van de Vijver DA, van Sighem A, Reiss P, et al. Characteristics and short-and long-term direct medical costs among adults with timely and delayed presentation for HIV care in the Netherlands. *PLoS ONE.* (2023) 18:e0280877. doi: 10.1371/journal.pone.0280877

51. Tonen-Wolyec S, Sarassoro A, Muwonga Masidi J, Twite Banza E, Nsiku Dikumbwa G, Maseke Matondo DM, et al. Field evaluation of capillary blood and oral-fluid HIV self-tests in the democratic republic of the Congo. *PloS ONE.* (2020) 15:e0239607-e. doi: 10.1371/journal.pone.0239607

52. Government of Canada. *Canada and the Global Fund to Fight AIDS, Tuberculosis and Malaria.* (2023). Available online at: https://www.international.gc.ca/world-monde/international_relations-relations_internationales/multilateral-multilateraux/gfatm-fmstp.aspx?lang=eng (accessed January 5, 2024).

53. Centers for Disease Control. *Diagnostic Sensitivity and Specificity for Clinical Laboratory Testing.* (2022). Available online at: https://www.cdc.gov/labtraining/docs/job_aids/additional_resources/Sensitivity_and_Specificity_Final_5_23_2022_508.pdf%20 (accessed May 23, 2022).

54. National AIDS & STI Control Program, Kenya Ministry of Health. *Kenya HIV Testing Services Operational Manual 2022 Edition.* (2022). Available online at: https://www.prepwatch.org/wp-content/uploads/2023/04/Kenya-HTS-Manual.pdf (accessed December 21, 2023).

55. World Health Organization. *Consolidated Guidelines on HIV Testing Services: 5Cs: Consent, Confidentiality, Counselling, Correct Results and Connection 2015.* Geneva: World Health Organization (2015).

56. Bhattacharya R, Barton S, Catalan J. When good news is bad news: psychological impact of false positive diagnosis of HIV. *AIDS Care.* (2008) 20:560–4. doi: 10.1080/09540120701867206

57. WHO. *Information Note on Preventing HIV Misdiagnosis.* Geneva: WHO (2023).

58. WHO. *WHO Outlines Considerations for Regulation of Artificial Intelligence for Health.* Geneva: WHO (2023).

59. Chee FYMS. *EU Lawmakers Vote for Tougher AI Rules as Draft Moves to Final Stage.* (2023). Available online at: https://www.reuters.com/technology/eu-lawmakers-agree-changes-draft-artificial-intelligence-rules-2023-06-14/#:$\sim$:text=BRUSSELS%2FSTOCKHOLM%2C%20June%2014%20,to%20disclose%20AI%2Dgenerated%20content (accessed June 14, 2023).

60. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med.* (2021) 27:2176–82. doi: 10.1038/s41591-021-01595-0

61. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit Med.* (2023) 6:113. doi: 10.1038/s41746-023-00858-z

62. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med.* (2021) 1:25. doi: 10.1038/s43856-021-00028-w

63. Sourlos N, Wang J, Nagaraj Y, van Ooijen P, Vliegenthart R. Possible bias in supervised deep learning algorithms for CT lung nodule detection and classification. *Cancers.* (2022) 14:867. doi: 10.3390/cancers14163867

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 22 February 2024
DOI 10.3389/fpubh.2024.1337395

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Jay J. Park,
Stanford University, United States
Jernej Zavrsnik,
Health Center dr Adolf Drolc, Slovenia

*CORRESPONDENCE
Christopher Hartnick
Christopher_Hartnick@meei.harvard.edu

†These authors have contributed equally to this work and share senior authorship

RECEIVED 13 November 2023
ACCEPTED 22 January 2024
PUBLISHED 22 February 2024

CITATION
Benboujja F, Hartnick E, Zablah E, Hersh C, Callans K, Villamor P, Yager PH and Hartnick C (2024) Overcoming language barriers in pediatric care: a multilingual, AI-driven curriculum for global healthcare education. *Front. Public Health* 12:1337395. doi: 10.3389/fpubh.2024.1337395

Overcoming language barriers in pediatric care: a multilingual, AI-driven curriculum for global healthcare education

Fouzi Benboujja[1], Elizabeth Hartnick[2], Evelyn Zablah[1], Cheryl Hersh[3], Kevin Callans[1,4], Perla Villamor[5,6], Phoebe H. Yager[7†] and Christopher Hartnick[1*†]

[1]Department of Otolaryngology, Mass Eye and Ear, Harvard Medical School, Boston, MA, United States, [2]Brown University, Providence, RI, United States, [3]Pediatric Airway, Voice and Swallowing Center, Massachusetts General Hospital for Children, Boston, MA, United States, [4]Massachusetts General Hospital for Children, Boston, MA, United States, [5]Hospital Serena del Mar, Cartagena, Colombia, [6]Hospital Infantil Napoleón Franco Pareja, Cartagena, Colombia, [7]Pediatric Intensive Care Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Background: Online medical education often faces challenges related to communication and comprehension barriers, particularly when the instructional language differs from the healthcare providers' and caregivers' native languages. Our study addresses these challenges within pediatric healthcare by employing generative language models to produce a linguistically tailored, multilingual curriculum that covers the topics of team training, surgical procedures, perioperative care, patient journeys, and educational resources for healthcare providers and caregivers.

Methods: An interdisciplinary group formulated a video curriculum in English, addressing the nuanced challenges of pediatric healthcare. Subsequently, it was translated into Spanish, primarily emphasizing Latin American demographics, utilizing OpenAI's GPT-4. Videos were enriched with synthetic voice profiles of native speakers to uphold the consistency of the narrative.

Results: We created a collection of 45 multilingual video modules, each ranging from 3 to 8 min in length and covering essential topics such as teamwork, how to improve interpersonal communication, "How I Do It" surgical procedures, as well as focused topics in anesthesia, intensive care unit care, ward nursing, and transitions from hospital to home. Through AI-driven translation, this comprehensive collection ensures global accessibility and offers healthcare professionals and caregivers a linguistically inclusive resource for elevating standards of pediatric care worldwide.

Conclusion: This development of multilingual educational content marks a progressive step toward global standardization of pediatric care. By utilizing advanced language models for translation, we ensure that the curriculum is inclusive and accessible. This initiative aligns well with the World Health Organization's Digital Health Guidelines, advocating for digitally enabled healthcare education.

KEYWORDS

pediatric, global health, generative language models, digital health, education, artificial intelligence (AI), curriculum, healthcare

1 Introduction

In pediatric healthcare, each stage of the patient journey—from initial assessments to post-treatment convalescence to eventual safe discharge home—poses its own set of unique challenges. These complexities are universal, transcending geographical and cultural boundaries (1). The COVID-19 pandemic has further spotlighted the crucial impact of cultural and linguistic differences on various aspects of healthcare (2). From determining patient access (3) to affecting continuity of care (4), these elements have been empirically shown to significantly influence health outcomes (5–10), especially during critical health events such as global pandemics (11–13).

Building on the understanding of the role of cultural and linguistic elements, a particularly pressing challenge arises in the domain of healthcare education (1, 14). Caregivers, parents, and healthcare professionals often face communication and comprehension barriers when the predominant language of instruction is not their native language (15). This issue becomes especially acute in pediatric care, where accuracy and in-depth understanding are imperative.

Considering the well-documented gaps in communication within healthcare settings (15, 16) and their subsequent ramifications (4, 17), there is a clear and pressing need for multimedia resources that are not just comprehensive but also culturally and linguistically tailored. In addressing this need, CareWays Collaborative was founded in 2022 with a mission to partner with health teams around the globe to transform the culture and delivery of surgical care so that all patients can achieve the most favorable outcome. This group of healthcare professionals based in Boston and affiliated with Massachusetts Eye and Ear (MEE) and Massachusetts General Hospital (MGH), aims to create and disseminate educational materials that bridge communication gaps and distribute knowledge effectively across diverse healthcare settings. This effort aligns with the World Health Organization's Digital Health Guidelines, which underscores the importance of leveraging digital interventions for health worker training and education (18). In this vein, CareWays Collaborative created video-based educational content aided with artificial intelligence (AI) to ensure both clarity and cultural relevance for various healthcare provider and caregiver demographics.

AI is playing an increasing role in bridging gaps in healthcare education (19, 20). With the emergence of diverse techniques, from computer vision (21, 22) and data analytics (23) to Natural Language Processing (NLP) (24, 25), AI offers promising solutions to longstanding challenges related to language and cultural barriers. Notably, innovative tools such as Nvidia's Riva (26) and state-of-the-art generative language models—including OpenAI's ChatGPT 4 (27), Meta's LLaMA 2 (28), and Microsoft's PaLM 2 (29)—are enabling real-time translation. This advancement not only makes medical education more accessible but also emphasizes the essential role of AI in enhancing these resources.

Drawing on a decade of empirical research, a curriculum has been developed using artificial intelligence and voice cloning technologies to address communication gaps in pediatric healthcare. It offers insights into effective healthcare practices and highlights strategies to mitigate potential adverse events. In the form of an educational video series, the curriculum extends beyond recorded lectures or textbook material by including live demonstrations of important techniques and procedures used in operating rooms and at patients' bedsides. The videos focus on very subspecialized critical care and surgical techniques geared toward critical care nurses, physicians and respiratory therapists, pediatric anesthesiologists and surgeons, and pediatric speech-language pathologists—all crucial members of the specialized team caring for pediatric patients with airway pathology and/or critical illness. Each of these videos has been organized into sub-specialty-specific chapters within the Canvas educational platform to serve as a resource. For example, there are 19 instructional videos for Pediatric Intensivists, including video instruction on "how to tape and secure an endotracheal tube," "how to dress, access and care for a central venous catheter," and "how to prevent ventilator-related pneumonias." These chapters were developed after qualitative interviews with intensivists in low/middle-resource countries (30). By dividing the content into these targeted volumes, we believe that healthcare providers can access the in-depth material they need, tailored to their specific specialty, to address any knowledge gaps they wish to fill. The curriculum complements what you can read in a book and offers greater accessibility than a standard textbook—users can scan a QR code and watch a video for immediate guidance before diagnosing and assessing a patient, after an initial diagnosis when specific questions arise due to a knowledge or educational gap, or before performing a specific procedure.

This study undertakes an exploratory effort to leverage Generative Language Models (GLMs) for translation and voice synthesis to create a more inclusive and accessible educational environment in pediatric healthcare. The integration of software tools with AI functionalities forms a strategic blueprint for creating medical content with various languages and cultural backgrounds. While we initially targeted Spanish due to its global prevalence, the same methodology can be applied to diverse languages. This study primarily aims to provide an in-depth, sub-specialty-specific educational curriculum for critical care nurses, physicians, respiratory therapists, pediatric anesthesiologists, surgeons, and pediatric speech-language pathologists. It is designed to bridge educational gaps identified by their peers and to help improve the quality and care they can provide to their patients.

2 Materials and methods

Guided by the results of a learning behavior assessment survey administered to providers in a Central America-based public children's hospital within a low-resource setting (30), a narrated online video format was selected to ensure learning continuity even after the departure of a mission team. During two recent missions, first to Guatemala and then to Colombia, a comprehensive needs assessment was undertaken by an interdisciplinary team of experienced nurses (15), respiratory therapists (4), speech-language pathologists (2), anesthesiologists (10), intensivists (10), general practitioners (2), and surgeons (8) to identify gaps and process maps that could be addressed through video content (30). Using these insights, concise narrated videos were crafted at Mass General Brigham sites. These videos encapsulate key aspects of pediatric care, such as surgical procedures, perioperative care, patient journeys, and educational resources with best-practice guidelines for patients and caregivers. With durations ranging from 3 to 8 min, these high-quality videos, embedded with QR codes, were made freely available on learning platforms such as Canvas.

To improve accessibility, English narratives embedded in the videos were transcribed utilizing the "Dictate" function of Microsoft Word 365, prioritizing linguistic accuracy and contextual fidelity. Subsequently, OpenAI's Generative Pre-trained Transformer (GPT-4) was employed to facilitate the translation of content into Spanish. Native Spanish-speaking medical professionals, who were selected from the missions and well-versed in the challenges of low-resource settings, meticulously reviewed the translated content to ensure that the tone and sentiment aligned with the original material.

Building on the contextual translation, we focused on transmitting an authentic and relatable auditory experience. We collected voice samples from five candidates, each reciting the "Rainbow Passage" for 2 min. The Rainbow Passage is a standard reading passage used in speech-language pathology and voice assessment. It contains many sounds and intonations found in the English language, which improves the quality of the AI-driven synthetic voices. These recordings, captured via a mobile phone in lossless quality, were processed using the ElevenLabs Python library (elevenLabs_multilingual_v1). The model is derived from an extensive dataset of audio recordings and audiobooks designed to generate synthetic voices that closely emulate human speech characteristics. Voice synthesis was calibrated with tunable parameters emphasizing clarity (60%) and stability (30%). These parameters can be adjusted to fine-tune the synthetic speech output. Once synthesized, the audio was rendered as an mp3 file with a 44.1 KHz sampling rate. This synthesized audio track was integrated into the original video using Apple's iMovie software (v.10.3.8). To achieve optimal synchronization between visual and auditory elements, the playback speed of the synthesized audio was occasionally adjusted using the original English audio track as a reference.

Figure 1 outlines the procedure for creating a 5-min educational video, which typically requires about an hour to produce. A significant part of this time is allocated to achieving narrative consistency. Our rationale for selecting the software tools used in this study was guided by their widespread accessibility, simplicity, and advanced capabilities. Additionally, many offer APIs that facilitate the adaptation of these tools to suit diverse workflows.

3 Results

We have developed a series of videos aimed at enhancing pediatric care, covering crucial areas from surgical interventions to nursing protocols. To enhance accessibility, these videos were translated into Spanish using state-of-the-art generative models, focusing on the primary demographic of North and Latin America, regions where our non-profit CareWays Collaborative frequently operates. Figure 2 displays snapshots from various curriculum chapters. For a detailed breakdown, refer to Supplementary Figure S1.

Subsequent sections outline the objectives and content of individual chapters.

Introduction: Provides an overview of the curriculum, highlighting the motivations and emphasizing the patient's journey across various stages of their care, including initial evaluation, surgery, intensive care, swallowing therapy, safe discharge planning, the transition from hospital to home, including caregiver responsibilities, and patient-caregiver interactions. The introduction is showcased in the English version (Supplementary Video S1) and in the machine-learning-enabled Spanish version (Supplementary Video S2). Although these examples are provided in English and Spanish, the methodology allows the creation of similar content with any other supported languages.

Ward nursing: Discusses transitional patient care from an intensive care unit to a general ward, emphasizing parent-centric care and individualized insights to improve patient outcomes. The nursing introduction here is presented both in the English version (Supplementary Video S3) and the machine-learning-enabled Spanish version (Supplementary Video S4).

Surgery: Offers a curriculum for specific aspects of global health surgery, covering various surgical techniques.

Intensive care unit: Presents a comprehensive guide for intensive care training, detailing best practices for respiratory management, intravenous procedures, and patient transition from the operating room to the intensive care unit (ICU).

Speech therapy: Introduces the role of Speech-Language Pathology in patient care, covering preoperative and postoperative dysphagia screening.

Anesthesia: Focuses on the preparation and execution of bronchoscopy cases, stressing equipment familiarity and best practices.

Patient or caregiver: Provides a tailored educational approach for caregivers and patients, covering the entire journey of surgery and postoperative care, aiming to prepare them emotionally and intellectually for the medical process and for parental care responsibilities after discharge.

4 Discussion

Global healthcare presents universal challenges that go beyond language and cultural boundaries. This study aims to address these challenges by creating a comprehensive curriculum tailored for a true global collaborative where education can be shared across various languages. The curriculum is built on a decade of research into effective healthcare practices. With a particular focus on Latin America, it emphasizes the importance of a holistic, patient-centered approach to global healthcare. It covers various topics, from surgery and intensive care to guidance for parents and caregivers. This broad scope ensures that the educational material has the potential to significantly impact healthcare professionals and caregivers by providing them with tailored information.

A key innovation of this study is the use of artificial intelligence, specifically generative language models, to surmount existing language and cultural barriers. While our initial focus is on Spanish translations, integrating GLMs allows an easy transition to other languages and cultures. We plan to extend our curriculum to

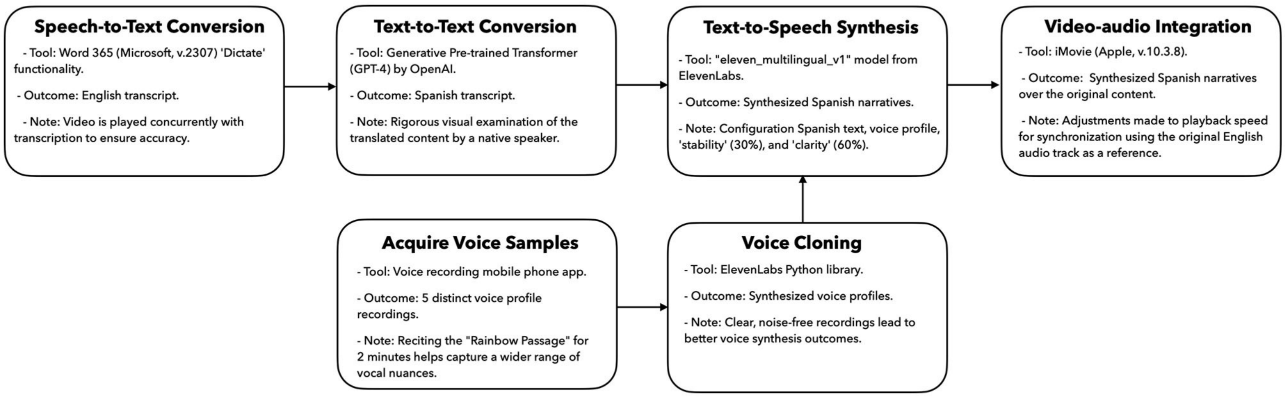


FIGURE 1
English to Spanish video translations using generative language models.

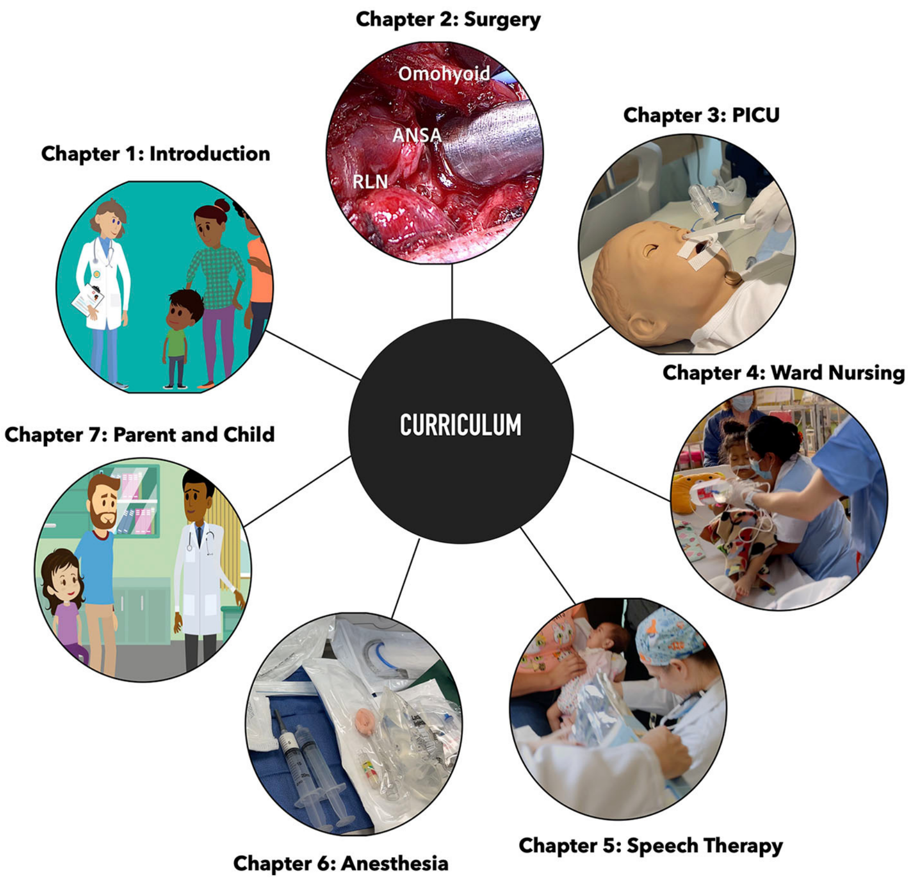


FIGURE 2
Snapshots from curriculum chapters.

include French, Portuguese, and Arabic translations, aiming for an even greater linguistic and cultural footprint. Additionally, our animated content is designed to mirror global diversity, creating a universal connection that is relatable for learners worldwide. However, incorporating AI into the curriculum development requires oversight due to the potential for inaccuracies and the lack of validation mechanisms (ground truth). In this study, we use native speakers to maintain linguistic precision and cultural nuances.

To evaluate the potential of translations produced by these GLMs, we randomly selected three surgical videos translated by AI and had their original English versions reviewed by medical translators. The translations from the medical professionals and the AI were then assessed by a Spanish-speaking surgeon in

Colombia. She reported that both sets of translated videos would provide equal value in supporting her in treating her patients. The Spanish translations completed by the medical professionals were subsequently re-translated into English using ChatGPT and evaluated by a surgeon in Boston. He determined that both the re-translated videos and the original English source material, which had undergone translation and back-translation, were equally informative and offered the same level of support for treating his patients.

There are several potential avenues for enhancing the current workflow. Firstly, the utilization of GLMs for translation can be refined by parallelizing the process and incorporating efficient speech-to-text tools, such as open-source solutions like SpeechBrain, DeepSpeech, or APIs like AWS Transcribe, known for its expertise with medical content. Secondly, the text-to-speech translation can be accelerated by leveraging solutions like ElevenLabs or IBM Watson TTS API.

In conclusion, generative language models can potentially transform global healthcare education, provided they are applied with the necessary safeguards, including data privacy measures and ethical use (31). Our initiative leverages machine learning to create educational resources that are both globally relevant and locally sensitive.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

FB: Formal analysis, Methodology, Visualization, Writing—original draft. EH: Formal analysis, Methodology, Visualization, Writing—review & editing. EZ: Formal analysis, Project administration, Validation, Visualization, Writing—review & editing. CHe: Formal analysis, Resources, Visualization, Writing—review & editing. KC: Methodology, Validation, Visualization, Writing—review & editing. PV: Formal analysis, Methodology, Visualization, Writing—review & editing. PY: Conceptualization, Formal analysis, Methodology, Supervision, Visualization, Writing—review & editing. CHa: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Visualization, Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1337395/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Curriculum breakdown.

SUPPLEMENTARY VIDEO S1
Overall Introduction to the Video Curriculum.

SUPPLEMENTARY VIDEO S2
Introducción al video currículum.

SUPPLEMENTARY VIDEO S3
Introduction Nursing Care Home to Hospital to Home.

SUPPLEMENTARY VIDEO S4
Visión general de enfermería UCIP a la sala a casa.

References

1. Marmot M, Friel S, Bell R, Houweling TA, Taylor S, Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet*. (2008) 372:1661–69. doi: 10.1016/S0140-6736(08)61690-6

2. Blumenthal D, Fowler EJ, Abrams M, Collins SR. COVID-19—implications for the health care system. *N Engl J Med*. (2020) 383:1483–8. doi: 10.1056/NEJMsb2021088

3. Newacheck PW, Hughes DC, Stoddard JJ. Children's access to primary care: differences by race, income, and insurance status. *Pediatrics*. (1996) 97:26–32.

4. Stevens GD, Seid M, Mistry R, Halfon N. Disparities in primary care for vulnerable children: the influence of multiple risk factors. *Health Serv Res*. (2006) 41:507–31. doi: 10.1111/j.1475-6773.2005.00498.x

5. Smedley BD, Stith AY, Nelson AR. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press (2003). doi: 10.17226/12875

6. Chan KS, Keeler E, Schonlau M, Rosen M, Mangione-Smith R. How do ethnicity and primary language spoken at home affect management practices and outcomes in children and adolescents with asthma? *Arch Pediatr Adolesc Med*. (2005) 159:283–9. doi: 10.1001/archpedi.159.3.283

7. Muñoz-Price LS, Nattinger AB, Rivera F, Hanson R, Gmehlin CG, Perez A, et al. Racial disparities in incidence and outcomes among patients with COVID-19. *JAMA Netw Open*. (2020) 3:e2021892. doi: 10.1001/jamanetworkopen.2020.21892

8. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a systematic review. *Oman Med J*. (2020) 35:1–7. doi: 10.5001/omj.2020.40

9. Cohen S, Moran-Ellis J, Smaje C. Children as informal interpreters in GP consultations: pragmatics and ideology. *Sociol Health Illn*. (1999) 21:163–86. doi: 10.1111/1467-9566.00148

10. Jacobs EA, Shepard DS, Suaya JA, Stone E-L. Overcoming language barriers in health care: costs and benefits of interpreter services. *Am J Public Health*. (2004) 94:866–9. doi: 10.2105/AJPH.94.5.866

11. Ryan JM. *COVID-19*. London: Routledge (2021). doi: 10.4324/9781003155911

12. Lemkow–Tovías G, Lemkow L, Cash-Gibson L, Teixidó-Compaño E, Benach J. Impact of COVID-19 inequalities on children: an intersectional analysis. *Sociol Health Illn*. (2022) 45:145–62. doi: 10.1111/1467-9566.13557

13. Mishra V, Seyedzenouzi G, Almohtadi A, Chowdhury T, Khashkhusha A, Axiaq A, et al. Health inequalities during COVID-19 and their effects on morbidity and mortality. *J Healthc Leadersh*. (2021) 13:19–26. doi: 10.2147/JHL.S270175

14. MacLachlan, M. *Culture and Health: A Critical Perspective towards Global Health*, 2nd ed. Hoboken, NJ: John Wiley and Sons Ltd (2006).

15. Cohen AL, Christakis DA. Primary language of parent is associated with disparities in pediatric preventive care. *J Pediatr*. (2006) 148:254–8. doi: 10.1016/j.jpeds.2005.10.046

16. Becker Herbst R, Margolis KL, Millar AM, Muther EF, Talmi A. Lost in translation: identifying behavioral health disparities in pediatric primary care. *J Pediatr Psychol*. (2016) 41:481–91. doi: 10.1093/jpepsy/jsv079

17. Flores G, Tomany-Korman SC. The language spoken at home and disparities in medical and dental health, access to care, and use of services in US children. *Pediatrics*. (2008) 121:e1703–14. doi: 10.1542/peds.2007-2906

18. World Health Organization. WHO *Guideline: Recommendations on Digital Interventions for Health System Strengthening*. Geneva World Health Organization (2019).

19. Seneviratne M, Shah N, Chu L. Bridging the implementation gap of machine learning in healthcare. *BMJ Innov*. (2019) 6:bmjinnov-2019. doi: 10.1136/bmjinnov-2019-000359

20. Amisha F, Malik P, Pathania M, Rathaur V. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. (2019) 8:2328. doi: 10.4103/jfmpc.jfmpc_440_19

21. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med*. (2021) 4:5. doi: 10.1038/s41746-020-00376-2

22. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: a survey. *arXiv*. (2021). doi: 10.1145/3505244

23. Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records *Nat Commun*. (2020) 11:3852. doi: 10.1038/s41467-020-17431-x

24. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*. (2011) 306:848–55. doi: 10.1001/jama.2011.1204

25. Yim W, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. *JAMA Oncol*. (2016) 2:797–804. doi: 10.1001/jamaoncol.2016.0213

26. Li J, Lavrukhin V, Ginsburg B, Leary R, Kuchaiev O, Cohen JM, Nguyen H, et al. Jasper: an end-to-end convolutional neural acoustic model. *arXiv [Preprint]. arXiv: abs/1904.03288* (2019).

27. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. GPT-4 technical report. *arXiv [Preprint]. arXiv: abs/2303.08774* (2023).

28. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: open foundation and fine-tuned chat models. *arXiv [Preprint].arXiv: abs/2307.09288* (2023).

29. Anil R, Dai AM, Firat O, Johnson M, Lepikhin D, Passos A, et al. PaLM 2 technical report. *arXiv [Preprint].arXiv: abs/2305.10403* (2023).

30. Jayawardena ADL, Ghersin ZJ, Mirambeaux M, Bonilla JA, Quiñones E, Zablah E, et al. A sustainable and scalable multidisciplinary airway teaching mission: the operation airway 10-year experience. *Otolaryngol Head Neck Surg*. (2020) 163:971–8. doi: 10.1177/01945998209 35042

31. Doo FX, Cook TS, Siegel EL, Joshi A, Parekh V, Elahi A, et al. Exploring the clinical translation of generative models like ChatGPT: promise and pitfalls in radiology, from patients to population health. *J Am Coll Radiol*. (2023) 20:877–85. doi: 10.1016/j.jacr.2023. 07.007

Advanced Optical Technologies

TYPE Original Research
PUBLISHED 23 February 2024
DOI 10.3389/aot.2024.1306142

OPEN ACCESS

EDITED BY
Saman Ghaffarian,
University College London, United Kingdom

REVIEWED BY
Marcel Leutenegger,
Max Planck Institute for Multidisciplinary Sciences, Germany
Hanumantha Raju,
BMS Institute of Technology, India

*CORRESPONDENCE
Liudmyla Kryvenko,
ls.kryvenko@knmu.edu.ua

RECEIVED 11 October 2023
ACCEPTED 05 February 2024
PUBLISHED 23 February 2024

CITATION
Kryvenko L, Krylova O, Lukin V and Kryvenko S (2024), Intelligent visually lossless compression of dental images.
Adv. Opt. Technol. 13:1306142.
doi: 10.3389/aot.2024.1306142

Intelligent visually lossless compression of dental images

Liudmyla Kryvenko[1]*, Olha Krylova[2], Vladimir Lukin[3] and Sergii Kryvenko[3]

[1]Department of Pediatric Dentistry and Implantology, Kharkiv National Medical University, Kharkiv, Ukraine, [2]Department of Therapeutic Dentistry, Kharkiv National Medical University, Kharkiv, Ukraine, [3]Department of Information and Communication Technologies, National Aerospace University, Kharkiv, Ukraine

Background: Tendencies to increase the mean size of dental images and the number of images acquired daily makes necessary their compression for efficient storage and transferring via communication lines in telemedicine and other applications. To be a proper solution, lossy compression techniques have to provide a visually lossless option (mode) where a desired quality (invisibility of introduced distortions for preserving diagnostically valuable information) is ensured quickly and reliably simultaneously with a rather large compression ratio.

Objective: Within such an approach, our goal is to give answers to several practical questions such as what encoder to use, how to set its parameter that controls compression, how to verify that we have reached our ultimate goal, what are additional advantages and drawbacks of a given coder, and so on.

Methods: We analyze the performance characteristics of several encoders mainly based on discrete cosine transform for a set of 512 × 512 pixel fragments of larger size dental images produced by Morita and Dentsply Sirona imaging systems. To control the visual quality of compressed images and the invisibility of introduced distortions, we have used modern visual quality metrics and distortion invisibility thresholds established for them in previous experiments. Besides, we have also studied the so-called just noticeable distortions (JND) concept, namely, the approach based on the first JND point when the difference between an image subject to compression and its compressed version starts to appear.

Results: The rate-distortion dependences and coder setting parameters obtained for the considered approaches are compared. The values of the parameters that control compression (PCC) have been determined. The ranges of the provided values of compression ratio have been estimated and compared. It is shown that the provided CR values vary from about 20 to almost 70 for modern coders and almost noise-free images that is significantly better than for JPEG. For images with visible noise, the minimal and maximal values of produced CR are smaller than for the almost noise-free images. We also present the results of the verification of compressed image quality by specialists (professional dentists).

Conclusion: It is shown that it is possible and easy to carry out visually lossless compression of dental images using the proposed approaches with providing quite high compression ratios without loss of data diagnostic value.

KEYWORDS

just noticeable distortions, dental image, visually lossless compression, quality metrics, intelligent image processing

1 Introduction

Imaging systems have become a conventional tool for getting valuable diagnostic information in medicine (Guy and Ffytche, 2005; Prince and Links, 2006; White and Pharoah, 2014; Suetens, 2017). They are used in ophthalmology, gastroenterology, dentistry, and other areas (Baghaie et al., 2015; Jayachandran, 2017; Federle et al., 2018). Due to the increase in spatial resolution, acquired images are usually quite large and their size often exceeds 1 MB (Anthony Seibert, 2020; Sridhar et al., 2022). This relates to medical images of different types including dentistry (Mohammad-Rahimi et al., 2023), which is of prime attention in this paper. The large size of images causes problems in their storage (Slone et al., 2000; Johnson et al., 2009; HBC, 2023) and/or transferring via communication lines in telemedicine (Fornaini and Rocca, 2022). This leads to the necessity to carry out efficient image compression (Koff and Shulman, 2006; Sanchez Silva, 2010; Flint, 2012).

Archiving and compression of medical images have a long story. Twenty years ago, many specialists insisted that only lossless compression could be applied (Fidler and Likar, 2007; Suapang et al., 2010; Liu et al., 2017). The problem of lossless compression is that the attained compression ratio is usually small and this does not satisfy specialists that exploit images in practice. After intensive discussions, it was decided that lossy compression could be used but only under the condition that compression is near-lossless or visually lossless, i.e., does not introduce visible distortions and, thus, does not result in losing diagnostically valuable information (Kocsis et al., 2003; Wu et al., 2003; Fidler and Likar, 2007; Kim et al., 2010; Ye et al., 2019).

This has led to studies intended on the design of appropriate techniques (see Foos et al., 1999; Wu et al., 2003; Fidler and Likar, 2007; Kim et al., 2010; Georgiev et al., 2013; Al-Shebani et al., 2019; Ye et al., 2019 and references therein). The influence of lossy compression on image diagnostic properties has been investigated (Eraso et al., 2002; Lehmann et al., 2006; Braunschweig et al., 2009). The appropriateness of the idea of visually lossless compression has been confirmed (Slone et al., 2000; Kocsis et al., 2003). However, a question was how to provide this in practice. The problem is that the visibility of distortions depends, at least, on three factors. The first factor is a used coder and the peculiarities of distortions introduced by it. As known, JPEG introduces blocking effects (artifacts) (Slone et al., 2000; Afnan et al., 2023) and this is undesired [similar effects, but to a lesser degree, can be observed for other coders based on discrete cosine transform (DCT) (Ponomarenko et al., 2005); because of this, image deblocking is often used after decompression]. In turn, wavelet-based coders such as, e.g., JPEG 2000 (Christopoulos et al., 2000) and SPIHT (Kim and Pearlman, 1997) produce ringing artifacts (Punchihewa et al., 2005; Kim et al., 2010; Zhang et al., 2012) and this is undesired as well. The second factor is image complexity (Lukin et al., 2022) where, on the one hand, a simple structure image can be compressed with a larger compression ratio (CR) without visible distortions, and, on the other hand, complex structure images are characterized by a better property of distortion masking (Ponomarenko et al.). The third factor are viewing conditions (Mikhailiuk et al., 2021).

Then, one possible approach is to determine the maximally possible CR for a given class of images and a given coder when distortions are invisible for any image. Such an approach needs special experiments with observers (experts) carried out in advance for a rather large set of images typical for a given application (Slone et al., 2000; Kocsis et al., 2003; Wolski et al., 2018). In addition, whilst it is easy to set and provide a desired CR for JPEG2000 or SPIHT, it is not easy to do for JPEG and other DCT-based coders since CR for them depends on image properties and varies in wide limits for a given value of parameter that controls compression (PCC) such as quality factor (QF) for JPEG or quantization step (QS) for the coder AGU (Ponomarenko et al., 2005) (this will be shown later). Another drawback of this approach is that there could be images for which the chosen (recommended) CR produces lossy compression at the edge of distortion invisibility whilst for other images there could be a large "reserve," i.e., a larger CR can be attained without visual loss of image quality.

Then, another idea arises—we should compress images adaptively considering their complexity and/or other properties with control of visual quality (Wu et al., 2003; Lastri et al., 2005; Ponomarenko et al., 2011; Võ et al., 2011; Ponomarenko et al., 2013). In Võ et al. (2011), the authors exploit different peculiarities of masking in heterogeneous image regions, edge/detail neighborhoods, and textures to appropriately set coder parameters. In Ponomarenko et al. (2013), the authors show that noise intensity and image blurriness determine distortion visibility threshold and, thus, JPEG QF has to be set adaptively. Noise type and its spatial-spectral properties are taken into consideration in (Lastri et al., 2005; Ponomarenko et al., 2011) to provide invisibility of distortions. Correlation between image quality metrics and distortion visibility threshold has been studied (Kim et al., 2010; Wolski et al., 2018; Afnan et al., 2023). It has been shown that visual quality metrics, both widely known and the ones designed recently (Johnson et al., 2011; Wolski et al., 2018; Afnan et al., 2023) perform better than conventional peak signal-to-noise ratio (PSNR). Note that the papers (Lastri et al., 2005; Ponomarenko et al., 2011; Võ et al., 2011; Ponomarenko et al., 2013) deal with other than dental types of images. This shows that, on the one hand, the task of providing visually lossless compression is quite general. On the other hand, it is worth using experience gained in other areas in the design of visually lossless techniques for dental image compression.

As it was already mentioned, in providing the desired visual quality of compressed images, it has become popular to apply visual quality metrics (Wang et al., 2003; Zemliachenko et al., 2016; Blau and Michaeli, 2019; Mantiuk et al., 2023). Their benefits compared to conventional metrics such as mean square error (MSE) and peak signal-to-noise ratio were confirmed in numerous papers (see Jayaraman et al., 2012; Ponomarenko et al., 2015a; Matsumoto, 2018 and references therein). Then, it is also assumed that the distortion visibility threshold for a given visual quality metric is already established (Ponomarenko et al., 2015a). Hence, the task in compression of a given image by a chosen coder is to provide a chosen metric value not worse than the corresponding threshold. This task can be solved by several practical procedures. One way is to apply iterative compression (Zemliachenko et al., 2016). This approach provides accurate solutions, but it might require too many iterations of compression and decompression leading to inappropriate time expenses. There are also two ways to reach the vicinity of the distortion visibility threshold approximately (with less accuracy). One way is to apply a two-step approach (Li, 2022; Li et al., 2022) based on the average rate-distortion curve obtained in advance, image compression/decompression at the first step and PCC refining with the final compression at the second step. Another way is to set a fixed PCC providing, on average, a slightly better value of the used quality metric than for distortion invisibility threshold. Both approaches will be further analyzed and discussed in the remainder part of this paper. The latter one has been intensively studied in our recent papers (Krivenko et al., 2020; Krylova et al., 2021; Kryvenko et al., 2022) for three different DCT-based coders, namely, ADCTC (advanced DCT coder) (Ponomarenko et al., 2007), AGU-M (Zemliachenko et al., 2016), and better portable graphics (BPG) (BPG Image format, 2022) encoders, respectively. It has been shown that by setting a proper PCC [QS for the ADCTC, scaling factor (SF) for the AGU-M coder, and parameter Q for the BPG encoder] it is possible to provide the metric PSNR-HVS-M (Ponomarenko et al.) in the range 40 . . . 46 dB with the mean value equal of about 42.5 dB where distortion visibility threshold for the metric PSNR-HVS-M is about 41 dB for noise-free images subject to lossy compression (Ponomarenko et al., 2015a).

Here it is worth saying that different imaging systems produce dental images of different quality that also depend on a chosen imaging mode (Flynn et al., 1996; Huda and Abrahams, 2015; Abramova et al., 2020). In particular, the system Morita (Diagnostic and Imaging Equipment, 2020) produces spatially correlated signal-dependent noise (Abramova et al., 2020) that is visible, especially in homogeneous image regions. Lossy compression of noisy images has several specific features (Al-Shaykh and Mersereau, 1998; Zemliachenko et al., 2015; Naumenko et al., 2022) including the so-called noise filtering effect. In our case, we do not need to have the noise filtering effect due to lossy compression appearing itself to full extent. Instead, we prefer to have such a compression that does not allow an observer to see changes (distortions) due to lossy compression that can be provided under certain conditions (Ponomarenko et al., 2020) discussed later.

Finally, there is an approach based on just noticeable distortions (JND) (Liu et al., 2020; Bondžulić et al., 2021; Testolina et al., 2023), namely, the first just noticeable difference point. The authors of Bondžulić et al. (2021) state that there is a high correlation between certain image features and the position of the first JND point (certain QF value) for JPEG. Then, by calculating such features, it becomes possible to properly set QF. However, this approach has not been yet applied to more modern DCT-based coders.

The paper's contributions consist in the following. First, we carry out a comparison of the performance of the ADCT, AGU-M, and BPG coders as well as JPEG for a set of 512 × 512 pixel fragments produced by the Morita system with setting the fixed values of the corresponding PCC. Second, we analyze what benefits can be provided if the two-step approach to providing a desired visual quality is applied. Third, we test the considered approaches for a set of image fragments produced by the system Dentsply Sirona where the noise intensity is less than in images produced by the Morita system.

The paper is structured as follows. Section 2 describes the possible modes of image analysis by specialists that determine requirements for compressed image quality. Image/noise properties are discussed as well. Section 3 analyzes the approach to visually lossless image compression based on setting the fixed PCC for different coders. The results for the two-step approach are given in Section 3. This section also contains initial data for the compression based on JND. Section 4 deals with lossy compression of dental images that are almost noise-free. The results of statistical verification are presented in Section 4. Finally, the conclusions are given.

2 Methods

2.1 Methodology of dental image receiving, analysis and basic mage/noise properties

To understand when lossy compression can be applied and how it can influence image quality, let us briefly consider a typical procedure of image acquisition, transferring, and analysis. As an example, we consider such a procedure for one dental center in Ukraine although, according to our knowledge, the procedures in other countries are similar.

The procedure related to patients and X-ray receiving has been performed at the University Dental Center, at the Department of Pediatric Dentistry and Implantology of Kharkiv National Medical University, Kharkiv, Ukraine, after obtaining informed consent from all patients. All actions in the dental office have been carried out using the protocols for providing dental care to the population in Ukraine (www.moz.gov.ua). Orthodontic patients undergo standard diagnostic procedure which is similar in most countries (see, for example, American Association of Orthodontists' instructions given at https://www2.aaoinfo.org/practice-management/cpg/) and include panoramic and cephalometric X-rays.

The present study has included only adult (>18 years) patients referred from general dental practitioners to the University Dental Center for diagnosis and treatment of orthodontic pathology. A designated expert committee (composed of the four local dental clinicians involved in the study) has checked the suitability of patients for the study, and the inclusion and exclusion criteria in a predefined clinical examination schedule that was agreed upon in advance. The successful candidates were scheduled for 30-min appointments at the University Dental Centre and fully assessed.

In total, 40 patients have been included in the investigation. The inclusion criteria for the patients were the following: the age is more than 18, the need for orthodontic treatment, the patients who had not received it previously, and the patients without acute tooth pain or acute health problems. The exclusion criteria were as follows: the non-adult patients (under 18), the patients with acute or chronic periodontal problems, the patients currently undergoing cancer treatment, and pregnant women. Thus, the homogenous group of population was presented that minimized a possible influence of general factors on X-ray image quality and made it possible to obtain the statistically significant results. The personnel included a team of 4 designated registered dental doctors (orthodontists and general practitioners), all calibrated and trained in advance.

The standard procedure starts with interviewing a patient. Then, a clinical examination with standard equipment and indicating the mode of X-ray examination (panoramic X-ray, lateral cephalography produced by Morita and Dentsply Sirona systems) is performed. The decision concerning X-ray type is undertaken by a

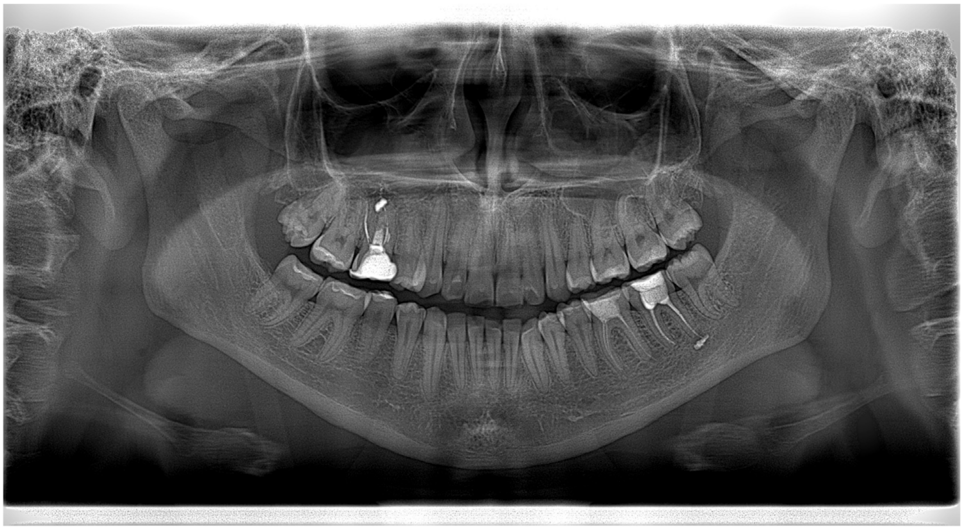


FIGURE 1
Large size image produced by the Morita system.

dental specialist, both orthodontist and general practitioner. The decision is undertaken on whether the X-ray is necessary for getting a correct diagnosis and for adequate treatment planning. According to the decision, a patient is sent to visit the diagnostic X-ray laboratory, where he/she is subject to an X-ray examination according to the indication list. The basic purpose of X-ray diagnostics is to detect dental pathology, to diagnose orthodontic pathology, and to indicate proper treatment.

After this, the results of X-ray examination are commonly sent to a dentist by e-mail [information such as the patient's name and sex do not accompany the image(s) in order to protect his/her privacy]. Just at this stage, a performer of X-ray might use lossy compression of acquired images or attach uncompressed images to e-mail.

In our study, we needed images suitable for diagnostics. Because of this, the "entrance control" has been performed. Four previously trained dentists have analyzed 65 images produced by the Morita and Dentsply Sirona systems (40 panoramic X-ray and 25 cephalometric X-ray images). The Clinical Image Quality Evaluation Chart was used for the evaluation of the quality of original (uncompressed) images. At the stage of anonymous image evaluation, 61 images have been recognized as "optimal for obtaining diagnosis," 3 have been considered "adequate for diagnosis," 1 has been treated as "poor but diagnosable," and there were no images classified as "unrecognizable."

For all such images, the compression declared as visually lossless should not result in decreasing their diagnostic quality.

Dental images are usually viewed and analyzed by specialists without applying automatic means of image processing and interpreting. Just because of this, the original (acquired) and compressed image visual quality is of prime importance. Meanwhile, the size of the original images, noise level, and methodology of image representation and analysis can be different. These factors determine the requirements to image visually lossless compression that have to be recalled.

First of all, image size can be quite large and depends on the imaging system mode. Figure 1 shows image acquired by the Morita system, panoramic X-ray (Vera-viewepocs 3D R100 J) (Diagnostic and Imaging Equipment, 2020). The size of the image presented in Figure 1 is 2761 × 1504 pixels, it occupies a few Megabytes. The images acquired by the Dentsply Sirona (Orthophos S) have a slightly smaller size of 2048 × 1087 pixels.

Then, the acquired images can be exploited in a different manner. First, they can be saved in a clinic depository and/or passed to a doctor and/or to a patient. Saving in a depository is desired since a patient or a doctor might need this image later or for some other purposes. Passing to a patient can be done because the patient might go to another clinic or another doctor. In both cases, image lossy compression is possible and even sometimes needed (if a great number of images are obtained in a laboratory or clinic). However, it should be visually lossless compression and no visible distortions should be seen (detectable) in any part of an image of a large size. Note that lossy compression can be also desired if communication lines have a limited bandwidth, a user or a clinic pays for Internet traffic, etc.

Although acquired images or their fragments under interest can be visualized on screens of very different devices, it is recommended and common to use laptops and stationary computers (monitors) with large sizes and appropriate quality screens. Note also that doctors can look at and analyze images in different scales using the maximal resolution scale for image fragments under interest. Because of its large size, a dental image has to be scrolled for analysis in maximal resolution scale to see the smallest details.

Since it is possible to expect that the monitor type has an impact on image perception, good screens are mainly used for visualization and analysis of an image in aggregate or its parts. In our study, the evaluators used the following monitors: (a) a monitor of laptop ASUS (15.6″, 1920 × 1080, Full HD, IPS), (b) an HP monitor (with a diagonal 27.1″, 1920 × 1080, Full HD, IPS), (c) a monitor of IPad (diagonal 10.2″, 2160 × 1620, IPS). Since distortions due to lossy

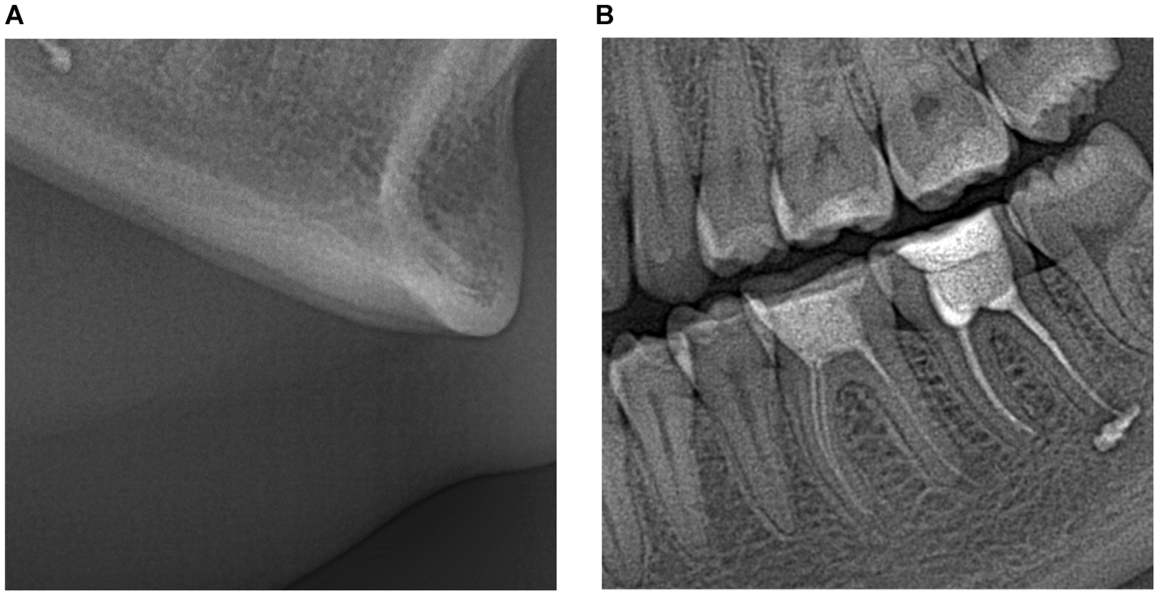


FIGURE 2
512 × 512 pixel fragments of different complexity taken from large size image in **(A,B)**.

compression appear themselves more for the maximal resolution scale, the doctors evaluated the images visually using the maximum zooming of images for the aforementioned monitors. Then, we need to provide visually lossless compression for the maximal resolution scale for fragments of the size of a few hundred to a few hundred pixels. Keeping this in mind, we have to carry out analysis correspondingly, i.e., for image fragments. We have chosen their size to be 512 × 512 pixels since it is convenient for DCT-based coders. In addition, such size fragments can be easily and conveniently placed nearby to each other for comparing the original and compressed versions according to recommendations in Testolina et al. (2021). Examples of such fragments taken from the image in Figure 1 are presented in Figure 2.

As it is seen, image fragments can be of different complexity where the fragment in Figure 2B contains more details, edges, and textures compared to a rather simple image fragment in Figure 2A. Besides, noise can be easily noticed in the latter image, especially in homogeneous areas having medium intensity. This is not surprising because of the following reasons. First, for the safety of patients, X-ray images are low dose and this leads to the presence of quite intensive noise (Lee et al., 2018). Second, noise is signal-dependent (Lee et al., 2018; Abramova et al., 2020) and this explains why the noise is better seen in homogeneous image regions having a larger local mean. Third, the noise is spatially correlated (Abramova et al., 2020) and this is one reason why noise is visible—spatially correlated noise is more visible than white noise of the same intensity. Meanwhile, noise characteristics also depend on the imaging mode (Abramova et al., 2020). The most adequate model of the noise occurred to be $\hat{\sigma}^2 = \hat{\sigma}^2_\mu \cdot I^2_{ij} + \hat{k} \cdot I_{ij} + \hat{\sigma}^2_a$ where $\hat{\sigma}^2_\mu$ and $\hat{\sigma}^2_a$ are the estimates of multiplicative and additive noise variances, respectively, $\hat{k}$ is the estimate of the quasi-Poissonian component parameter, and I_{ij} denotes image true value in the ij-th pixel. Then, for a given fragment, equivalent variance can be estimated as

$$\hat{\sigma}^2_{eq} = \hat{\sigma}^2_a + \hat{\sigma}^2_\mu \sum_{i=1}^{N} \sum_{j=1}^{M} I^2_{ij} \Big/ NM + \hat{k} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} \Big/ NM \quad (1)$$

where N and M in Eq. 1 define the fragment size. Then, for one mode of the Morita system operation, $\hat{\sigma}^2_{eq}$ varies from about 10 to 60 depending on fragment mean intensity. For the second mode, $\hat{\sigma}^2_{eq}$ varies from 30 to almost 200 (for the fragment in Figure 2B) (Krivenko et al., 2020). In any case, the noise occurs to be visible.

Meanwhile, for images acquired by the Dentsply Sirona system, the noise is hardly noticed.

2.2 Providing visually lossless compression by setting the fixed PCC

2.2.1 Used metrics and considered coders

Recall that for a visually lossless approach we need some adequate visual quality metric and the corresponding distortion invisibility threshold. Metric adequateness for a given type of distortion is usually determined by analysis of the Spearman rank order correlation coefficient (SROCC) between metric values and mean opinion score (MOS) for image databases that contain images with the considered type of distortions. The image database TID 2013 (Ponomarenko et al., 2015b) is a good option that contains images distorted by lossy compression. It has been established that, for many metrics including PSNR, SROCC exceeds 0.9, i.e., distortions due to lossy compression are quite adequately characterized. Meanwhile, there are several metrics for which SROCC is between 0.96 and 0.97 including PSNR-HVS-M (Ponomarenko et al.), feature similarity (FSIM) index (Zhang et al., 2011), mean deviation similarity index (MDSI) (Ziaei Nafchi et al., 2016), Haar wavelet-based perceptual similarity index (HaarPSI) (Reisenhofer et al., 2018), and some others

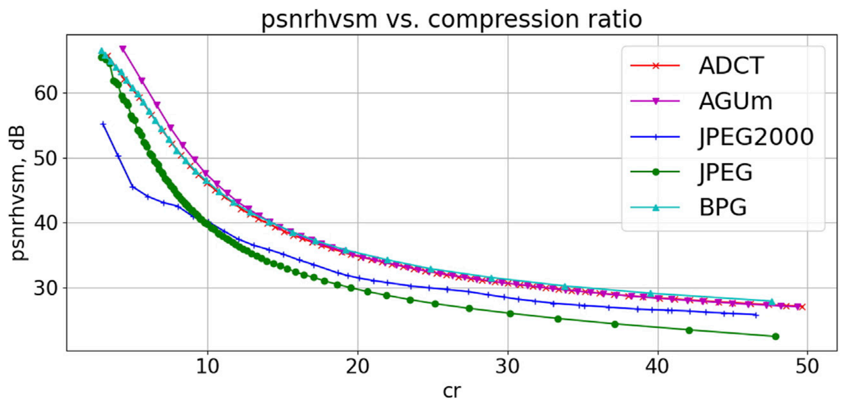


FIGURE 3
Dependences PSNR-HVS-M(CR) for a fragment of dental image in Figure 1.

(Wang et al., 2004). For some of them, the invisibility threshold has been determined. For example, for PSNR-HVS-M, the threshold is $T_{PHVSM} \approx 40$ dB (Ponomarenko et al., 2015a). In turn, for MDSI the threshold is $T_{MDSI} \approx 0.22$ (Li et al., 2022). So, let us rely on the metric PSNR-HVS-M expressed for 8-bit images as

$$PSNR - HVS - M^n = 10\,log_{10}\left(\frac{255^2}{MSE - HVS - M^n}\right), \quad (2)$$

where *MSE-HVS-M*^{n} in Eq. 2 is calculated in 8 × 8 blocks in the DCT domain taking into consideration two peculiarities of the human vision system (HVS)—the lower sensitivity of HVS to distortions in high spatial frequencies than distortions in low spatial frequencies and masking effect. Similarly to PSNR, PSNR-HVS-M is expressed in dB and its larger values relate to better visual quality. For additive white Gaussian noise and similar distortions, PSNR-HVS-M occurs to be slightly larger than the corresponding PSNR due to the masking effect. This property can indirectly describe the properties of distortions introduced by image lossy compression (Abramova et al., 2023).

In this paper, we consider four DCT-based coders. One of them is JPEG controlled by quality factor (QF). Smaller QF values are associated with larger CR and greater introduced distortions.

ADCTC (Ponomarenko et al., 2007) employs a partition scheme to adapt to image content and uses rectangular shape blocks where all sizes of block sides are powers of two to ensure the possibility of using fast DCT algorithms. The coder is not fast since partition scheme optimization needs some time, the decompression is faster than compression.

The AGU-M coder uses 32 × 32 pixel blocks and an advanced algorithm of bit-plane coding of quantized DCT coefficients. In opposition to the standard AGU (https://ponomarenko.info/#dow), AGU-M employs different quantization steps for different spatial frequencies and uses scaling factor (SF, analog of QS) as PCC. The larger SF results in larger CR and greater distortions introduced.

The better portable graphics (BPG) encoder is a part of the HEVC video coder and it has several advantages. In particular, the BPG encoder provides higher CR compared to JPEG and many other methods for the same quality characterized by PSNR. Its available versions can operate with data from 8 to 14 bits per channel. Here, we present the results obtained for the grayscale BPG version 0.9.8 offered at https://bellard.org/bpg/. The parameter Q (that can be only integer and varies from 1 to 51) plays the role of PCC. Its larger values correspond to a larger CR and greater distortions.

Figure 3 allows comparing the coders' performance for one fragment of the dental image. Dependencies for all coders are given as functions of CR to offer an opportunity to compare the results (recall that the coders have different PCCs).

Analysis of data in Figure 3 shows the following tendencies. AGU-M coder produces the best results in the area of interest (CR=10–15, PSNR-HVS-M>40 dB). The ADCT and BPG encoders perform closely to AGU-M. JPEG and JPRG2000 produce significantly worse results. Thus, the plots in Figure 3 explain one more time why we have paid attention to the analysis of the ADCT, BPG, and AGU-M in our previous and current studies. Similarly, according to PSNR, ADCTC is the best for small CR whilst the BPG encoder is the best for large CR. AGU-M produces results similar to the ADCT and BPG coders. JPEG and JPEG2000 perform considerably worse, especially in the area of interest (CR=10–20, PSNR>35 dB for the three best coders).

2.2.2 The use of fixed PCC for DCT-based coders

Different reasoning can be put into the basis of setting some fixed PCC for DCT-based coders for the considered application. Let us start our analysis for the standard JPEG. It is sometimes supposed that setting QF = 75 practically guarantees that distortions are invisible (Bondžulić et al., 2023). We have applied JPEG with QF = 75 to 20 fragments of the size 512 × 512 pixels taken from the image in Figure 1. It occurred that the minimal and maximal CRs are equal to 5.45 and 11.37, respectively. Minimal and maximal PSNR-HVS-M values are 45.26 dB and 50.65 dB, respectively, i.e., the difference is about 5 dB. Note that the largest PSNR-HVS-M is observed just for the fragment having the smallest CR. These data show that QF can be smaller since there is a reserve for decreasing the PSNR-HVS-M values.

To see what QF can be set, we have calculated the mean PSNR-HVS-M for 20 fragments of Morita images compressed by JPEG with different QF. It follows that mean PSNR-HVS-M equals 40 dB for QF about 49. However, in this case, there are fragments having PSNR-HVS-M smaller and larger than 40 dB (approximate

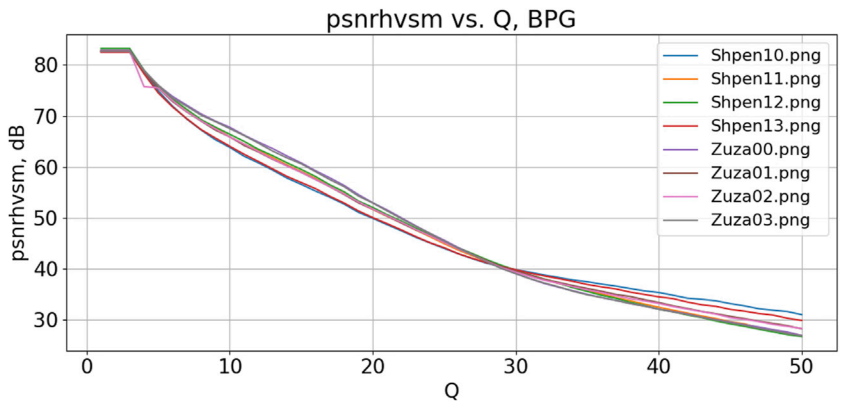


FIGURE 4
Dependencies of PSNR-HVS-M on Q for the BPG encoder for eight image fragments.

threshold of distortion invisibility). Assuming that PSNR-HVS-M can differ from its mean values by ±2.5 dB, we should set such QF that provides mean PSNR-HVS-M equal to 42.5 dB. This takes place for QF = 60.

Thus, we set QF = 60 and determined CR and PSNR-HVS-M values. The CR values vary from 6.8 to 19.7, PSNR-HVS-M values vary from 41.2 dB to 44.8 dB. This means that by setting QF = 60 the requirement to have invisible distortions for JPEG is satisfied.

A similar analysis has been done for three other coders. For ADCTC it was recommended to set QS = 12 (Krivenko et al., 2020). In this case, for the same set of 20 image fragments, the mean PSNR-HVS-M was the same as above (42.5 dB) where PSNR-HVS-M varied in the limits from 40.5 dB to 45.6 dB (i.e., in slightly larger interval than for JPEG, this is a drawback) and CR varied from 7.5 to 20.6 (i.e., CR values are larger than for JPEG, this is an advantage).

For the AGU-M coder, we recommended setting SF = 8.8 to have the same mean PSNR-HVS-M (Krylova et al., 2021). Then, PSNR-HVS-M was in the limits from 41.1 dB to 45.1 dB (this result is better than for ADCTC) whilst CR varied from 9.4 to 35.0 (both minimal and maximal values are larger than the corresponding values in the previous cases).

Finally, for the BPG encoder, the results are the following (Kryvenko et al., 2022). It was recommended to set Q = 28 (analysis of several dependencies for particular image fragments presented in Figure 4 shows that this is a correct decision). This led to PSNR-HVS-M within the limits from 41.8 dB to 45.9 whilst CR varies from 8.6 to 16.2. This is slightly worse than for the AGU-M coder.

Note that, since Q for the BPG encoder can be only an integer and Q increasing by unity leads to PSNR-HVS-M reduction by about 1.5 dB, it is difficult to compare the obtained results to the corresponding results for other coders. Because of this, here we also present the results for Q = 29. PSNR-HVS-M varies within the limits from 40.3 dB to 44.4 dB, CR is from 9.5 to 18.2, i.e., CR has improved by the expense of a lower visual quality and the performance is at approximately the same level as for the ADCTC.

Above, we have considered approaches to lossy compression based on the fixed setting of PCC for the DCT-based coders. Let us denote them as JPEG-FS, ADCTC-FS, AGU-M-FS, and BPG-FS, respectively. A common advantage is that no decompression is used and, thus, the compression procedures are relatively fast (except the ADCT coder for which partition scheme optimization takes significant time). Meanwhile, one general conclusion that follows from the presented results is that, for the fixed PCC, the PSNR-HVS-M values vary in some limits and this opens a certain room for further improvement.

In particular, in the paper (Kryvenko et al., 2022), it was proposed to use the following procedure - apply compression with Q = 28 at the first step and determine PSNR-HVS-M_1 after decompression. If PSNR-HVS-M_1 is within the limits from 41.75 dB to 43.25 dB, leave the compression result. If PSNR-HVS-M_1 is outside these limits, then calculate Q as Q = 28 + [(PSNR-HVS-M_1 – 42.5)/1.5] where [ ] denotes rounding to the nearest integer. This provides a mean PSNR-HVS-M of about 42.5 dB and narrower limits of its variation after the final step. Similar procedures called two-step can be realized for other coders. They are considered in the next section.

3 Results

3.1 Two-step providing of appropriate visual quality

The basic idea of the two-step procedures (Li, 2022; Li et al., 2022) is the following. For a rather small interval for PCC variation, dependencies of PSNR-HVS-M on PCC are almost linear and they are "almost parallel" to each other for particular images. This is seen well in Figure 4 for the BPG coder if Q is in the interval under interest (Q from 25 to 33). Then, knowing a metric value for a given PCC and having some estimate of the derivative of the corresponding rate-distortion curve, it becomes possible to find a PCC that approximately corresponds to the desired value of the considered metric $Metr_{des}$.

More in detail, suppose that, in advance (off-line), the average rate-distortion curve $Metr_{av}$ (PCC) has been obtained. Then, it can be used for two purposes: first—to determine PCC_1, for which $Metr_{av}$ (PCC) ≈ $Metr_{des}$, and, second, to determine $M' = dMetr_{av}/dPCC$. The first step is to compress an image using PCC_1, to decompress it and to determine $Metr_1$. This value can be smaller or larger than $Metr_{des}$ and, knowing M', it is possible to calculate PCC_2 as

$$PCC_2 = PCC_1 + (Metr_{des} - Metr_1)\big/ M'. \tag{3}$$

Note that, in Eq. 3, $Metr_{av}$ is often used instead of $Metr_1$ (Li et al., 2020a). These two values are the same if PCC can fall into any value. However, for the JPEG and BPG coders, PCC values can be only integer and, thus, PCC_1 can be only integer too. Then, the expression (Suetens, 2017) is more general.

Then, assuming the Gaussian distribution of residual errors of providing $Metr_{des}$ after the second step, we can set $Metr_{des} \approx Metr_{thr} + 3\sigma_M$ where σ_M is the standard deviation of residual errors for the two-step method. σ_M depends on the following main factors—a metric used and its desired value and a coder considered. For the metric PSNR-HVS-M, the corresponding studies have been carried out for the ADCT (Li et al., 2020a) and BPG (Li et al., 2020b) coders. For ADCTC for PSNR-HVS-M_{des} about 40 dB, $\sigma_M \approx 0.55\,dB$ for highly textured images and smaller for simpler structure ones. So, we can set PSNR-HVS-M_{des} = 41.65 dB for the two-step procedure. For the BPG encoder, $\sigma_M \approx 0.6\,dB$ and, thus, it is reasonable to set PSNR-HVS-M_{des} = 41.8 dB. Note that in both cases, PSNR-HVS-M_{des} is smaller than the average PSNR-HVS-M (42.5 dB) we provided for the PCC fixed setting (see previous Section).

The advantages of the two-step approach are that it usually provides minimal PSNR-HVS-M larger than for the case of the fixed PCC setting and maximal PSNR-HVS-M smaller than for the fixed PCC setting. In the first case, this results in smaller probability that distortions are visible. In the second case, a larger CR is provided.

We have not carried out experiments for JPEG and AGU-M coders intended to determine σ_M for them for different PCCs. However, taking into account the limits of variation of PSNR-HVS-M for the fixed PCC for average PSNR-HVS-M equal to 42.5 dB reported in the previous Section, it is possible to set PSNR-HVS-M_{des} = 41.65 dB for both JPEG and AGU-M. We have also checked if this setting is correct by calculating the residual errors in experiments with the set of dental image fragments.

Let us start with the results for the AGU-M coder. For the two-step procedure, SF at the second step is from 8.4 to 11.7, the minimal PSNR-HVS-N has occurred to be equal to 41.4 dB whilst the maximal is of about 41.8 dB. This means that the variations of image visual quality according to the metric PSNR-HVS-M are very small. CR values are from 11.1 to 38.3, i.e., minimal and maximal CRs are better than for any other approach considered above.

For the BPG-encoder, the final values of Q are from 26 to 30, the minimal and maximal PSNR-HVS-M are from 41.8 dB to 43.3 dB (i.e., the interval is larger than for AGU-M), and CRs are in the limits from 9.9 to 19.3. In other words, the results are more stable than for the fixed Q according to PSNR-HVS-M and slightly better according to CR.

The compression parameters for the ADCT coder are the following. QS at the second step varies in the limits from 11.0 to 16.3, PSNR-HVS-M is in the limits from 41.3 dB to 41.8 dB, and minimal and maximal CRs are equal to 8.3 and 28.9. Totally, the results are at the same level as for the BPG encoder and worse than for the AGU-M coder.

Finally, for JPEG, QF is from 47 to 61, PSNR-HVS-M is in the limits from 41.3 dB to 41.7 dB and CR varies from 7.9 to 20.4. This is, in general, worse than for all coders considered above. According to the case of fixed QF setting (see Section 3), there is a small benefit in CR values.

Summarizing the results of the two-step procedure, we can state that it provides more stable values of PSNR-HVS-M (they are very close to the desired PSNR-HVS-M) and larger values of minimal and maximal CR. The payment for these improvements is an increase in computations since more time is spent on image compression, decompression, and final compression at the second step.

3.2 JPEG-compression based on JND

Let us first recall some results presented in the papers (Bondžulić et al., 2021; Bondžulić et al., 2023; Testolina et al., 2023). In Bondžulić et al. (2023), analysis of QF values for the first JND point (JNDP1) has been carried out for two image databases specially designed for this purpose, MCL-JCI and JND-Pano (panoramic images). Although color image compression has been studied, the obtained results seem valuable for our case. It has been shown that PSNR for JNDP1 varies in very large limits—from 27.6 to 46.0 dB for the MCL-JCI database and from 20.9 to 44.7 dB for the JND-Pano database. QF varies from 25 to 70 and from 38 to 75, respectively. This shows that the adequateness of PSNR and distortion invisibility threshold for it [about 36 dB according to Ponomarenko et al. (2015a)] is of doubt. In turn, PSNR-HVS-M for JNDP1 varies from 36.2 to 48.1 dB for the MCL-JCI database and from 39.9 to 49.2 dB for the JND-Pano database. Thus, PSNR-HVS-M is more adequate in characterizing visual quality (its limits of variation are considerably narrower than for PSNR) although PSNR-HVS-M and its invisibility threshold are not perfect.

In Ponomarenko et al. (2020) it has been also shown that, for noisy images, the difference between the same image contaminated by the noise of different intensities, becomes visible if intensities differ by 10%–20%. Then, since we have equivalent variances from 10 to 200 (see Section 2.1), the MSE of distortions introduced by lossy compression should be from 1 to 20. For images compressed with MSE = 1, distortions are not seen, they start to be visible for MSE ≈ 3 in the worst case. Thus, the PSNR of noisy images compressed in a visually lossless manner can be from approximately 35 dB–43 dB depending on the noise intensity. These results are in good agreement with the results reported above for both databases.

There are approaches to the prediction of QF for JNDP1 (Lin et al., 2020; Stojanovic et al., 2022). The method (Stojanovic et al., 2022) is based on exploiting a simple parameter called mean gradient magnitude (MGM) able to characterize an image to be compressed. It is shown that the RMSE of such a prediction (PSNR is predicted for JNDP1 based on MGM) is about 1 dB. MGM can be defined as:

$$MGM = \frac{1}{NM} \sum\nolimits_{\forall n,m} \frac{1}{g_{max}} \sqrt{g_x^2(n,m) + g_y^2(n,m)}, \tag{4}$$

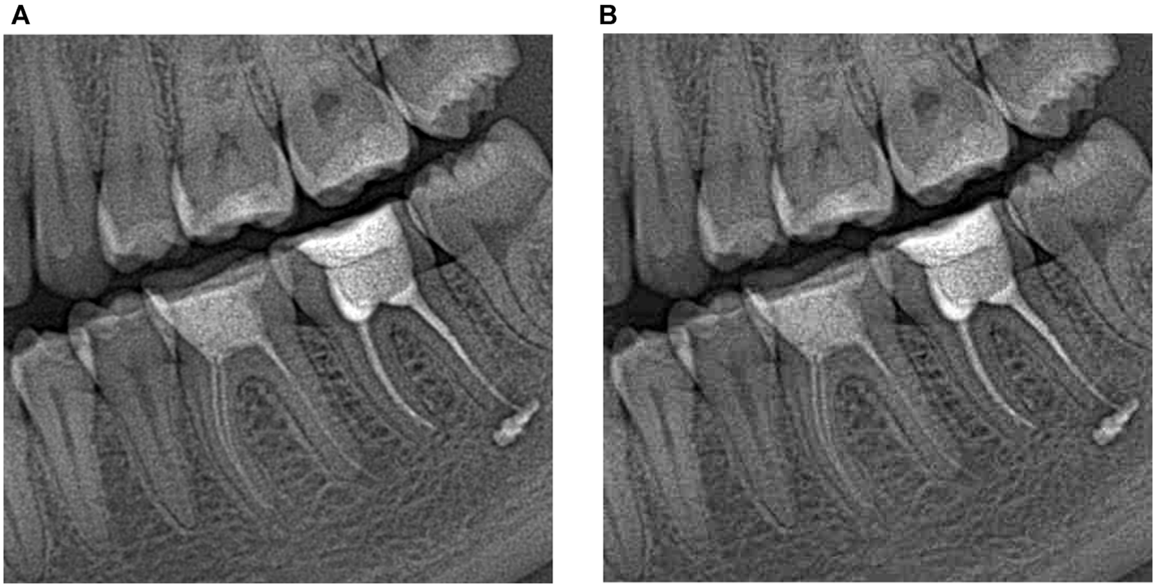


FIGURE 5
Original **(A)** and compressed **(B)** versions of image fragments.

where g_{max} in Eq. 4 is the maximum magnitude value, g_{max} = 4.472 for grayscale images with a dynamic range from 0 to 1.

MGM values are in the limits from 0 to 0.17 although they mainly concentrate in the limits from 0.01 to 0.07. For smaller MGM that corresponds to simpler structure images without noise, PSNRs for JNDP1 are larger. Whilst, for MGM ≈ 0.07, PSNRs for JNDP1 are of about 32 dB, they are of about 42 dB for MGM ≈ 0.01. The formula for PSNR prediction obtained in Bondžulić et al. (2021) is the following:

$$PSNR = \begin{cases} 2115.5MGM^2 - 377MGM + 46.4, MGM \le 0.0896 \\ 29.58, MGM > 0.0896 \end{cases}, \quad (5)$$

where for MGM = 0.0896 the mapping function in Eq. 5 reaches its minimum value equal to $PSNR_{min}$ = 29.58 dB.

We have decided to calculate MGM for our 512 × 512 fragments of Morita images. The MGM values are from 0.031 to 0.071. Then, PSNR values for JNDP1 should be from 30 dB to 35.5 dB. For each considered image fragment, a desired PSNR can be provided by the two-step procedure. For 4 out of 20 test fragments, the difference between the desired PSNR (recommended by Eq. 5) and PSNR provided by the two-step procedure exceeded 1 dB—the maximal difference was equal to 1.3 dB. Then, we have also calculated CR and PSNR-HVS-M. CR has varied in the limits from 6.4 to 20.3, i.e., approximately in the same limits as for approaches considered above. However, problems have arisen with QF and PSNR-HVS-M of compressed images. QF varied from 16 to 84. PSNR corresponding to small QF was about 30 dB, which, according to our experience, is too small. This is confirmed by image fragments in Figure 5 where original (Figure 5A) image and the corresponding compressed one (Figure 5B) are represented. Distortions (mainly, blocking artifacts) are seen in compressed image. Thus, we do not deal with the desired visually lossless compression.

Meanwhile, there are also fragments that have been compressed with PSNR in the limits from 38 dB to 42 dB and PSNR-HVS-M from 44.5 dB to 49.5 dB. For them, no distortions are visible and the recommended QFs exceed 70. According to previous experience (see Section 3), smaller QF values can be used while keeping the introduced distortions invisible.

Thus, we can state that the approach based on JNDP1 prediction does not perform satisfactorily and, at the moment, cannot be recommended for practical use. To our opinion, there are several reasons behind this imperfection. First, there are several factors that lead to errors in setting QF. They are imperfect dependence of PSNR for JNDP1 on MGM where some points differ from the fitted curve by a few dB and the limited accuracy of the two-step approach that provides the desired PSNR with errors exceeding 1 dB. Second, the fitted curve (Prince and Links, 2006) has been obtained for color images without noise and we have used it for grayscale noisy images. Recall here that PSNR of about 30 dB corresponds to MSE of introduced losses of about 65 (for 8-bit images that we have in our experiments). This means that MSE is comparable to noise intensity and, thus, it is not surprising that the introduced losses are visible.

The obtained results do not mean that the approach based on JNDP1 has no potential for the considered application. However, the dependence of PSNR (or PSMR-HVS-M) on one parameter or parameters describing image characteristics has to be additionally studied and made more accurate.

3.3 Visually lossless compression of almost noise-free images

As it has been mentioned above, some X-ray imagers produce almost noise-free images. Example of such images having large sizes is given in Figure 6. As one can see, noise is not visible.

FIGURE 6
Example of large size dental images produced by the Imager Dentsly Sirona (https://www.dentsplysirona.com/en/discover/discover-by-brand/orthophos-e.html).

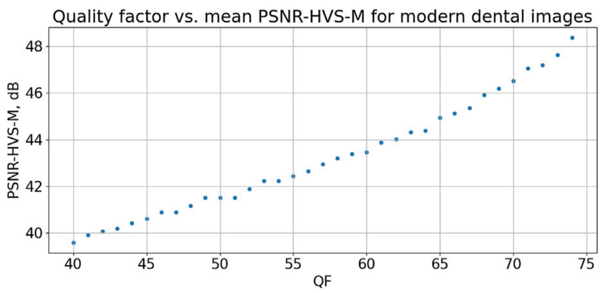


FIGURE 7
Dependence of mean PSNR-HVS-M on QF for JPEG for fragments for almost noise-free dental images.

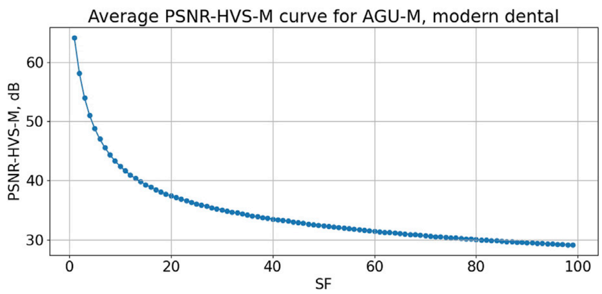


FIGURE 8
Dependence of mean PSNR-HVS-M on SF for AGU-M for fragments for almost noise-free dental images.

We have briefly analyzed the rate-distortion characteristics for fifteen 512 × 512 fragments of almost noise-free images. The goal for this was to analyze do they differ a lot from the dependencies earlier obtained for image fragments considered above. Since for all approaches we need average rate-distortion curves, let us obtain them and compare them to the previously used ones. Figure 7 presents the average dependence of PSNR-HVS-M on QF for JPEG. It is monotonous and almost linear in the area of interest. The derivative dPSNR-HVS-M/dQF is practically the same as for the Morita image fragments—about 0.22. This means that, if we would like to provide an average PSNR-HVS-M equal to 42.5 dB for fixed QF, we can set QF = 55 (for the considered type of dental images).

Let us see, what are the provided compression characteristics in this case. PSNR-HVS-M values are from 41.9 dB to 42.8 dB, i.e., in rather narrow interval. CR values are from 14.2 to 24.6, i.e., both minimal and maximal CRs are larger than the corresponding ones for the Morita image fragments.

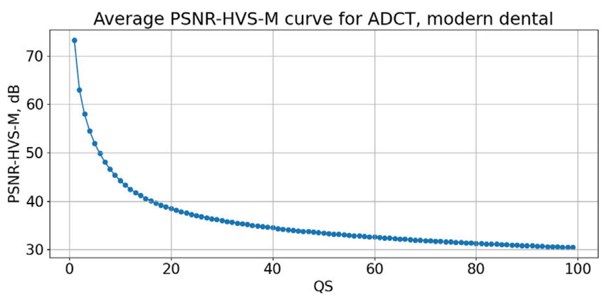


FIGURE 9
Dependence of mean PSNR-HVS-M on QS for ADCTC for fragments for almost noise-free dental images.

In turn, Figure 8 represents the dependence of PSNR-HVS-M on SF for the AGU-M coder. Similarly to the dependences in Figure 4, it is monotonous. The values of average PSNR-HVS-M are slightly larger than for Morita image fragments. If one would like to set a fixed SF for providing average PSNR-HVS-M about 42.5 dB, then SF=9.93. Setting this SF leads to the following results: PSNR-HVS-M varies in the limits from 41.9 dB to 42.9 dB, i.e., in a narrow range, CR is in the limits from 22.8 to 52.5, i.e., the CR values are considerably better than for Morita system image fragments. We associate this with the practical absence of noise. The CR values are also considerably better (larger) than for the JPEG data given above.

Figure 9 presents the average dependence of PSNR-HVS-M on QS for the ADCT coder. According to this curve, one has to set QS = 11.98 to provide the average PSNR-HVS-M equal to 42.5 dB. For this QS used as the fixed setting, the results are the following. PSNR-HVS-M varies from 41.8 dB to 43.2 dB, i.e., in quite narrow (appropriate) limits. CR is from 21.4 to 59.2, i.e., they are comparable to the interval of CR variation for the AGU-M coder.

Finally, we have checked the results for the BPG encoder. According to the average curve, Q was set equal to 27 to ensure the average PSNR-HVS-M of about 42.5 dB. As the results, PSNR-HVS-M values vary from 41.9 dB to 43.0 dB (in appropriately narrow intervals) and CRs are from 22.4 to 71.6. This means that, for more complex structure images, CR is approximately the same as for AGU-M and ADCT coders whilst, for simple structure images, there is a certain benefit in CR provided by the BPG coder.

Let us summarize the results given above in this Section. First, we have checked whether distortions are visible for some fragments compressed by all four coders providing the average PSNR-HVS-M = 42.5 dB and have not found such cases. Second, it has been established that for the fixed values of PCC, the differences in PSNR-HVS-M values for particular fragments do not differ a lot (the variation interval widths are about 1 dB). Then, it is possible to expect that the 2-step procedure is able to provide even narrower variation intervals. Thus, for the 2-step procedure, we have decided to set the desired PSNR-HVS-M equal to 42 dB and to check what results can be obtained in this case for all four coders.

For JPEG, the obtained results are the following. The provided PSNR-HVS-M varies from 41.6 dB to 42.1 dB, QF values are mostly equal to 52 although there are a few images for which QF equals either 51 or 53. The provided CR is from 17.1 to 25.5. The positive feature is that the variation range for PSNR-HVS-M has decreased. CR values have slightly increased whilst the mean PSNR-HVS-M has decreased which can be expected. Taking into account that the two-step procedure requires two compressions and one decompression, we do not see an essential difference between applying the fixed (properly set) QF or using the two-step procedure for JPEG.

For the AGU-M coder, PSNR-HVS-M varies from 41.96 dB to 42.04 dB, i.e., very high accuracy is provided. SF is from 9.8 to 11.0, i.e., SF adaptation to image properties takes place. Finally, CR values are from 23.3 to 58.5, i.e., they are considerably better than for JPEG and slightly better than for the case of setting fixed SF for the AGU-M coder (see the corresponding data above).

The coder ADCT has produced similar results. PSNR-HVS-M is in the limits from 41.97 dB to 42.08 dB, i.e., the desired PSNR-HVS-M is provided with high accuracy. This is due to the adaptation of QS to image content—QS varies from 11.8 to 13.5. CR values are from 22.7 to 63.2, i.e., the minimal CR is slightly less and the maximal CR is greater than for the AGU-M coder.

Finally, the BPG coder produces PSNR-HVS-M in the limits from 41.9 dB to 42.9 dB. As one can see, the provided mean PSNR-HVS-M is slightly larger than the desired one. This results from the fact that Q can be only integer. The Q values for almost all image fragments are equal to 27. CR values are from 22.5 to 58.7, i.e., practically the same as for the AGU-M coder. No improvement compared to the fixed setting of Q is offered. This is explained by two reasons. First, the BPG coder produces quite close values of PSNR-HVS-M for fixed Q that differ from each other by about 1–1.2 dB, at least, if the desired PSNR-HVS-M are in the range of interest (40–44 dB). Changing of Q by 1 leads to PSNR-HVS-M changing by about 1.5–1.7 dB, i.e., by more than the aforementioned range width. This means that the two-step procedure produces a limited improvement of accuracy for the considered situation and it is not worth employing it for the BPG coder.

Summarizing the obtained results, we can state that the use of the two-step procedure for providing the desired PSNR-HVS-M offers some benefits for the coders AGU-M and ADCT since the desired PSNR-HVS-M can be provided with higher accuracy and this leads to a certain increase in CR. Meanwhile, there are no obvious reasons to apply the two-step procedure for JPEG and BPG encoders since the accuracy of providing the desired PSNR-HVS-M for them does not improve a lot because of setting PCCs as only integer values.

We have also tested the approach based on MGM calculation, PSNR prediction and its providing by the two-step method. MGM values are smaller than for image fragments with visible noise considered in Sections 3, 4, they are in the limits from 0.012 to 0.026. Then, the predicted PSNRs for JNDP1 are larger—from 38.1 dB to 42.0 dB. They have been provided by the two-step procedure with errors not exceeding 1 dB. As a result, CRs are from 23 to 37 (good results), but PSNR-HVS-M are within the limits from 35.7 dB to 38.7 dB, i.e., below the distortion visibility threshold. Then, we have checked the compressed image fragments. It has occurred that for many of them the introduced distortions are visible.

Figure 10 shows the fragment obtained after visually lossless compression using the two-step procedure for the four considered coders (further denoted as JPEG2st, AGU-M2st, ADCT2st, and BPG2st, respectively). As one can see, it is difficult to find differences between the compressed images (maybe, the image compressed by ADCTC is slightly sharper). There is practically no difference compared to the original image.

Comparing the results presented in this Section for almost noise-free images to the results for noisy images in the previous two Sections, the following two conclusions can be drawn. First, average rate-distortion curves differ a little. Average values of PSNR-HVS-M for the same PCC for almost noise-free images are slightly larger for noise-free images (this difference does not lead to any problem if the two-step procedure is applied). Second, on average, larger final values of CR are obtained for noise-free images—this is not surprising, see the results presented in Krivenko et al. (2018).

Above, in the design of visually lossless compression of dental images and comparison of performance characteristics for different coders, we have relied on the following. First, the results on the distortion visibility threshold presented for images in the database TID2013 (Ponomarenko et al., 2015a) have been taken into account. Second, we have taken into account the results of verification experiments earlier carried out for the ADCT, AGU-M, and BPG

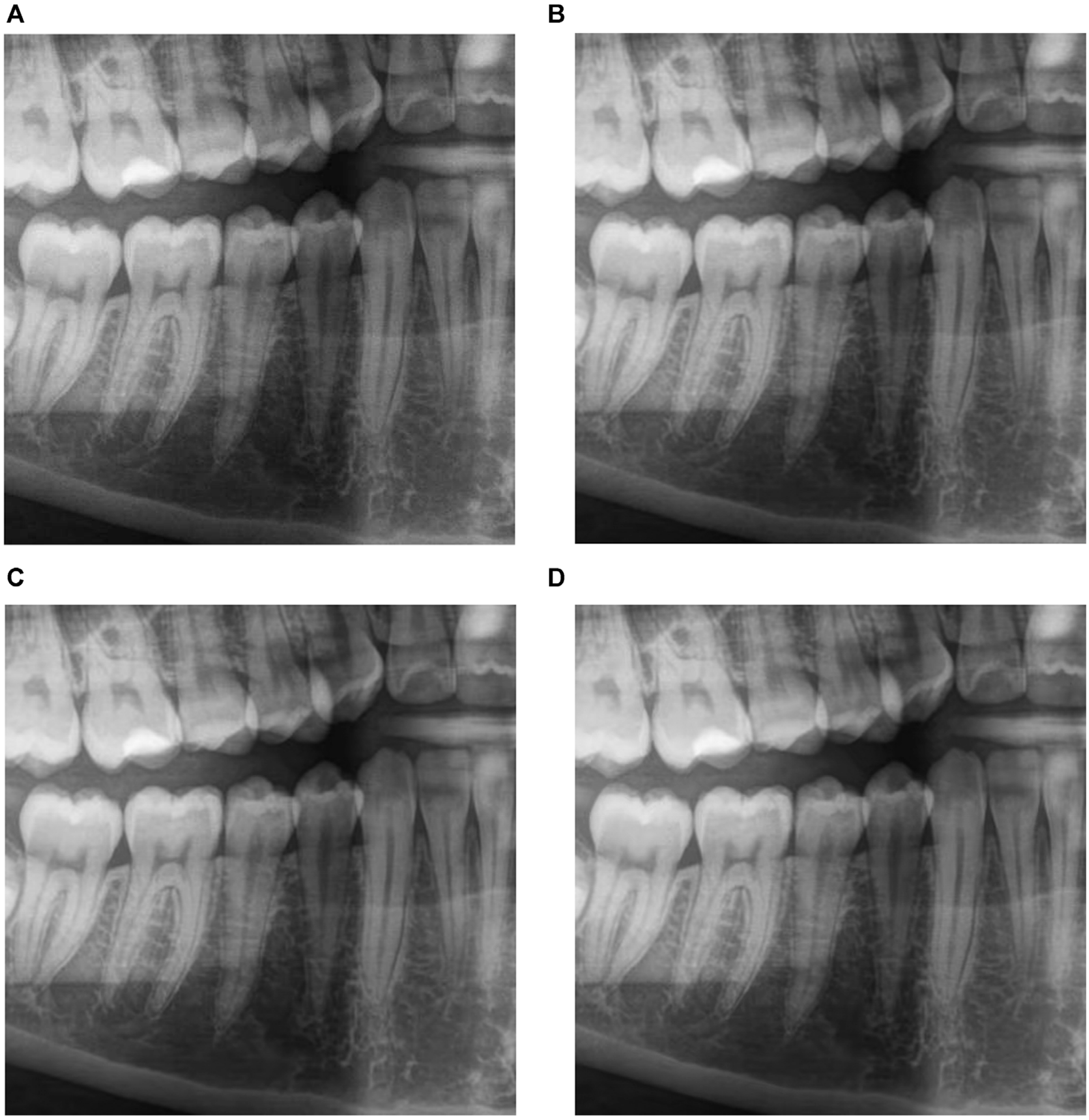


FIGURE 10
The fragment obtained after visually lossless compression using the two-step procedure for JPEG, CR = 14.7, PSNR-HVS-M = 41.97 dB **(A)**, AGU-M, CR = 23.5, PSNR-HVS-M = 42.00 dB **(B)**, ADCT, CR = 23.2, PSNR-HVS-M = 42.00 dB **(C)** and BPG, CR = 26.3, PSNR-HVS-M = 41.37 dB **(D)** coders.

coders with the fixed PCCs described in our papers (Krivenko et al., 2020; Krylova et al., 2021; Kryvenko et al., 2022). However, the threshold of distortion invisibility according to the PSNR-HVS-M metric is approximate and this has been confirmed by experiments in Bondžulić et al. (2021). Thus, we need to be sure that the proposed approaches to visually lossy compression do not lead to a reduction of the diagnostic value of compressed images.

4 Discussion

We have carried out experiments with compressed image fragments to understand the following: 1) how often a specialist (dentist) sees the differences between original and compressed image fragments; 2) do these differences have an impact on the diagnostic value of compressed images; 3) do the results for the considered coders differ between each other. Preparing this experiment, we have taken into account the recommendations for performing such experiments and previous experience. First, an observer's distance to the monitor has to be fixed and it should be convenient for a particular observer. Second, it is usually enough to have about 15 s to decide if there are differences in the viewed images and if are there artifacts in the compressed image (that, in our case, influence its diagnostic value). Third, the experiment should not be too long since an observer becomes tired and starts to perform his/her task improperly. Fourth, between pairs of images following each other in comparison, there should be a small break (dark screen) "to remove the previous pair from human memory."

 10.3389/aot.2024.1306142

TABLE 1 The obtained experimental data for the Morita imager.

Compression approach and its parameters	P_d	Range of PSNR-HVS-M variation (dB)	Range of CR variation
JPEG-FS (QF = 60)	0.165	40.8–45.3	6–22
ADCTC-FS (QS = 12)	0.085	40.3–46.1	7–27
AGU-M-FS (SF = 8.8)	0.140	40.7–45.7	8–39
BPG-FS (Q = 28)	0.150	41.1–46.4	8–23
JPEG2st (PSNR-HVS-M_{des} = 41.65 dB)	0.175	41.1–42.0	7–22
ADCTC2st (PSNR-HVS-M_{des} = 41.65 dB)	0.075	41.2–42.0	8–37
AGU-M2st (PSNR-HVS-M_{des} = 41.65 dB)	0.155	41.3–41.9	10–39
BPG2st (PSNR-HVS-M_{des} = 41.8 dB)	0.090	41.0–43.8	9–28

Thus, each pair of image fragments was shown for comparison for 5–20 s and an observer had to press either the button "Identical" or the button "Different." In the latter case, the observer had to press either the button "Appropriate" or the button "Inappropriate" where the latter means that the distortions introduced by lossy compression have led to lost diagnostic information. Between the subsequent pairs, 5-s break with the dark screen was offered. Each observer has been given 50 pairs of image fragments to be compared. As a result, the experiment did not last more than 25 min (in fact, no one experiment lasted more than 22 min with an average duration of 16 min). The fragments have been randomly chosen from the considered large-size images in advance, compressed, and saved in the folder used in the experiments. This was done since the compression and decompression time could be larger than the time taken for comparison of each pair of image fragments. Note that compression and decompression time depends on several factors including the coder used (the largest compression/decompression time is for the ADCT coder), computer characteristics, and software realization of a given coder.

The original image fragment was always placed left with respect to the corresponding compressed fragment. The buttons were put below the visualized image fragments. Taking into account that each of the four doctors carried out 50 comparisons, 200 comparisons have been performed total (for each coder). Thus, the probability that two image fragments are different (P_d) could be determined. No considerable difference in the results for different monitors have been observed. Similarly, no significant difference has been noticed in the results for doctors participating in experiments. Comparisons for different coders (and their variants) were done on different days. The doctors were not told what coder is under study at the current moment. Each doctor participated in experiments carried out on different computers. Monitors have been viewed approximately from the distance $1.6D_v$ where D_v is the monitor height.

The comparison results for images acquired by the Morita system are presented in Table 1. In addition to P_d, we give the additional data that allow comparing the considered approaches to compression: the range of PSNR-HVS-M values and the range of CR values for the analyzed image fragments. Keeping in mind that P_d is estimated with RMSE about 0.02, the conclusions that can be drawn are the following:

1) The fixed setting of the PCC leads to approximately the same P_d as the two-step approach for a given coder; the difference for them is in narrower ranges of PSNR-HVS-M variation and slightly larger minimal and maximal CR values for the two-step approach;
2) The ADCTC with fixed QS setting and with the use of the two-step procedure produces slightly smaller P_d, a more thorough comparison has shown that object edges are preserved by this coder better than by other considered coders;
3) Wider ranges of CR variation have been observed in experiments compared to the cases analyzed in previous Sections; this can be explained by wider variations of image fragment content (200 fragments have been considered instead of 15–20 analyzed in previous experiments);
4) Slightly larger CR values have been observed for the two-step procedure; the largest values took place for the AGU-M coder although the differences in CR values are not significant.

It is rather important to notice that in none of the cases, the quality of compressed images for the methods listed in Table 1 was treated as inappropriate. Meanwhile, we have also tested the compression procedure for JPEG based on JNDP1. For this one, P_d is equal to 0.27, i.e., the difference between original and compressed image fragments has been found considerably more often. Moreover, with probability of 0.08, the compressed images were treated as inappropriate, mainly because of visible blocking effects (artifacts).

Since the results for the fixed setting of PCC and two-step procedure were close to each other, only the two-step procedure has been used in experiments for almost noise-free images. The obtained results are presented in Table 2. The desired PSNR-HVS-M was equal to 42 dB for all considered coders.

As one can see, the values of P_d are approximately in the same range as in Table 1 where the best result (the smallest P_d) is again provided by the ADCT coder. Besides, the ADCT and AGU-M coders provide the smallest variations of PSNR-HVS-M whilst JPEG and the BPG encoders are characterized by wider variations (since PCC values for them can be only integer). No cases when the quality of compressed images has been considered inappropriate have been detected.

The CR values for three modern coders are significantly larger than for JPEG and they are considerably larger than 10. The CR values are also, in general, considerably larger than the CR values in Table 1 and this can be explained by the fact that images acquired by the Dentsply Sirona system contain less noise.

TABLE 2 The obtained experimental data for the Dentsply Sirona imager.

Compression approach and its parameters	P_d	Range of PSNR-HVS-M variation (dB)	Range of CR variation
JPEG2st (PSNR-HVS-M_{des} = 42 dB)	0.160	41.5–42.4	14–29
ADCTC2st (PSNR-HVS-M_{des} = 42 dB)	0.080	41.9–42.1	21–67
AGU-M2st (PSNR-HVS-M_{des} = 42 dB)	0.125	41.9–42.1	22–62
BPG2st (PSNR-HVS-M_{des} = 42 dB)	0.145	41.3–42.8	21–64

Thus, we can recommend using PSNR-HVS-M_{des} = 42 dB and the two-step procedure for providing either visually lossless compression or lossy compression for which introduced distortions can be hardly noticed and do not affect the diagnostic value of dental images acquired by modern systems.

5 Conclusion

The task of visually lossless compression of dental images acquired by two modern systems is considered. The images usually have a large size and, due to this, it is worth applying lossy compression for their transferring via communication lines and storage. Three approaches to visually lossless compression are considered, namely, the use of the fixed PCC, the use of the two-step compression, and the approach based on the first just noticeable distortion point. It is shown by experiments carried out by four qualified dentists for three types of monitors with 200 image fragments used in comparisons that the latter approach is not perfect at the moment and it produces the largest percentage of situations when distortions in compressed image fragments are noticeable and able to negatively affect the diagnostic value of dental images.

For the two former approaches, the compression characteristics depend on whether the noise is visible or not. For images with visible noise, the minimal and maximal values of produced CR are smaller than for the almost noise-free images. Note that in the latter case, the provided CR varies from about 20 to almost 70 for modern coders that significantly outperform JPEG (in the sense of larger CR for approximately the same visual quality and P_d). If one would like to have smaller P_d than in our experiments, it is possible to increase PSNR-HVS-M_{des} for the two-step approach or to use larger fixed QF for JPEG and smaller fixed values of SF, QS, and Q for AGU-M, ADCT, and BPG coders, respectively. However, smaller minimal and maximal CR values are then produced.

In the future, we plan to consider other modern coders and quality metrics. Images produced by other dental systems can be studied as well. Considering the obtained results, it will be interesting to perform multi-centered research, increase the number of patients and collect data from dental centers with different geographical locations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethics committee of Kharkiv National Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

LK: Writing–original draft, Writing–review and editing, Data curation, Software, Supervision, Validation, Visualization, Methodology. OK: Methodology, Software, Validation, Visualization, Writing–review and editing. VL: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Writing–original draft, Writing–review and editing. SK: Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

References

Abramova, V., Krivenko, S., Lukin, V., and Krylova, O. (2020). "Analysis of noise properties in dental images," in 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine (IEEE), 511–515. Available from: https://ieeexplore.ieee.org/document/9088768/ (Accessed October 19, 2020).

Abramova, V., Lukin, V., Abramov, S., Kryvenko, S., Lech, P., and Okarma, K. (2023). A fast and accurate prediction of distortions in DCT-based lossy image compression. *Electronics* 12 (11), 2347. doi:10.3390/electronics12112347

Afnan, A., Ullah, F., Yaseen, Y., Lee, J., Jamil, S., and Kwon, O. J. (2023). Subjective assessment of objective image quality metrics range guaranteeing visually lossless compression. *Sensors* 23 (3), 1297. doi:10.3390/s23031297

Al-Shaykh, O. K., and Mersereau, R. M. (1998). Lossy compression of noisy images. *IEEE Trans Image Process* 7 (12), 1641–1652. doi:10.1109/83.730376

Al-Shebani, Q., Premaratne, P., Vial, P. J., and McAndrew, D. J. (2019). The development of a clinically tested visually lossless Image compression system for capsule endoscopy. *Signal Process. Image Commun.* 76, 135–150. doi:10.1016/j.image.2019.04.008

Anthony Seibert, J. (2020). Archiving, chapter 2: medical image data characteristics. Society for imaging informatics in medicine. Available from: https://siim.org/page/archiving_chapter2.

Baghaie, A., Yu, Z., and D'Souza, R. M. (2015). State-of-the-art in retinal optical coherence tomography image analysis. *Quantitative Imaging Med. Surg.* 5 (4), 603–617. doi:10.3978/j.issn.2223-4292.2015.07.02

Blau, Y., and Michaeli, T. (2019). "Rethinking lossy compression: the rate-distortion-perception tradeoff," in International Conference on Machine Learning, Long Beach, USA (PMLR), 675–685.

Bondžulić, B., Lukin, V., Bujaković, D., Li, F., Kryvenko, S., and Pavlović, B. (2023). *On visually lossless JPEG image compression* in 2023 Zooming Innovation in Consumer Technologies Conference (ZINC), Novi Sad, Serbia (IEEE), 113–118. Available from: https://ieeexplore.ieee.org/document/10174090/ (Accessed September 4, 2023).

Bondžulić, B., Stojanović, N., Petrović, V., Pavlović, B., and Miličević, Z. (2021). Efficient prediction of the first just noticeable difference point for JPEG compressed images. *Acta Polytech. Hung.* 18 (8), 201–220. doi:10.12700/aph.18.8.2021.8.11

BPG Image format (2022). BPG Image format. Available from: https://bellard.org/bpg/ (Accessed May 5, 2022).

Braunschweig, R., Kaden, I., Schwarzer, J., Sprengel, C., and Klose, K. (2009). Image data compression in diagnostic imaging: international literature review and workflow recommendation. *RoFo Fortschritte dem Geb. Rontgenstrahlen Nukl.* 181 (7), 629–636. doi:10.1055/s-0028-1109341

Christopoulos, C., Skodras, A., and Ebrahimi, T. (2000). The JPEG2000 still image coding system: an overview. *IEEE Trans. Consum. Electron* 46 (4), 1103–1127. doi:10.1109/30.920468

Diagnostic and Imaging Equipment (2020). Diagnostic and imaging equipment | MORITA. Available from: https://www.jmoritaeurope.de/en/products/diagnostic-and-imaging-equipment-overview/ (Accessed October 19, 2020).

Eraso, F. E., Analoui, M., Watson, A. B., and Rebeschini, R. (2002). Impact of lossy compression on diagnostic accuracy of radiographs for periapical lesions. *Oral Surg. Oral Med. Oral Pathology, Oral Radiology, Endodontology* 93 (5), 621–625. doi:10.1067/moe.2002.122640

M. P. Federle, S. R. Sinha, and P. D. Poullos (2018). *Imaging in gastroenterology* (Philadelphia, PA: Elsevier).

Fidler, A., and Likar, B. (2007). What is wrong with compression ratio in lossy image compression? *Radiology* 245 (1), 299–300. author reply 299-300. doi:10.1148/radiol.2451062005

Flint, A. C. (2012). Determining optimal medical image compression: psychometric and image distortion analysis. *BMC Med. Imaging* 12, 24. doi:10.1186/1471-2342-12-24

Flynn, M. J., Hames, S. M., Wilderman, S. J., and Ciarelli, J. J. (1996). Quantum noise in digital X-ray image detectors with optically coupled scintillators. *IEEE Trans. Nucl. Sci.* 43 (4), 2320–2325. doi:10.1109/23.531897

Foos, D. H., Slone, R. M., Whiting, B. R., Kohm, K. S., Young, S. S., Muka, E., et al. (1999). in *Dynamic viewing protocols for diagnostic image comparison*. Editor E. A. Krupinski (San Diego, CA), 108–120. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=981183 (Accessed December 18, 2023).

Fornaini, C., and Rocca, J. P. (2022). Relevance of teledentistry: brief report and future perspectives. *Front. Dent.* 19, 25. doi:10.18502/fid.v19i25.10596

Georgiev, V. T., Karahaliou, A. N., Skiadopoulos, S. G., Arikidis, N. S., Kazantzi, A. D., Panayiotakis, G. S., et al. (2013). Quantitative visually lossless compression ratio determination of JPEG2000 in digitized mammograms. *J. Digit. Imaging* 26 (3), 427–439. doi:10.1007/s10278-012-9538-7

Guy, C., and Ffytche, D. (2005). *An introduction to the principles of medical imaging*. London : Singapore ; Hackensack, NJ: Imperial College Press ; Distributed by World Scientific Pub, 374. Rev.

HBC (2023). Dental imaging in the cloud: overcoming data management and storage challenges. *Healthc. Bus. Club*. Available from: https://healthcarebusinessclub.com/articles/healthcare-provider/technology/dental-imaging-in-the-cloud-overcoming-data-management-and-storage-challenges/ (Accessed September 4, 2023).

Huda, W., and Abrahams, R. B. (2015). Radiographic techniques, contrast, and noise in X-ray imaging. *Am. J. Roentgenol.* 204 (2), W126–W131. doi:10.2214/ajr.14.13116

Jayachandran, S. (2017). Digital imaging in dentistry: a review. *Contemp. Clin. Dent.* 8 (2), 193–194. doi:10.4103/ccd.ccd_535_17

Jayaraman, D., Mittal, A., Moorthy, A. K., and Bovik, A. C. (2012). "Objective quality assessment of multiply distorted images," in 2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR) (IEEE), 1693–1697.

Johnson, J. P., Krupinski, E. A., Nafziger, J. S., Yan, M., and Roehrig, H. (2009). in *Visually lossless compression of breast biopsy virtual slides for telepathology*. Editors B. Sahiner and D. J. Manning (Lake Buena Vista, FL), 72630N. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.813786 (Accessed December 18, 2023).

Johnson, J. P., Krupinski, E. A., Yan, M., Roehrig, H., Graham, A. R., and Weinstein, R. S. (2011). Using a visual discrimination model for the detection of compression artifacts in virtual pathology images. *IEEE Trans. Med. Imaging* 30 (2), 306–314. doi:10.1109/tmi.2010.2077308

Kim, B.-Jo, and Pearlman, W. A. (1997). "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in Proceedings DCC '97 Data Compression Conference, Snowbird, UT, USA (IEEE Comput. Soc. Press), 251–260. Available from: http://ieeexplore.ieee.org/document/582048/ (Accessed September 4, 2023).

Kim, K. J., Kim, B., Mantiuk, R., Richter, T., Lee, H., Kang, H.-S., et al. (2010). A comparison of three image fidelity metrics of different computational principles for JPEG2000 compressed abdomen CT images. *IEEE Trans. Med. Imaging* 29 (8), 1496–1503. doi:10.1109/tmi.2010.2049655

Kocsis, O., Costaridou, L., Mandellos, G., Lymberopoulos, D., and Panayiotakis, G. (2003). Compression assessment based on medical image quality concepts using computer-generated test images. *Comput. Methods Programs Biomed.* 71 (2), 105–115. doi:10.1016/s0169-2607(02)00090-1

Koff, D. A., and Shulman, H. (2006). An overview of digital compression of medical images: can we use lossy image compression in radiology? *Can. Assoc. Radiol. J.* 57 (4), 211–217.

Krivenko, S., Lukin, V., Krylova, O., Kryvenko, L., and Egiazarian, K. (2020). A fast method of visually lossless compression of dental images. *Appl. Sci.* 11 (1), 135. doi:10.3390/app11010135

Krivenko, S. S., Krylova, O., Bataeva, E., and Lukin, V. V. (2018). SMART LOSSY COMPRESSION OF IMAGES BASED ON DISTORTION PREDICTION. *Telecom Rad. Eng.* 77 (17), 1535–1554. doi:10.1615/telecomradeng.v77.i17.40

Krylova, O., Kryvenko, L., Krivenko, S., and Lukin, V. (2021). "A fast noniterative visually lossless compression of dental images using AGU-M coder," in 2021 IEEE 16th International Conference on the Experience of Designing and Application of CAD Systems (CADSM) (IEEE), 6–10.

Kryvenko, L., Krylova, O., Krivenko, S., and Lukin, V. (2022). "A fast noniterative visually lossless compression of dental images using BPG coder," in 2022 11th Mediterranean Conference on Embedded Computing (MECO) (IEEE), 1–6.

Lastri, C., Aiazzi, B., Alparone, L., and Baronti, S. (2005). Virtually lossless compression of astrophysical images. *EURASIP J. Adv. Signal Process.* 2005, 192492–192515. doi:10.1155/asp.2005.2521

Lee, S., Lee, M., and Kang, M. (2018). Poisson–Gaussian noise analysis and estimation for low-dose X-ray images in the NSCT domain. *Sensors* 18 (4), 1019. doi:10.3390/s18041019

Lehmann, T. M., Abel, J., and Weiss, C. (2006). in *The impact of lossless image compression to radiographs*. Editors S. C. Horii and O. M. Ratib (San Diego, CA), 614516. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.651697 (Accessed September 4, 2023).

Li, F. (2022). Adaptive two-step method for providing the desired visual quality for SPIHT. *Radioelectron. Comput. Syst.* (1), 195–205. doi:10.32620/reks.2022.1.15

Li, F., Krivenko, S., and Lukin, V. (2020a). "A two-step procedure for image lossy compression by ADCTC with a desired quality," in 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine (IEEE), 307–312. Available from: https://ieeexplore.ieee.org/document/9125000/ (Accessed October 22, 2020).

Li, F., Krivenko, S., and Lukin, V. (2020b). "An approach to better portable graphics (BPG) compression with providing a desired quality," in 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine (IEEE), 13–17. Available from: https://ieeexplore.ieee.org/document/9349289/ (Accessed September 4, 2023).

Li, F., Lukin, V., Ieremeiev, O., and Okarma, K. (2022). Quality control for the BPG lossy compression of three-channel remote sensing images. *Remote Sens.* 14 (8), 1824. doi:10.3390/rs14081824

Lin, H., Hosu, V., Fan, C., Zhang, Y., Mu, Y., Hamzaoui, R., et al. (2020). SUR-FeatNet: predicting the satisfied user ratio curve for image compression with deep feature learning. *Qual. User Exp.* 5 (1), 5. doi:10.1007/s41233-020-00034-1

Liu, F., Hernandez-Cabronero, M., Sanchez, V., Marcellin, M. W., and Bilgin, A. (2017). The current role of image compression standards in medical imaging. *Inf. (Basel)* 8 (4), 131. doi:10.3390/info8040131

Liu, H., Zhang, Y., Zhang, H., Fan, C., Kwong, S., Kuo, C. C. J., et al. (2020). Deep learning-based picture-wise just noticeable distortion prediction model for image compression. *IEEE Trans Image Process* 29, 641–656. doi:10.1109/tip.2019.2933743

Lukin, V., Krivenko, S., Li, F., Abramov, S., and Makarichev, V. (2022). "On image complexity in viewpoint of image processing performance," in 2022 3rd International Workshop on Intelligent Information Technologies and Systems of Information Security, 35.

Mantiuk, R. K., Hammou, D., and Hanji, P. (2023). HDR-VDP-3: a multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. Available from: https://arxiv.org/abs/2304.13625 (Accessed December 18, 2023).

Matsumoto, R. (2018). Introducing the perception-distortion tradeoff into the rate-distortion theory of general information sources. *IEICE Commun. Express* 7 (11), 427–431. doi:10.1587/comex.2018xbl0109

Mikhailiuk, A., Ye, N., and Mantiuk, R. K. (2021). The effect of display brightness and viewing distance: a dataset for visually lossless image compression. *ei* 33 (11), 152–158. doi:10.2352/issn.2470-1173.2021.11.hvei-152

Mohammad-Rahimi, H., Vinayahalingam, S., Mahmoudinia, E., Soltani, P., Bergé, S. J., Krois, J., et al. (2023). Super-resolution of dental panoramic radiographs using deep learning: a pilot study. *Diagnostics* 13 (5), 996. doi:10.3390/diagnostics13050996

Naumenko, V., Lukin, V., and Krivenko, S. (2022). "Analysis of noisy image lossy compression by BPG," in *Integrated computer technologies in mechanical engineering - 2021. Lecture notes in networks and systems*. Editors M. Nechyporuk, V. Pavlikov, and D. Kritskiy (Cham: Springer International Publishing), 367, 911–923. Available from: https://link.springer.com/10.1007/978-3-030-94259-5_71 (Accessed May 5, 2022).

Ponomarenko, M., Ieremeiev, O., Lukin, V., and Egiazarian, K. (2020). An expandable image database for evaluation of full-reference image visual quality metrics. *Electron. Imaging* 32 (10), 137–146. doi:10.2352/issn.2470-1173.2020.10.ipas-137

Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., et al. (2015b). Image database TID2013: peculiarities, results and perspectives. *Signal Process. Image Commun.* 30, 57–77. doi:10.1016/j.image.2014.10.009

Ponomarenko, N., Lukin, V., Astola, J., and Egiazarian, K. (2015a). "Analysis of HVS-metrics' properties using color image database TID2013,". ACIVS in *Advanced concepts for intelligent vision systems* (Lecture Notes in Computer Science: Springer), 613–624. Available from: https://link.springer.com/chapter/10.1007%2F978-3-319-25903-1_53.

Ponomarenko, N., Lukin, V., Egiazarian, K., and Astola, J. (2005). "DCT based high quality image compression," in *Image analysis. Lecture notes in computer science*, H. Kalviainen, J. Parkkinen, A. Kaarna, D. Hutchison, T. Kanade, and J. Kittler, (Berlin, Heidelberg: Springer Berlin Heidelberg), 3540, 1177–1185. Available from: http://link.springer.com/10.1007/11499145_119 (Accessed May 11, 2021).

Ponomarenko, N., Silvestri, F., Egiazarian, K., Carli, M., Astola, J., and Lukin, V. *ON between-coefficient contrast masking of DCT basis functions*, 4.

Ponomarenko, N. N., Egiazarian, K. O., Lukin, V. V., and Astola, J. T. (2007). High-quality DCT-based image compression using partition schemes. *IEEE Signal Process Lett.* 14 (2), 105–108. doi:10.1109/lsp.2006.879861

Ponomarenko, N. N., Lukin, V. V., and Egiazarian, K. O. (2011). "Visually lossless compression of synthetic aperture radar images," in 2011 VIII International Conference on Antenna Theory and Techniques, Kyiv, Ukraine (IEEE), 263–265. Available from: http://ieeexplore.ieee.org/document/6170755/ (Accessed September 4, 2023).

Ponomarenko, N. N., Lukin, V. V., Egiazarian, K. O., and Lepisto, L. (2013). Adaptive visually lossless JPEG-based color image compression. *SIViP* 7 (3), 437–452. doi:10.1007/s11760-013-0446-1

Prince, J. L., and Links, J. (2006). *Medical imaging signals and systems*. Upper Saddle River, New Jersey: Prentice Hall, 496. Pearson Prentice Hall bioengineering.

Punchihewa, G. A. D., Bailey, D. G., and Hodgson, R. M. (2005). Objective evaluation of edge blur and ringing artefacts: application to JPEG and JPEG2000 image codecs. Image and Vision Computing New Zealand. IVCNZ, 61–66.

Reisenhofer, R., Bosse, S., Kutyniok, G., and Wiegand, T. (2018). A Haar wavelet-based perceptual similarity index for image quality assessment. Signal Processing: image Communication. *Signal Process. Image Commun.* 61, 33–43. doi:10.1016/j.image.2017.11.001

Sanchez Silva, V. F. (2010). Advances in medical image compression: novel schemes for highly efficient storage, transmission and on demand scalable access for 3D and 4D medical imaging data. Available from: https://doi.library.ubc.ca/10.14288/1.0064771 (Accessed October 19, 2020).

Slone, R. M., Foos, D. H., Whiting, B. R., Muka, E., Rubin, D. A., Pilgram, T. K., et al. (2000). Assessment of visually lossless irreversible image compression: comparison of three methods by using an image-comparison workstation. *Radiology* 215 (2), 543–553. doi:10.1148/radiology.215.2.r00ap47543

Sridhar, C., Pareek, P. K., Kalidoss, R., Jamal, S. S., Shukla, P. K., and Nuagah, S. J. (2022). Optimal medical image size reduction model creation using recurrent neural network and GenPSOWVQ. *J. Healthc. Eng.* 2022, 1–8. doi:10.1155/2022/2354866

Stojanovic, N., Bondzulic, B., Pavlovic, B., Petrovic, V., and Zelmati, O. (2022). "Just noticeable difference prediction and image quality assessment," in 2022 IEEE Zooming Innovation in Consumer Technologies Conference (ZINC), Novi Sad, Serbia (IEEE), 234–238. Available from: https://ieeexplore.ieee.org/document/9840625/ (Accessed September 4, 2023).

Suapang, P., Dejhan, K., and Yimmun, S. (2010). "Medical image archiving, processing, analysis and communication system for teleradiology," in TENCON 2010 - 2010 IEEE Region 10 Conference, Fukuoka (IEEE), 339–345. Available from: http://ieeexplore.ieee.org/document/5686025/ (Accessed September 4, 2023).

Suetens, P. (2017). *Fundamentals of medical imaging*. Third edition. Cambridge: Cambridge University Press, 257.

Testolina, M., and Ebrahimi, T. (2021). "Review of subjective quality assessment methodologies and standards for compressed images evaluation," in *Applications of digital image processing XLIV*. Editors A. G. Tescher, and T. Ebrahimi (San Diego, United States: SPIE), 37. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11842/2597813/Review-of-subjective-quality-assessment-methodologies-and-standards-for-compressed/10.1117/12.2597813.full (Accessed December 18, 2023).

Testolina, M., Hosu, V., Jenadeleh, M., Lazzarotto, D., Saupe, D., and Ebrahimi, T. (2023). "JPEG AIC-3 dataset: towards defining the high quality to nearly visually lossless quality range," in 2023 15th International Conference on Quality of Multimedia Experience (QoMEX), Ghent, Belgium (IEEE), 55–60. Available from: https://ieeexplore.ieee.org/document/10178554/ (Accessed September 4, 2023).

Vō, D. T., Lertrattanapanich, S., and Kim, Y.-T. (2011). Low line memory visually lossless compression for color images using non-uniform quantizers. *IEEE Trans. Consum. Electron* 57 (1), 187–195. doi:10.1109/tce.2011.5735501

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13 (4), 600–612. doi:10.1109/tip.2003.819861

Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). "Multiscale structural similarity for image quality assessment," in The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA (IEEE), 1398–1402. Available from: http://ieeexplore.ieee.org/document/1292216/ (Accessed May 11, 2021).

White, S. C., and Pharoah, M. J. (2014). Oral radiology: principles and interpretation. Available from: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1167459 (Accessed October 19, 2020).

Wolski, K., Giunchi, D., Ye, N., Didyk, P., Myszkowski, K., Mantiuk, R., et al. (2018). Dataset and metrics for predicting local visible differences. *ACM Trans. Graph* 37 (5), 1–14. doi:10.1145/3196493

Wu, D., Tan, D. M., and Wu, H. R. (2003). "Visually lossless adaptive compression of medical images," in Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia Proceedings of the 2003 Joint, Singapore (IEEE), 458–463. Available from: http://ieeexplore.ieee.org/document/1292494/ (Accessed September 4, 2023).

Ye, N., Perez-Ortiz, M., and Mantiuk, R. K. (2019). "Visibility metric for visually lossless image compression," in 2019 Picture Coding Symposium (PCS), Ningbo, China IEEE), 1–5. Available from: https://ieeexplore.ieee.org/document/8954560/ (Accessed September 4, 2023).

Zemliachenko, A., Lukin, V., Ponomarenko, N., Egiazarian, K., and Astola, J. (2016). Still image/video frame lossy compression providing a desired visual quality. *Multidim Syst. Sign Process* 27 (3), 697–718. doi:10.1007/s11045-015-0333-8

Zemliachenko, A. N., Abramov, S. K., Lukin, V. V., Vozel, B., and Chehdi, K. (2015). Lossy compression of noisy remote sensing images with prediction of optimal operation point existence and parameters. *J. Appl. Remote Sens.* 9 (1), 095066. doi:10.1117/1.jrs.9.095066

Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. *IEEE Trans Image Process* 20 (8), 2378–2386. doi:10.1109/tip.2011.2109730

Zhang, Y., Reinhard, E., and Bull, D. R. (2012). "Perceptually lossless high dynamic range image compression with JPEG 2000," in 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA (IEEE), 1057–1060. Available from: http://ieeexplore.ieee.org/document/6467045/ (Accessed December 18, 2023).

Ziaei Nafchi, H., Shahkolaei, A., Hedjam, R., and Cheriet, M. (2016). Mean deviation similarity index: efficient and reliable full-reference image quality evaluator. *IEEE Access* 4, 5579–5590. doi:10.1109/access.2016.2604042

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 20 March 2024
DOI 10.3389/fpubh.2024.1365479

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Vinod Kumar,
Institute of Microbial Technology (CSIR), India
Udhaya Kumar,
Baylor College of Medicine, United States

*CORRESPONDENCE
Jianwen Chen
✉ jwchen@hutb.edu.cn
Jiangang Wang
✉ 600410@csu.edu.cn

†These authors have contributed equally to this work

RECEIVED 04 January 2024
ACCEPTED 23 February 2024
PUBLISHED 20 March 2024

CITATION
Miao R, Dong Q, Liu X, Chen Y, Wang J and Chen J (2024) A cost-effective, machine learning-driven approach for screening arterial functional aging in a large-scale Chinese population.
Front. Public Health 12:1365479.
doi: 10.3389/fpubh.2024.1365479

A cost-effective, machine learning-driven approach for screening arterial functional aging in a large-scale Chinese population

Rujia Miao[1], Qian Dong[2], Xuelian Liu[1], Yingying Chen[1], Jiangang Wang[1*†] and Jianwen Chen[2*†]

[1]Health Management Medicine Center, The Third Xiangya Hospital, Central South University, Changsha, China, [2]School of Science, Hunan University of Technology and Business, Changsha, China

Introduction: An easily accessible and cost-free machine learning model based on prior probabilities of vascular aging enables an application to pinpoint high-risk populations before physical checks and optimize healthcare investment.

Methods: A dataset containing questionnaire responses and physical measurement parameters from 77,134 adults was extracted from the electronic records of the Health Management Center at the Third Xiangya Hospital. The least absolute shrinkage and selection operator and recursive feature elimination-Lightweight Gradient Elevator were employed to select features from a pool of potential covariates. The participants were randomly divided into training (70%) and test cohorts (30%). Four machine learning algorithms were applied to build the screening models for elevated arterial stiffness (EAS), and the performance of models was evaluated by calculating the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy.

Results: Fourteen easily accessible features were selected to construct the model, including "systolic blood pressure" (SBP), "age," "waist circumference," "history of hypertension," "sex," "exercise," "awareness of normal blood pressure," "eat fruit," "work intensity," "drink milk," "eat bean products," "smoking," "alcohol consumption," and "Irritableness." The extreme gradient boosting (XGBoost) model outperformed the other three models, achieving AUC values of 0.8722 and 0.8710 in the training and test sets, respectively. The most important five features are SBP, age, waist, history of hypertension, and sex.

Conclusion: The XGBoost model ideally assesses the prior probability of the current EAS in the general population. The integration of the model into primary care facilities has the potential to lower medical expenses and enhance the management of arterial aging.

KEYWORDS

machine learning, XGBoost, arterial stiffness, physical examination, questionnaire, feature

1 Introduction

Vascular aging, regardless of the presence of atherosclerosis, is characterized by intimal and medial thickening and a loss of arterial elasticity, leading to vascular stiffness (1, 2). Population-based studies indicated that vascular aging should no longer be only considered a part of normal aging but rather influenced by industrialized lifestyle (3) and increased with the level of urbanization (4). Aortic pulse wave velocity (PWV) is considered a physiological method for quantifying arterial functional aging (5, 6) and serves as a surrogate marker of arterial stiffness, which is strongly related to cardiovascular diseases (CVDs) morbidity (7, 8).

Approximately 35% of individuals aged less than 40 years present a PWV value that was above the 90th percentile of PWV expected for their age, according to the European Reference Values Collaboration (9), and widespread PWV screening remains lacking in practice (10). Vascular aging is strongly influenced by acquired risk factors, primarily related to lifestyle choices and cardio-metabolic indicators such as smoking (11, 12), high blood pressure (13, 14), and glucose levels (15). This implies that arterial aging, as measured by PWV, can be preemptively evaluated with the aid of a suitable algorithm. This might allow for the precise identification of high-risk individuals at minimal or even zero cost, optimizing the efficiency of screening initiatives and reducing the health investment required for the vast low-risk population. We hereby reintroduce the concept of pretest probability. For example, in clinical practice, before diagnosing coronary heart disease, it is necessary to assess the disease's prior probability based on symptoms and other factors. This assessment informs the next steps in patient testing (16). Similar to this scope, our testing employs cross-sectional big data and machine learning methods to determine the pretest probability of current vascular aging.

Machine learning (ML) has been successfully employed in medicine to establish and develop accurate models (17). It outperforms conventional statistical methods by automatically training itself and improving its performance without the need for intricate programming. ML has the capacity to learn from diverse data modules and model complex relationships, resulting in more accurate predictions (18). A gradient boosting-based model for arterial stiffness assessment using clinical characteristics was constructed in a cohort of 1,672 patients with diabetes (19), demonstrating effective classification for elevated arterial stiffness (EAS) within this specific group. Based on our speculation that the prevalence of elevated PWV would be higher in individuals with diabetes, and considering the distinct parameters and contributing features of the model in this population, we realized that the model for the general population would likely differ substantially.

By leveraging a vast dataset gathered from a substantial cohort of physical examinees, encompassing Pulse Wave Velocity (PWV) metrics, along with detailed insights into lifestyle patterns and fundamental clinical attributes, we employed machine learning (ML) techniques to craft a streamlined and economical screening model for arterial stiffness in the general populace. This approach enables more precise pinpointing of high-risk demographics. Such a strategy may significantly reduce the likelihood of overlooking potential diagnoses and curtail medical squandering, enhancing the overall efficiency and effectiveness of health screenings for arterial aging.

2 Methods

2.1 Study population and data source

The retrospective dataset was extracted from the electronic healthcare records of the health management center of the Third Xiangya Hospital. The records collected spanned from 2015 to 2021 and constituted the original set. The dataset consisted of 77,191 physical examinees who underwent a brachial-ankle pulse wave velocity (baPWV) test at the health management center. The exclusion criteria were as follows: (1) individuals who were unable to sign informed consent; (2) patients with a diagnosis of end-stage renal disease and aortopathy. Samples with massive missing data or excessive outliers whose irrationality was justified by the clinician were deleted, resulting in a final dataset of 77,134 individuals. All participants signed the informed consent form. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Third Xiangya Hospital (No 2020-S609).

2.2 Feature characteristics and definitions

The dataset included a total of 82 features, which encompassed various types of information. These features included physical examination measurements such as body mass index (BMI), waist circumference (WC), and systolic/diastolic blood pressure (SBP/DBP). Laboratory indicators were also present in the dataset and included measurements such as fasting blood sugar (FBS), lipid profile, and renal function markers. Additionally, questionnaire information was collected, including previous diagnoses of diseases, lifestyle factors (diet preference, smoking, alcohol consumption, exercise habits, working and sedentary time, and sleep status), and health literacy regarding basic medical knowledge. For specific definitions within the dataset, alcohol consumption was defined as consuming more than 50 g of alcohol per week. The current smoking status was defined as smoking more than one cigarette per day. The baPWV measurements were obtained using a non-invasive measurement system called VP-2000 manufactured by Colin Co Ltd., Komaki, Japan. Trained medical staff performed the measurements in a room after the subjects had rested in a supine position for 10 min. Four pneumatic cuffs were attached to the bilateral arms and ankles to obtain pulse waves. The baPWV was automatically calculated using the formula (La-Lb)/Tba, where La represents the distance between the heart and ankle, Lb represents the distance between the heart and brachium, and Tba represents the time difference between the initial increase in the brachial waveform and that in the ankle waveform. Elevated arterial stiffness (EAS) was considered as baPWV≥1,400 cm/s (20).

2.3 Data processing and statistical analysis

Nineteen features in the original dataset are related to detailed questions about smoking, drinking, exercising, and work intensity, including types and frequency. To reduce the redundancy of the data, these data are not included in the analysis for now. If factors such as smoking and drinking prove important in later analyses, this part of the data will be specifically analyzed. However, since factors such as smoking and drinking were not considered highly important in subsequent analyses, no further specific analysis was conducted on these questions. If a sample has more than 30% of its variables missing, then the sample is deleted. Therefore, a total of 57 samples were deleted. For the remaining missing data, single-value imputation was used to handle features. Ultimately, 63 features and 77,134 samples were retained. The entire dataset was randomly divided into a training set (70%, 53,993 samples) and a test set (30%, 23,141 samples). In the test set, there were 13,600 samples for the non-EAS category and 9,541 samples for the EAS category.

Continuous measurement data conforming to the normal distribution were expressed as mean ± standard deviation, otherwise, the quartile was adopted, and the difference between categorical features was calculated using the chi-squared test. The difference of $p < 0.05$ on both sides was considered statistically significant.

2.4 Feature selection

In the field of machine learning, feature selection can eliminate irrelevant or redundant features, thereby reducing the number of features, improving model accuracy, and reducing runtime. Feature selection methods are mainly divided into three categories: filter, embedded, and wrapper methods.

Filter method: Features that did not show significant differences across categories meant they had no use for the prediction target and would be deleted.

Embedded: The features were input into the Lasso model for training after the filter method. The L1 regularization parameters were tuned using 10-fold cross-validation, and the coefficient would be shrunken to zero if their features were not important.

Wrapper: After filtering through Lasso regression, if there were still quite a number of features left, the wrapper method could be used for further selection. The working principle of recursive feature elimination (RFE) is to recursively remove features and build a model on the remaining features (in this case, the Lightweight Gradient Elevator (LGBM) is chosen), thereby determining which combinations of features contribute more significantly to the prediction results. Regarding the choice of the number of features, as we continuously increase the number of features, the AUC score of RFE-LGBM gradually increases and tends to stabilize. When the number of features reaches N, the AUC score reaches its maximum value. Adding more features will not enhance the model's predictive capability, so ultimately, N features are determined for model building through RFE-LGBM.

Based on the common usage and the situation of data being selected at each step, we ultimately employed these three methods.

2.5 Parameter optimization and model evaluation

Four machine learning algorithms (logistic regression [LR], random forest [RF], extreme gradient boosting [XGBoost], and light gradient boosting machine [LightGBM]) were used to develop predictive models on the training data. Weights of different classes were assigned by setting parameters in the trained models to deal with data imbalance. The parameters of LR, RF, and LightGBM models were class_weight = "balanced," and XGBoost's was scale_pos_weight = "ratio of majority and minority class." As to the other parameters that required adjustment, grid searching was adopted in the LR model and the Bayesian Optimization Algorithm was applied to the RF, XGBoost, and LightGBM models. To assess the model's optimization and improve its generalization ability, 10-fold cross-validation was performed. The model's performance was evaluated by calculating the accuracy, sensitivity, and specificity, and the distinguishing abilities of the risk assessment model were evaluated with the area under the receiver operator characteristic (AUC). The overview of the proposed ML algorithms is shown in Figure 1.

2.6 Feature importance ranking

Shapley additive explanations (SHAP) value was defined as the average marginal contribution of a feature value across all possible feature coalitions. It provides insights into the influence of each feature on individual samples, showcasing both positive and negative effects.

ML was implemented in Python (v 3.6) using the sklearn (v. 0.24.1) and XGBoost (v. 1.3.3) packages. Bayesian optimization based on the TPE toolbox (bayes_opt v. 1.2.0) was used to tune hyperparameters for learning algorithms when the best combination of parameters yielded a low model performance. SHAP (V 0.41.0) was for explainable ML.

3 Results

3.1 Characteristics of the study population

A total of 77,134 subjects were included, with a mean age of 48.6 ± 11.4 years, and 54.9 ± 10.6 years in subjects with EAS. The sample prevalence of EAS was 40.8%. Significant differences ($p < 0.01$) were observed in age, BMI, WC, SBP/DBP, indicators concerning renal function, FBS, total cholesterol (TC), triglyceride (TG), high−/low-density lipoprotein cholesterol (HDL-c/LDL-c) (Table 1), sex, history of cardiovascular diseases, smoking, exercise, work intensity, sedentary time, sleeping quality and time, 16 items of dietary habits, 9 items of negative emotions, and 12 items of health literacy (Table 2) between EAS and non-EAS.

3.2 Feature selection

Three health literacy items (active in the acquisition of medical knowledge and awareness of normal pulse rate/total cholesterol) with no statistical difference were removed from further analysis, and 60 features were input for subsequent feature selection.

The Lasso regression ranked the features based on their parameter values, and the bottom 25 features (Figure 2A) were visually observed. Subsequently, 21 features with zero parameter values were deleted, including LDL-c, history of other cardiovascular diseases (excluding hypertension and diabetes), awareness of normal salt intake/FBS/TG values, sunlight exposure, lack of enthusiasm, impatience, difficulty concentrating, nervousness, upset, meals on time, gluttony, frequency

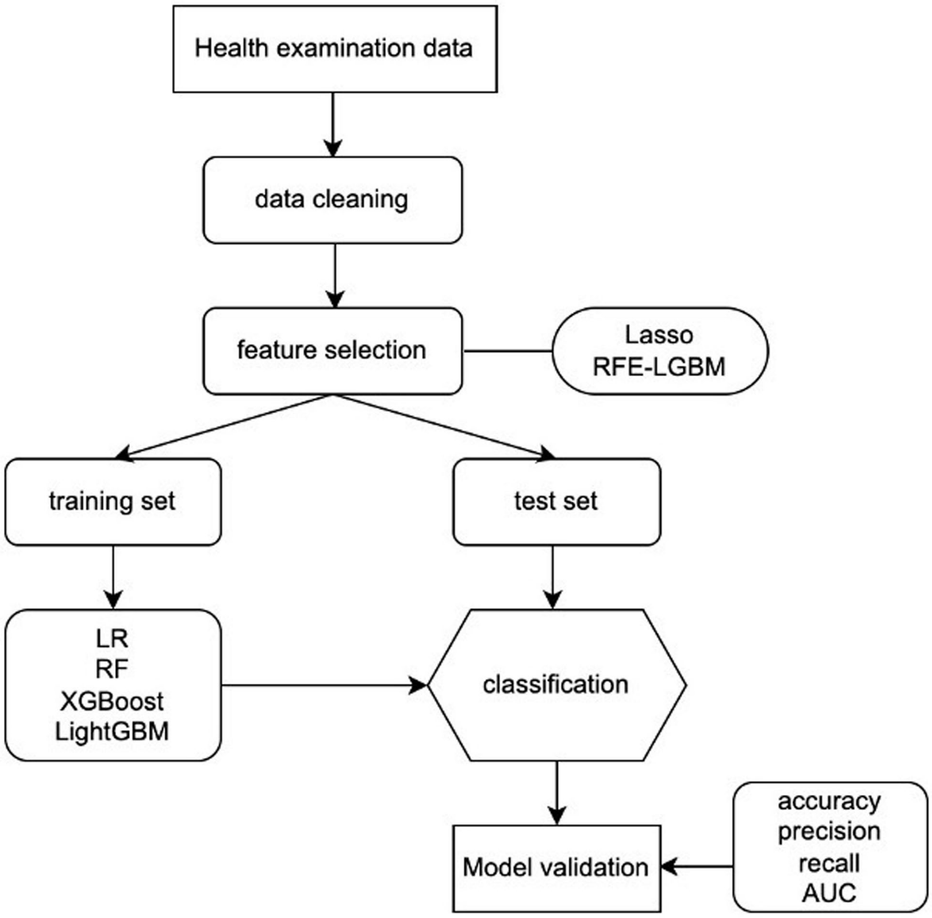


FIGURE 1
Machine learning flowchart of this study. LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting; LightGBM, Light gradient boosting machine; RFE-LGBM, recursive feature elimination-Lightweight Gradient Elevator; AUC, area under the receiver operating characteristic curve.

TABLE 1 Features of the participants that are continuous data in primary settings.

Feature	All (*n* = 77,134)	Non-AS (*n* = 45,688)	EAS (*n* = 31,446)	*p* value
Age	48.6 ± 11.4	44.2 ± 9.7	54.9 ± 10.6	<0.001
Body mass index	24.6 ± 3.2	24.2 ± 3.2	25.1 ± 3.1	<0.001
Waist circumference	84.0 ± 9.804	82.3 ± 9.9	86.4 ± 9.1	<0.001
Systolic blood pressure	125.3 ± 16.8	118.0 ± 12.7	135.9 ± 16.3	<0.001
Diastolic blood pressure	77.2 ± 11.5	73.1 ± 9.8	83.1 ± 11.3	<0.001
Blood urea nitrogen	4.9 ± 1.3	4.8 ± 1.2	5.1 ± 1.4	<0.001
Serum creatinine	74.0 ± 18.8	72.8 ± 16.3	75.9 ± 21.9	<0.001
Serum uric acid	347.9 ± 89.6	339.9 ± 89.2	359.6 ± 88.9	<0.001
Fasting blood sugar	5.7 ± 1.5	5.4 ± 1.0	6.1 ± 1.9	<0.001
Total cholesterol	5.1 ± 1.0	5.0 ± 0.9	5.3 ± 1.1	<0.001
Triglyceride	2.0 ± 1.9	1.8 ± 1.7	2.3 ± 2.2	<0.001
High-density lipoprotein cholesterol	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.3	<0.001
Low-density lipoprotein cholesterol	2.9 ± 0.9	2.9 ± 0.8	3.0 ± 0.9	<0.001

EAS, elevated arterial stiffness.

TABLE 2 Features of the participants that are categorical data in primary settings.

Features	Chi value	p-value	Features	Chi value	p-value
Sex	494.529	<0.001	Irritableness	408.974	<0.001
History of hypertension	6905.432	<0.001	Nervousness	523.143	<0.001
History of diabetes	1913.659	<0.001	Anxiousness	263.128	<0.001
History of other CVD	586.573	<0.001	Impatience	306.742	<0.001
Meals on time	828.782	<0.001	Lack of enthusiasm	562.609	<0.001
Eat midnight snack	2184.694	<0.001	Upset	374.698	<0.001
Gluttony	36.759	<0.001	Depression	513.749	<0.001
Dinner party	186.321	<0.001	Difficulty concentrating	378.7	<0.001
Drink milk	526.563	<0.001	Sleep quality	82.116	<0.001
Eating eggs	103.121	<0.001	Sleep duration	264.814	<0.001
Eat beans	13.101	0.001	Active acquisition of medical knowledge	0.2	0.655
Eat fruit	302.005	<0.001	Fasten seat belt	791.405	<0.001
Eat vegetables	415.069	<0.001	Observed stools	114.9	<0.001
Eat meat	100.03	<0.001	Self-measurement of BP/HR	4761.237	<0.001
Eat fatty meat	119.924	<0.001	Take first-aid medicine along	1136.878	<0.001
Eat animal offal	190.865	<0.001	Sunlight exposure	771.451	<0.001
Eat fish	178.008	<0.001	Awareness of normal BP	154.105	<0.001
Drink coffee	487.353	<0.001	Awareness of normal body temperature	13.242	<0.001
Sugary drinks	1276.974	<0.001	Awareness of normal pulse	0.01	0.919
Smoking	207.578	<0.001	Awareness of normal salt intake	79.905	<0.001
Drink alcohol	131.602	<0.001	Awareness of normal BMI	85.353	<0.001
Exercise	81.124	<0.001	Awareness of normal WC	25.583	<0.001
Work intensity	2985.317	<0.001	Awareness of normal FBS	156.494	<0.001
Sedentary duration	173.991	<0.001	Awareness of normal triglyceride	31.048	<0.001
Depressed	587.383	<0.001	Awareness of normal TC	1.525	0.217

BP, blood pressure; HR, heart rate; BMI, body mass index; WC, waist circumference; FBS, fasting blood sugar; TC, total cholesterol.

of consumption of eggs, vegetables, fat, animal offal, coffee, sugary drinks, sedentary duration, and dinner party.

The 39 reserved features were considered too much, and we adopted the LGBM model to train them further. With each addition, AUC scores for the test set were recorded. As depicted in Figure 2B, there was a discernible decrease in the score as we integrated 23 features one by one; continuous feature addition not only raised computational costs but also failed to enhance model performance. Consequently, we discarded the 16 non-contributory features. The remaining 23 significant features were sex, age, history of hypertension, SBP, DBP, BMI, WC, TC, TG, HDL-c, FBS, serum uric acid, serum creatinine, urea nitrogen, exercise, awareness of normal BP, the consumption frequency of items such as fruit, milk, bean products, alcohol, and smoking, work intensity, and irritableness. An important observation was the strong linear correlation between BMI and WC ($r = 0.86$) and between SBP and DBP ($r = 0.78$) (Supplementary Figure 1). For efficiency, we retained only WC and SBP, leading to a final 21 features for ML evaluation. It is noteworthy that we also developed a model that prioritizes features that are cost-effective and easily obtainable. Therefore, lab indicators were omitted, leaving us with 14 salient features.

3.3 Model performance and feature importance ranking

The performance of four classifiers—logistic regression (LR), random forest (RF), XGBoost, and LightGBM—across the training and test sets is detailed in Table 3 (for the 14-feature model) and Supplementary Table 1 (for the 21-feature model). Intriguingly, the incorporation of laboratory indicators in the 21-feature models resulted in only marginal improvements in performance compared to the 14-feature models. For instance, the AUC test scores showed negligible differences: LR (0.8746 vs. 0.8706), RF (0.8697 vs. 0.8695), XGBoost (0.8754 vs. 0.8710), and LGBM (0.8732 vs. 0.8702). This increment came at a significant cost to the testers, suggesting that classifiers with only 14 features are more suitable for broad public applications. In terms of overall performance, the XGBoost model

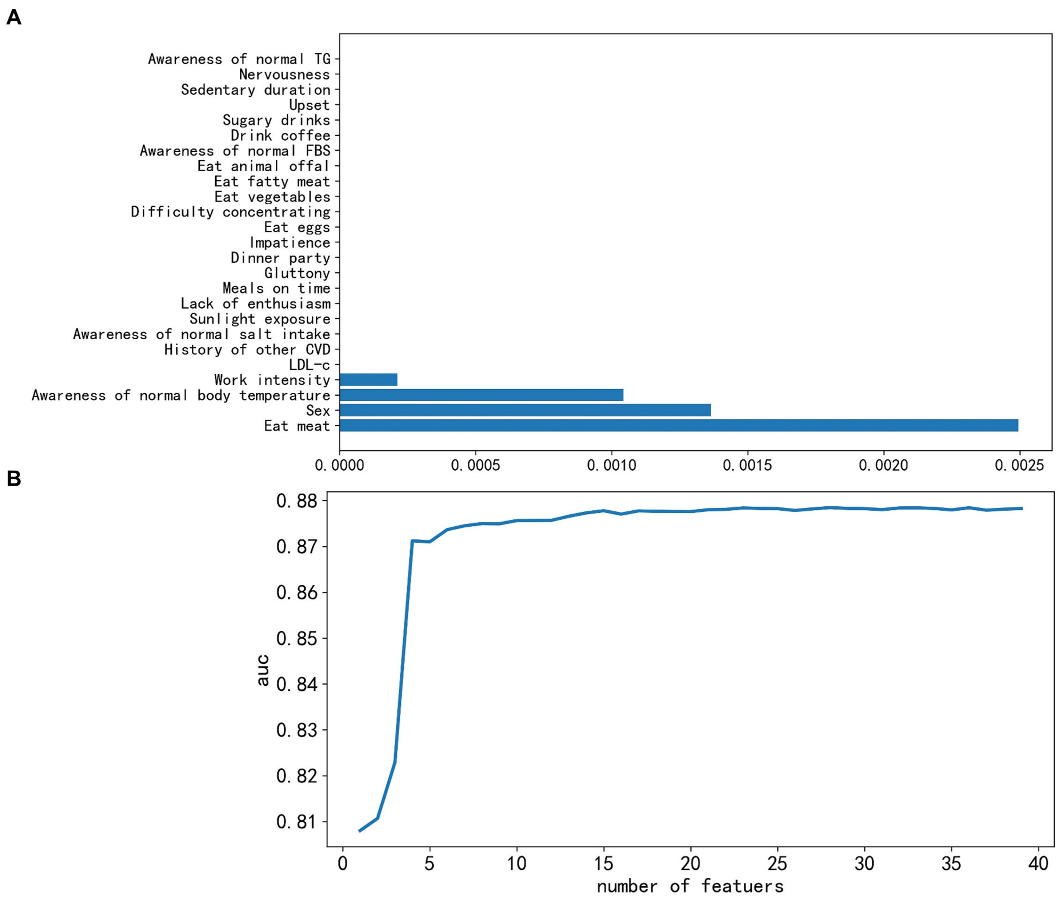


FIGURE 2
Feature selection successively using the Lasso algorithm and recursive feature elimination-lightweight gradient elevator (RFE-LGBM). **(A)** Coefficients profile of bottom 25 features resulting from the Lasso algorithm, where features with zero coefficients were removed. **(B)** Five-fold cross-validation criterion of RFE based on LGBM, 23 features were optimal and finally kept. AUC, area under the receiver operating characteristic curve.

TABLE 3 The results of classification algorithms based on 14 costless features.

Model	Accuracy_train	Accuracy_test	Sensitivity	Specificity	AUC_train	AUC_test
LR	0.7839	0.7877	0.78	0.72	0.8678	0.8706
RF	0.7852	0.7845	0.77	0.72	0.8678	0.8685
XGB	0.7870	0.7878	0.78	0.73	0.8722	0.8710
LGBM	0.7967	0.7865	0.77	0.73	0.8702	0.8702

LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting; LightGBM, Light gradient boosting machine; RFE-LGBM, recursive feature elimination-Lightweight Gradient Elevator; AUC, area under the receiver operating characteristic curve.

modestly outshone the other classifiers in multiple metrics, including accuracy (0.7878), sensitivity (0.78), specificity (0.73), and AUC (0.8710) on the test set.

Feature importance was assessed in the XGBoost model, and the top 10 most influential features were identified as SBP, age, history of hypertension, FBS, TG, sex, smoking, fruit consumption, awareness of normal BP, and exercise, as shown in Figure 3. Figure 4 visually presents the SHAP values corresponding to the 14 most important features. The *x*-axis represents the SHAP values, while the *y*-axis lists the features; each dot symbolizes a sample. The color scale varies from low (blue) to high (red), indicating the feature values. A positive SHAP value indicates contributing to EAS formation, whereas a negative value suggests inhibition. The top 5 features in terms of importance were SBP, age, WC, history of hypertension, and sex.

4 Discussion

In this study, we established four ML models for arterial stiffness screening in a large-scale physical examination population using easily accessible measurements and questionnaire indicators. Among them, XGBoost demonstrated superior performance in validation. Interestingly, while including optional laboratory features led to a

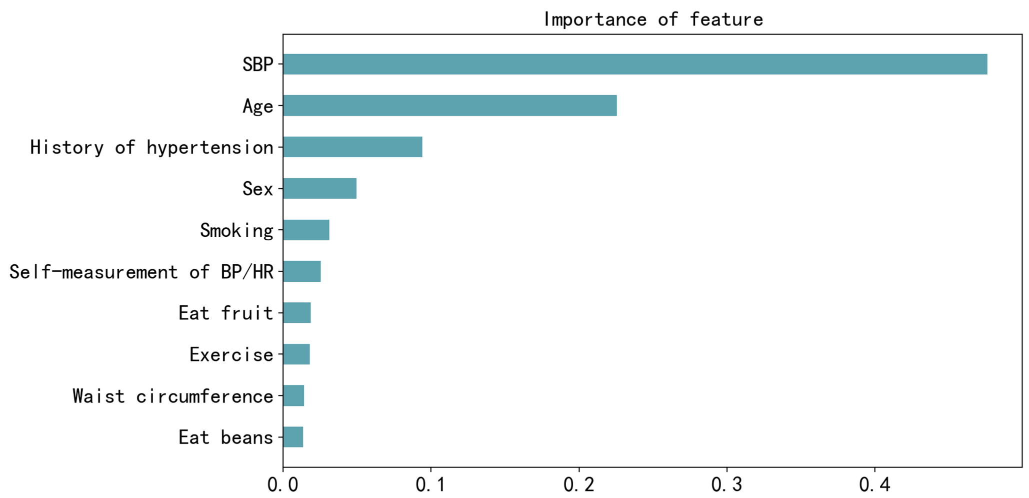


FIGURE 3
Feature importance plot of the best-performing extreme gradient boosting model measured by F-score. SBP, systolic blood pressure; BP, blood pressure; HR, heart rate.

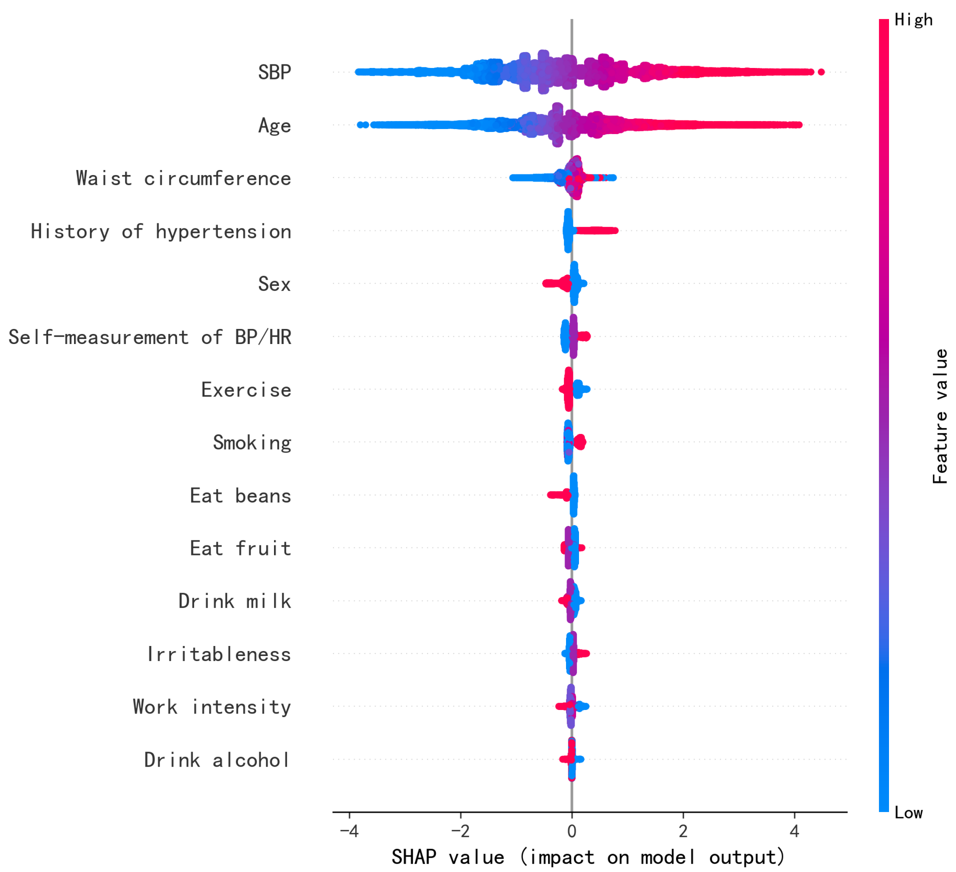


FIGURE 4
Feature weight sorting by Shapley additive explanations (SHAP) value. SBP, systolic blood pressure; BP, blood pressure; HR, heart rate.

minor enhancement in model performance, the gains may not justify the increased complexity and costs. This model holds promise for predicting early-stage vascular aging, particularly in regions with elevated epidemiological risks.

Traditional CVD risk factors played a noticeable role in affecting arterial stiffness (AS). A meta-analysis (21) highlighted age and blood pressure as the primary determinants of arterial stiffness. Additionally, history of hypertension and waist circumference were also influential factors. Regarding the more costly and invasive laboratory indices, abnormalities in glucose metabolism (22), triglyceride (23), HDL-c, uric acid (24), and glomerular filtration rate (GFR) (25) as determined by serum creatinine have all been linked to varying degrees with arterial stiffness. Our model, which incorporated 21 features—including both questionnaire items and select laboratory indices—aligns well with findings from previous studies. However, one exception was total cholesterol (26, 27), where its potential anti-stiffening effects remain a matter of debate. A notable model for assessing arterial stiffness is the SAGE scoring system. This model, derived from a cohort of 3,943 outpatients and using multiple logistic regression, incorporated variables such as SBP, age, glycemia, and GFR. Impressively, it achieved a 0.77 ROC in its validation cohort (28).

Most traditional models related to cardiovascular risk integrate both laboratory and clinical parameters (29–31). Securing these variables often entails financial expenditures, making the broad application of such models somewhat limited in the general population (32). In our earlier exploration, we compared results derived from both self-assessment items and blood tests with those garnered exclusively from self-assessment features. Intriguingly, the differences were minimal, which we attribute primarily to the overwhelming influence of factors such as blood pressure, age, and waist circumference. A notable strength of our study is its reliance on non-invasive, easily obtainable indicators that come at a minimal cost, which are sufficient to craft a scalable EAS model. However, using a vast number of these indicators poses challenges. These raw datasets tend to be non-linear, intricate, and possibly inter-correlated or influenced by a plethora of confounding variables (33). This complexity could lead to significant deviations and diminished accuracy when building models. Given this backdrop, ML algorithms emerge as a preferable approach, offering a remedy to these pitfalls. We initially employed LASSO and RFE-LGBM to filter out superfluous features and avert overfitting caused by collinearity. Ultimately, leveraging four distinct ML algorithms to assess EAS proved more effective than relying on a single model. Notably, XGBoost marginally surpassed the other three in metrics such as accuracy, sensitivity, specificity, and AUC.

Contrary to conventional methodologies such as logistic regression, which operate under the presumption of variables being independent and purely linear (34), XGBoost adopts a non-parametric approach. It integrates a regularized loss function and marries gradient-boosting algorithms with decision trees, preserving inter-feature correlations (35). We theorize that this unique characteristic underpins XGBoost's standout performance.

Another study that focused on assessing the risk of EAS in diabetic patients reported ROC values of 0.928 and 0.821 using a gradient-boosting algorithm for a discovery dataset of 760 Chinese individuals and a validation dataset of 912 Japanese individuals, respectively (19). In comparison, our study's significantly larger dataset ensured a more normal distribution and, consequently, more precise analyses. The ROC values in our study were 0.8826 and 0.8754 for the training and test datasets, respectively. The closer alignment of these values suggests a superior model fit (36). Furthermore, by targeting the general population rather than specialized groups, our model may boast broader applicability.

Currently, PWV assessment is used for patients suspected or diagnosed with CVD and individuals undergoing self-funded health examinations in Chinese hospitals. However, it is challenging to implement PWV screening for the general population outside of hospital settings to enhance the management of CVD. Moreover, the subject's general physical examination reveals the need for a more standardized approach in developing personalized medical examination programs, particularly concerning vessel checks. Currently, the formulation of these programs is largely influenced by the individual's economic status, resulting in either missed diagnoses due to inadequate testing or unnecessary resource wastage through over-testing. For instance, the number of physical examinees in China surged from 444 to 549 million in 3 years, according to the China Health Statistics Yearbook, with the average cost per examination amounting to 755.8¥ (37). This typically includes visceral color ultrasound and selected blood tests such as lipid, glucose, hepatic and renal function, and complete blood count (38). However, this cost does not necessarily cover more specialized, expensive tests such as arterial stiffness assessments, CT scans, or gastrointestinal endoscopies. This suggests that early identification measures in primary care remain inadequate, even for those who undergo routine physicals, let alone those who have not visited a hospital. Therefore, incorporating a convenient and easily accessible assessment model into primary care could offer a cost-free initial screening for the general population. This would allow healthcare providers to tailor preventive strategies to those at higher risk. Machine learning models serve as a viable pathway for achieving this, particularly for asymptomatic individuals. At the current stage, the process of transforming a model derived from the research process into a practical clinical tool involves a productization process. This includes placing the model in the cloud or making it into a web-based version, accessible from anywhere. It also involves connecting certain data links, allowing data to be transferred from medical devices to the servers that host the model. We plan to make this model into an easily accessible web-based version.

4.1 Limitations

Our study is not free from limitations. First, given its cross-sectional design, it is not possible to infer causal relationships from the data. However, previous studies had established definite causality that we could follow. Our initial intention for the research was to assess the current status of arterial aging, thus making the study design reasonable. Second, the standards for evaluating certain lifestyle factors, such as fruit and bean consumption or exercise habits, were not unified and objectively quantified. This could potentially decrease the prediction's precision. Finally, since the study relies on data from physical examinations conducted in China, there may be restrictions on the applicability of the results to other ethnicities or cultures.

5 Conclusion

The primary aim of our research was to provide an effective model for evaluating the risk of early Arterial Stiffening (EAS) using only questionnaires and physical measurements. Overall, we found that the model's diagnostic accuracy and assessment capabilities were satisfactory. Incorporating this risk assessment model into primary healthcare settings could significantly improve the prevention and management of arterial aging across the broader populace.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Third Xiangya Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

RM: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. QD: Formal analysis, Writing – review & editing. XL: Data curation, Investigation, Writing – review & editing. YC: Data curation, Investigation, Writing – review & editing. JW: Project administration, Supervision, Writing – review & editing. JC: Conceptualization, Formal analysis, Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Wisdom Accumulation and Talent Cultivation Project of the Third Xiangya Hospital of Central South University, the Natural Science Foundation of Hunan (2021JJ40941).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1365479/full#supplementary-material

References

1. Wang JC, Bennett M. Aging and atherosclerosis: mechanisms, functional consequences, and potential therapeutics for cellular senescence. *Circ Res*. (2012) 111:245–59. doi: 10.1161/CIRCRESAHA.111.261388

2. Zhou YF, Wang Y, Wang G, Zhou Z, Chen S, Geng T, et al. Association between statin use and progression of arterial stiffness among adults with high atherosclerotic risk. *JAMA Netw Open*. (2022) 5:e2218323. doi: 10.1001/jamanetworkopen.2022.18323

3. Nowak KL, Rossman MJ, Chonchol M, Seals DR. Strategies for achieving healthy vascular aging. *Hypertension*. (2018) 71:389–402. doi: 10.1161/HYPERTENSIONAHA.117.10439

4. Lemogoum D, Ngatchou W, Janssen C, Leeman M, Van Bortel L, Boutouyrie P, et al. Effects of hunter-gatherer subsistence mode on arterial distensibility in Cameroonian pygmies. *Hypertension*. (2012) 60:123–8. doi: 10.1161/HYPERTENSIONAHA.111.187757

5. Calimport SRG, Bentley BL, Stewart CE, Pawelec G, Scuteri A, Vinciguerra M, et al. To help aging populations, classify organismal senescence. *Science*. (2019) 366:576–8. doi: 10.1126/science.aay7319

6. Aging Biomarker ConsortiumZhang L, Guo J, Liu Y, Sun S, Liu B, et al. A framework of biomarkers for vascular aging: a consensus statement by the aging biomarker consortium. *Life Med*. (2023) 2:lnad033. doi: 10.1093/lifemedi/lnad033,

7. Mitchell GF, Hwang SJ, Vasan RS, Larson MG, Pencina MJ, Hamburg NM, et al. Arterial stiffness and cardiovascular events: the Framingham heart study. *Circulation*. (2010) 121:505–11. doi: 10.1161/CIRCULATIONAHA.109.886655

8. Vlachopoulos C, Terentes-Printzios D, Laurent S, Nilsson PM, Protogerou AD, Aznaouridis K, et al. Association of estimated pulse wave velocity with survival: a secondary analysis of SPRINT. *JAMA Netw Open*. (2019) 2:e1912831. doi: 10.1001/jamanetworkopen.2019.12831

9. Reference Values for Arterial Stiffness' Collaboration. Determinants of pulse wave velocity in healthy people and in the presence of cardiovascular risk factors: 'establishing normal and reference values'. *Eur Heart J*. (2010) 31:2338–50. doi: 10.1093/eurheartj/ehq165

10. Cunha PG, Boutouyrie P, Nilsson PM, Laurent S. Early vascular ageing (EVA): definitions and clinical applicability. *Curr Hypertens Rev*. (2017) 13:8–15. doi: 10.2174/1573402113666170413094319

11. Mahmud A, Feely J. Effect of smoking on arterial stiffness and pulse pressure amplification. *Hypertension*. (2003) 41:183–7. doi: 10.1161/01.hyp.0000047464.66901.60

12. Damay VA, Setiawan , Lesmana R, Akbar MR, Lukito AA, Tarawan VM, et al. Aerobic exercise versus electronic cigarette in vascular aging process: first histological insight. *Int J Vasc Med*. (2023) 2023:8874599. doi: 10.1155/2023/8874599

13. Keehn L, Mangino M, Spector T, Chowienczyk P, Cecelja M. Relation of pulse wave velocity to contemporaneous and historical blood pressure in female twins. *Hypertension*. (2023) 80:361–9. doi: 10.1161/HYPERTENSIONAHA.122.19311

14. Li A, Yan J, Zhao Y, Yu Z, Tian S, Khan AH, et al. Vascular aging: assessment and intervention. *Clin Interv Aging*. (2023) 18:1373–95. doi: 10.2147/cia.S423373

15. Stacey RB, Bertoni AG, Eng J, Bluemke DA, Hundley WG, Herrington D. Modification of the effect of glycemic status on aortic distensibility by age in the multi-ethnic study of atherosclerosis. *Hypertension*. (2010) 55:26–32. doi: 10.1161/HYPERTENSIONAHA.109.134031

16. Taylor AJ, Cerqueira M, Hodgson JM, Mark D, Min J, O'Gara P, et al. ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 appropriate use criteria for cardiac computed tomography. A report of the American College of Cardiology Foundation appropriate use criteria task force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. *J Am Coll Cardiol*. (2010) 56:1864–94. doi: 10.1016/j.jacc.2010.07.005

17. Barrios JP, Tison GH. Advancing cardiovascular medicine with machine learning: progress, potential, and perspective. *Cell Rep Med*. (2022) 3:100869. doi: 10.1016/j.xcrm.2022.100869

18. Marsland S. *Machine learning: an algorithmic perspective*. Florida: CRC press (2015).

19. Li Q, Xie W, Li L, Wang L, You Q, Chen L, et al. Development and validation of a prediction model for elevated arterial stiffness in Chinese patients with diabetes using machine learning. *Front Physiol*. (2021) 12:714195. doi: 10.3389/fphys.2021.714195

20. Yamashina A, Tomiyama H, Arai T, Hirose K-i, Koji Y, Hirayama Y, et al. Brachial-ankle pulse wave velocity as a marker of atherosclerotic vascular damage and cardiovascular risk. *Hypertens Res*. (2003) 26:615–22. doi: 10.1291/hypres.26.615

21. Cecelja M, Chowienczyk P. Dissociation of aortic pulse wave velocity with risk factors for cardiovascular disease other than hypertension: a systematic review. *Hypertension*. (2009) 54:1328–36. doi: 10.1161/HYPERTENSIONAHA.109.137653

22. Tomiyama H, Hashimoto H, Hirayama Y, Yambe M, Yamada J, Koji Y, et al. Synergistic acceleration of arterial stiffening in the presence of raised blood pressure and raised plasma glucose. *Hypertension*. (2006) 47:180–8. doi: 10.1161/01.HYP.0000198539.34501.1a

23. Wang L, Zhi F, Gao B, Ni J, Liu Y, Mo X, et al. Association between lipid profiles and arterial stiffness: a secondary analysis based on a cross-sectional study. *J Int Med Res*. (2020) 48:300060520938188. doi: 10.1177/0300060520938188

24. Saijo Y, Utsugi M, Yoshioka E, Horikawa N, Sato T, Gong YY, et al. Relationships of C-reactive protein, uric acid, and glomerular filtration rate to arterial stiffness in Japanese subjects. *J Hum Hypertens*. (2005) 19:907–13. doi: 10.1038/sj.jhh.1001913

25. Yoshida M, Tomiyama H, Yamada J, Koji Y, Shiina K, Nagata M, et al. Relationships among renal function loss within the normal to mildly impaired range, arterial stiffness, inflammation, and oxidative stress. *Clin J Am Soc Nephrol*. (2007) 2:1118–24. doi: 10.2215/CJN.01880507

26. Liu J, Fan F, Liu B, Li K, Jiang Y, Jia J, et al. Association between remnant cholesterol and arterial stiffness in a Chinese community-based population: a cross-sectional study. *Front Cardiovasc Med*. (2022) 9:993097. doi: 10.3389/fcvm.2022.993097

27. Chen H, Chen Y, Wu W, Cai Z, Chen Z, Yan X, et al. Total cholesterol, arterial stiffness, and systolic blood pressure: a mediation analysis. *Sci Rep*. (2021) 11:1330. doi: 10.1038/s41598-020-79368-x

28. Xaplanteris P, Vlachopoulos C, Protogerou AD, Aznaouridis K, Terentes-Printzios D, Argyris AA, et al. A clinical score for prediction of elevated aortic stiffness: derivation and validation in 3943 hypertensive patients. *J Hypertens*. (2019) 37:339–46. doi: 10.1097/HJH.0000000000001904

29. Liu X, Shen P, Zhang D, Sun Y, Chen Y, Liang J, et al. Evaluation of atherosclerotic cardiovascular risk prediction models in China: results from the CHERRY study. *JACC Asia*. (2022) 2:33–43. doi: 10.1016/j.jacasi.2021.10.007

30. Xing X, Yang X, Liu F, Li J, Chen J, Liu X, et al. Predicting 10-year and lifetime stroke risk in Chinese population. *Stroke*. (2019) 50:2371–8. doi: 10.1161/STROKEAHA.119.025553

31. Hippisley-Cox J. C Coupland, and P brindle, development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ*. (2017) 357:j2099. doi: 10.1136/bmj.j2099

32. Duval S, Van't JR. Hof, LM Steffen, and RV Luepker, estimation of cardiovascular risk from self-reported knowledge of risk factors: insights from the Minnesota heart survey. *Clin Epidemiol*. (2020) 12:41–9. doi: 10.2147/clep.S219708

33. Batool S, Taj IA, Ghafoor M. Ejection fraction estimation from echocardiograms using optimal left ventricle feature extraction based on clinical methods. *Diagnostic*. (2023) 13:2155. doi: 10.3390/diagnostics13132155

34. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. (2011) 18:1099–104. doi: 10.1111/j.1553-2712.2011.01185.x

35. Thomas J, Hepp T, Mayr A, Bischl B. Probing for sparse and fast variable selection with model-based boosting. *Comput Math Methods Med*. (2017) 2017:1421409. doi: 10.1155/2017/1421409

36. Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, et al. ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol Evol*. (2014) 5:1198–205. doi: 10.1111/2041-210X.12261

37. Gao X, Wu F, Chu J, Wu S, Zeng Q. Development trend of physical examination volume and revenue in health-checkup institutions in China from 2010 to 2019. *Chin J Health Manage*. (2023) 17:96–101.

38. Zhang W, Chen Q, Yuan Z, Liu J, Du Z, Tang F, et al. A routine biomarker-based risk prediction model for metabolic syndrome in urban Han Chinese population. *BMC Public Health*. (2015) 15:64. doi: 10.1186/s12889-015-1424-z

Frontiers in Public Health

TYPE Methods
PUBLISHED 27 March 2024
DOI 10.3389/fpubh.2024.1342937

Check for updates

OPEN ACCESS

EDITED BY
Tetyana Chumachenko,
Kharkiv National Medical University, Ukraine

REVIEWED BY
Gilbert Yong San Lim,
SingHealth, Singapore
John Plodinec,
Independent Researcher, Aiken, SC,
United States

*CORRESPONDENCE
Viacheslav Moskalenko
✉ v.moskalenko@cs.sumdu.edu.ua

RECEIVED 22 November 2023
ACCEPTED 15 March 2024
PUBLISHED 27 March 2024

CITATION
Moskalenko V and Kharchenko V (2024) Resilience-aware MLOps for AI-based medical diagnostic system.
Front. Public Health 12:1342937.
doi: 10.3389/fpubh.2024.1342937

Resilience-aware MLOps for AI-based medical diagnostic system

Viacheslav Moskalenko[1]* and Vyacheslav Kharchenko[2]

[1]Department of Computer Science, Faculty of Electronics and Information Technologies, Sumy State University, Sumy, Ukraine, [2]Department of Computer Systems, Network and Cybersecurity, Faculty of Radio-Electronics, Computer Systems and Infocommunications, National Aerospace University "KhAI", Kharkiv, Ukraine

Background: The healthcare sector demands a higher degree of responsibility, trustworthiness, and accountability when implementing Artificial Intelligence (AI) systems. Machine learning operations (MLOps) for AI-based medical diagnostic systems are primarily focused on aspects such as data quality and confidentiality, bias reduction, model deployment, performance monitoring, and continuous improvement. However, so far, MLOps techniques do not take into account the need to provide resilience to disturbances such as adversarial attacks, including fault injections, and drift, including out-of-distribution. This article is concerned with the MLOps methodology that incorporates the steps necessary to increase the resilience of an AI-based medical diagnostic system against various kinds of disruptive influences.

Methods: *Post-hoc* resilience optimization, *post-hoc* predictive uncertainty calibration, uncertainty monitoring, and graceful degradation are incorporated as additional stages in MLOps. To optimize the resilience of the AI based medical diagnostic system, additional components in the form of adapters and meta-adapters are utilized. These components are fine-tuned during meta-training based on the results of adaptation to synthetic disturbances. Furthermore, an additional model is introduced for *post-hoc* calibration of predictive uncertainty. This model is trained using both in-distribution and out-of-distribution data to refine predictive confidence during the inference mode.

Results: The structure of resilience-aware MLOps for medical diagnostic systems has been proposed. Experimentally confirmed increase of robustness and speed of adaptation for medical image recognition system during several intervals of the system's life cycle due to the use of resilience optimization and uncertainty calibration stages. The experiments were performed on the DermaMNIST dataset, BloodMNIST and PathMNIST. ResNet-18 as a representative of convolutional networks and MedViT-T as a representative of visual transformers are considered. It is worth noting that transformers exhibited lower resilience than convolutional networks, although this observation may be attributed to potential imperfections in the architecture of adapters and meta-adapters.

Conclusion: The main novelty of the suggested resilience-aware MLOps methodology and structure lie in the separating possibilities and activities on creating a basic model for normal operating conditions and ensuring its resilience and trustworthiness. This is significant for the medical applications as the developer of the basic model should devote more time to comprehending medical field and the diagnostic task at hand, rather than specializing in system resilience. Resilience optimization increases robustness to disturbances and speed of adaptation. Calibrated confidences ensure the recognition of a portion of unabsorbed disturbances to mitigate their impact, thereby enhancing trustworthiness.

KEYWORDS

MLOps, medical diagnosis, image recognition, robustness, resilience

1 Introduction

1.1 Motivation

The advent of Artificial Intelligence (AI) in healthcare has opened new horizons in medical diagnostics, offering more precise, efficient, and rapid techniques for detecting a wide range of diseases. However, the critical nature of healthcare imposes strict requirements on AI-based diagnostic systems, necessitating robust performance, high reliability, and stringent data security measures. Despite the attention to quality, security, and performance in traditional Machine Learning Operations (MLOps), an overlooked aspect remains—resilience to disturbances.

In healthcare applications, AI-based systems are exposed to numerous disturbances that can significantly impact their effectiveness. These disturbances may range from adversarial attacks designed to manipulate model outputs, to fault injections that can undermine system integrity, and to drift phenomena where the model's performance degrades due to changing patterns in data distribution. Conventional MLOps methodologies focus extensively on data quality, model performance, and security but do not adequately address these resilience challenges. Given the potentially life-altering decisions that AI-based medical diagnostic systems are entrusted with, a lack of resilience can have severe consequences, including inaccurate diagnoses and, consequently, improper treatment plans.

Moreover, the unique characteristics of the healthcare domain such as patient-specific variabilities, heterogeneous data sources, and strict regulatory constraints introduce distinctive kinds of disturbances that are not commonly observed in other sectors. Therefore, a "one-size-fits-all" approach from other domains cannot be directly applied here.

The ability of a system to be resilient—to absorb, detect, and adapt to disturbances—is particularly crucial in high-stakes environments like healthcare. A resilient-aware MLOps framework for AI-based medical diagnostic systems would not only improve their robustness but would also enhance trust among clinicians, healthcare providers, and patients, thus accelerating the adoption rate of AI in healthcare.

In light of these challenges and opportunities, this study aims to enrich MLOps methodology by incorporating resilience as a fundamental component. By identifying characteristic disturbances in healthcare and developing methods to ensure resilience, this study endeavors to elevate the reliability and trustworthiness of AI-based medical diagnostic systems, making them better equipped to provide quality healthcare solutions in dynamic and unpredictable environments.

1.2 State-of-the-art

The evolution of AI in healthcare has led to various significant advancements, many of which are integrated into existing MLOps frameworks (1). A plethora of research exists, focusing on improving data quality, model training, evaluation, and deployment in the healthcare domain (2, 3).

Machine learning operations have gained momentum in healthcare due to their potential to streamline the development, deployment, and maintenance of machine learning models (4). Several studies have delved into the unique requirements and challenges that healthcare poses to the MLOps methodology, such as patient data confidentiality (5), bias reduction (6), and compliance with health regulations like HIPAA (7). However, most existing MLOps frameworks are designed to ensure efficient operation rather than resilience to the various disturbances that healthcare environments may present.

The concept of resilience in AI systems is not new and has been examined across various fields, including cybersecurity, manufacturing, and even autonomous vehicles. Techniques like adversarial training (8, 9), robust optimization (10), and uncertainty quantification have been employed to improve resilience (11). The paper (12) proposes the concept of Secure Machine Learning Operations paradigm, but without proposals for combining different methods and aspects of protecting the same AI system from different threats. The issue of resolving the incompatibility of the selected approaches in the tasks of ensuring the resilience and efficiency of the AI system is not considered. The vast majority of researchers consider each type of disturbance for AI systems separately, and the question of compatibility of methods for ensuring resilience to each of these disturbances remains under-researched (13–15). In addition, although these methods provide a certain degree of perturbation absorption, they often do not ensure rapid adaptation and evolution in response to changing conditions.

Despite the abundance of work in MLOps, resilience, and AI-based medical diagnostics separately, there is a conspicuous absence of research focusing on integrating resilience into MLOps frameworks specifically designed for medical diagnostic systems. This gap highlights the need for a holistic approach that combines these elements to ensure not just efficiency and reliability, but also resilience against the myriad disturbances that these systems may encounter.

1.3 Objectives and contributions

The aim of this study is to develop a new MLOps methodology that ensures the resilience of a medical diagnostic system to such negative factors as adversarial attacks, failure injection, drift, and out-of-distribution of data. The key objectives are as follows:

- analysis of resilience issue of MLOps for healthcare;
- analysis of methods of ensuring the resilience of AI-systems;
- develop resilience-aware MLOps architecture for Medical Diagnostic Systems; and
- experimentally confirm the advantages of resilience-aware MLOps compared to the conventional approach.

Structurally, the work consists of the following sections. The analysis of methods of ensuring the resilience of MLOps and resilience-aware MLOps architecture for Medical Diagnostic Systems are presented in Section 2. The Section 3 describes the experimental results of testing and comparition of the proposed resilience-aware MLOps and Conventional MLOps. The research results are discussed in the Section 4. The Section 5 concludes the paper and describes the directions of future research.

The main contribution of the research includes architectures of resilience-aware MLOps for Medical Diagnostic Systems. In addition, the results of the comparison between the traditional and the proposed MLOps on the MedMNIST datasets are analyzed. It has been experimentally proven that the addition of resilience optimization, predictive uncertainty calibration, uncertainty monitoring, and graceful degradation makes a positive contribution to the robustness and performance recovery of a medical diagnostic system.

2 Architecting resilient MLOps-based medical diagnostic system

2.1 Resilience issue of MLOps for healthcare

Machine learning operations is the process of automating the lifecycle of machine learning models. It involves four main stages (Figure 1) (1, 2):

1. Data Preparation—gathering, cleaning, and transforming data for further model training.
2. Model Development, Training and Evaluation—building the architecture, training, and testing the model on prepared data.
3. Model Deployment—integrating the trained model into a production environment.
4. Performance Monitoring—tracking the model's metrics in operation and providing feedback to the data preparation stage.

Conventional approaches to MLOps often do not pay enough attention to the resilience of machine learning systems to the perturbations inherent in the medical domain. They do not focus on absorbing and detecting disturbances and adapting to them quickly. However, for medical applications, these aspects are extremely important, as human lives depend on the recommendations of ML systems. Disturbances in an intelligent system can be caused by both intentional attacks and natural causes. Examples of natural disturbances include noise in the data, sudden hardware faults (memory faults), and data drift over time (16). Intentional attacks can also include fault injections and data manipulation in the form of so-called adversarial attacks.

Drift is particularly relevant to the healthcare industry, as disease patterns can change due to new strains of viruses and bacteria and disease patterns can change due to changes in treatment protocols (17). Data characteristics may also change due to improvements in medical equipment, changes in data collection methods, and changes in demographics. In addition, the emergence of new data, the identification of previously unknown relationships and factors, and the refinement of disease taxonomies are additional sources of concept drift. The main problem with drift adaptation is the delay in the arrival of labeled data after drift occurs, so the ability to quickly adapt to a small amount of labeled data is a very relevant property.

Adversarial attacks in AI-based medical diagnostic systems refer to deliberate manipulations of input data (such as medical images) designed to deceive AI algorithms (18). These attacks exploit vulnerabilities in the AI's learning process, where slightly perturbed images, indistinguishable to the human eye, can lead to incorrect diagnoses or assessments. The source of these attacks can vary, ranging from external threats aiming to undermine the system's reliability to internal errors in training data or algorithm design. However, to protect against adversarial attacks, the initial development of models and methods for training intelligent diagnostic systems may be complicated by the need to investigate the compatibility of various methods for enhancing the robustness and resilience of AI systems (19).

In the computational environment of an intelligent medical diagnostic system, malicious faults, commonly known as fault injections, can pose significant threats. These deliberate disruptions can be executed in various forms, targeting different components of the system. For instance, one notable type of attack is the "row hammer" attack on memory (20). This involves repeatedly accessing a row of memory cells to induce bit flips in adjacent rows, potentially corrupting data or causing system crashes. The existing efforts to increase fault tolerance and the adaptation rate to a certain amount of faults may not be compatible with protection against other types of disturbances.

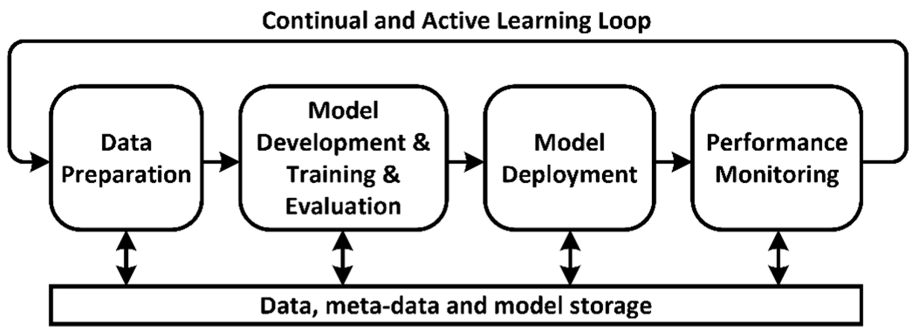


FIGURE 1
Basic stages of conventional MLOps.

Thus, AI algorithms used in the healthcare industry have numerous vulnerabilities that traditional MLOps approaches do not focus on. Therefore, a promising direction for the development of AI-based medical diagnostic systems is the use of MLOps with elements to ensure resilience to disturbances.

2.2 Methods of ensuring the resilience of AI systems

Table 1 shows the approaches to ensure the resilience to adversarial attacks, fault injection, and concept drifts for AI system. The ability to absorb disturbances (robustness), graceful degradation due to the impact of disturbances that could not be absorbed, and rapid adaptation to new disturbances are considered to be the key features of resilient system. “Graceful degradation” refers to a system being pre-configured with an organized set of less functional states. These states represent acceptable compromises between functionality, performance, and cost-effectiveness.

Implementation of out-of-domain generalization through domain randomization or domain augmentation increases model robustness to limited shifts in data distribution (21). Dynamic adjustment of model weights in an ensemble can mitigate a certain level of concept drift (32). In addition, the ability to detect drift or out-of-distribution data can provide graceful degradation by delegating control to a human or to an AI model that is more resistant to such a disturbance.

Robustness to adversarial attacks can be enhanced by protecting training data or restricting access to knowledge about the AI model. The protection of training data is usually achieved through data encryption and sanitization (36). The incorporation of randomization or non-differentiable input transformations into the model provides gradient masking to counteract adversarial evasion attacks (33). Moreover, adding different regularization and training on adversarial samples provides robust optimization (10). However, detecting adversarial examples can be an effective mechanism for graceful degradation by delegating control to humans or automatically switching to models that are more resilient to such a perturbation (22).

Redundancy in the form of N-versioning of AI model or duplication of critical neurons is the most common way to ensure robustness to faults (fault tolerance) (35). However, training under noise added to the gradient or neurons weights can also help to reduce the importance of individual neurons by providing fault masking, i.e., eliminating their influence (34). However, the use of error detection mechanisms for model weights can be combined with error correction mechanism or with downloading an uncorrupted version of the weights (24). If an error in the weights causes abnormal distortions in the feature space, it can increase predictive uncertainty and require a delegation of control to a human or switching to another model or model branch (27, 28).

Estimating model uncertainty is a useful function to identify the negative impact of any destructive influences on the AI system. However, by default, conventional AI model do not provide an efficient predictive uncertainty estimation and it requires calibration. In addition, detection of destructive disturbances affecting the AI model can be achieved through the mechanisms of AI explainability.

Graceful degradation can occur by increasing the resource consumption for disturbance processing, for example, by delegating control to an expert or large model. An expert can be reinforced by an AI explanation algorithm, while a large model is used with auxiliary semantic information (i.e., Zero-shot learning) (26). Switching to a simpler model can also be viewed as a graceful degradation, as a simple model is generally less sensitive to disturbances in the data, but produces more coarse or abstract predictions (27, 28).

Adaptation to disturbance typically occurs through retraining in the face of disturbance using Active Learning, Continuous Learning, and Few-Shot Learning methods (29–31). However, to increase the speed of adaptation, meta-learning and Parameter-Efficient Fine Tuning methods are widely used (37). In addition, meta-learning is also effective in optimizing robustness (38).

Annotation of training data for medical diagnostic system requires deep medical knowledge, while the knowledge is constantly expanding and updating. There is a need to integrate Active Learning in MLOps feedback to effectively use data and time of highly qualified experts (29). At the stage of training data preparation, an expert could annotate the most complex cases. Complicated cases where the AI system has the greatest uncertainty can be identified by a specified uncertainty indicator. The simplest and most logical way to measure uncertainty is to calculate Shannon’s Entropy Measure for classification model or using quantile regression for regression model.

Detection of out-of-distribution, concept drift, a certain fraction of adversarial attack and fault injection effects can be implemented by testing for exceeding the threshold value of the model uncertainty indicator. In this case, the threshold value of the uncertainty indicator can be defined as a 95% percentile on an augmented test or training dataset.

Thus, there are methods for implementing different resilience features for different disturbances. However, the vast majority of them require changing the model or learning algorithm, which complicates the responsibility separation in MLOps. In this case, making the AI system resilient will require additional research into the compatibility of different mechanisms for ensuring resilience to various disturbances and their mutual influence.

TABLE 1 Approaches to ensure the resilience of AI-systems.

Disturbance source	Resilience capabilities		
	Robustness	Graceful degradation	Adaptation
Drift, Out-of-distribution data	Out-of-domain generalization (21); Ensemble selection (21)	Detecting adversarial examples (22), drift, out-of-distribution data (23), and faults (24); AI Explanation mechanism (25); Zero-shot learning (26); Switch to simpler model (27) or Reduce prediction granularity (28)	Active Learning (29); Continual Learning (30); Few-Shot Learning (31).
Adversarial attack	Data encryption and sanitization (32); Gradient Masking (33); and Robustness Optimization (10)		
Fault injection	Fault masking (34); Explicit redundancy (35); and Error detection and correction (24)		

2.3 Resilience-aware MLOps architecture for medical diagnostic systems

Important principles in MLOps are the separation of responsibilities and collaboration between teams. Platform-level specialized solutions to ensure the resilience of any AI model delegates the updating and maintenance of this mechanism to a separate team of AI resilience experts. New MLOps stages for ensuring resilience aspects should be implemented as *post-hoc* procedures to maximize the separation of responsibilities.

Figure 2 shows a diagram of the proposed resilience-aware MLOps, which additionally includes the stages of *Post-hoc* Resilience Optimization, *Post-hoc* Uncertainty Calibration, and Uncertainty Monitoring and Graceful Degradation. In addition to Uncertainty Monitoring, the Explainable AI mechanism can be used to assist decision-making by the human to whom control is delegated in case of uncertainty. The article (39) questions the necessity and adequacy of existing methods of explaining decisions, so the explanation mechanism will be excluded from further consideration, but for generality, the diagram shows this MLOps stage.

At the stage of resilience optimization, it is proposed to attach computationally efficient (meta-) adapters to the original model in order to increase robustness and speed up the fine-tuning process (40). In this case, the weights of the original model remain frozen. The original model usually consists of certain blocks or modules, for example Convolutional Residual Block. To generalize, we will refer to these blocks as frozen operations and denote them as $OP(x)$. The parallel method of connecting an adapter to the frozen blocks of the model is the most convenient and versatile approach (Figure 3A). In this case, to ensure the properties of resilience, it is proposed to use three consecutive blocks of adapters at once, two of which are tuned during meta-training (40). To balance between different modules, we introduce a channel-wise scaling factor.

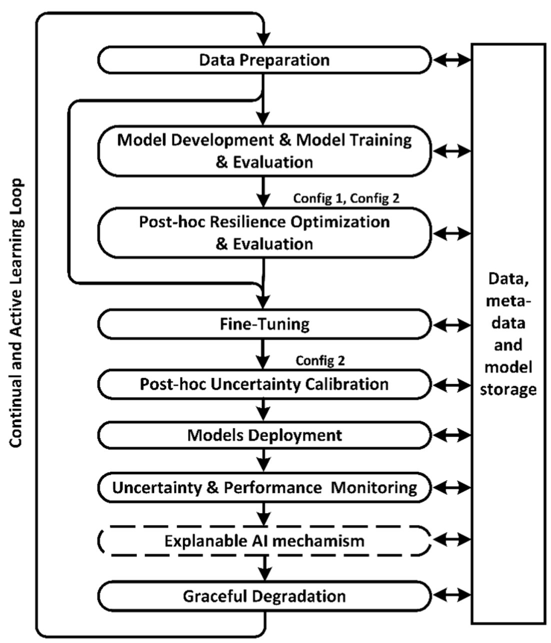


FIGURE 2
Basic stages of resilience-aware MLOps.

The adapter architectures depicted in Figure 3B are based on convolutional layers. The convolutional adapter shown in Figure 3B has a hyperparameter γ, which regulates channel compression by 1, 2, 4, or 8 times. However, adapters can also be implemented as a two-layer feed-forward network with a downward projection bottleneck or ResNet-like conversion.

The original model is trained on a dataset $D_{base} = \left\{D_{base}^{tr}; D_{base}^{val}\right\}$ to perform the main task under known conditions. Resilience optimization involves generating a set of synthetic disturbance implementations $\left\{\tau_i | i = \overline{1,N}\right\}$. As disturbances τ_i can be considered adversarial attacks, fault injection, or switching to a new task. In addition, it is necessary to provide datasets $D = \left\{D_k^{tr}; |D_k^{val}; |k = \overline{1,K}\right\}$, that solve other problems for K few-shot learning tasks, where fine-tuning data D_k^{tr} is used in the fine-tuning stage and validation set D_k^{val} is used in the meta-update stage. There is also a given set of parameters θ, ϕ, ω, and W, where θ are parameters of a pretrained and frozen base AI model, ϕ and ω are adaptation parameters of AI model backbone, and W are task specified parameters (model head parameters). Head weights W_{base} for the main task are pre-trained on the data D_{base}. If we reject the specialization of different parameters of the AI model and denote the set of all parameters as $\Xi = \langle \theta, \phi, \omega, W \rangle$, then the process of meta-learning for direct maximization of the expected resilience criterion can be described by the formula:

$$\Xi^* = \underset{\Xi}{argmax} \underset{\tau_i \sim p(\tau)}{E}\left[R_{\tau_i}\left(U\left(\Xi, D\right)\right)\right] = \underset{\Xi}{argmax}\, F\left(\Xi\right) \quad (1)$$

where U is an operator that combines disturbance generation and adaptation in T steps, which maps the current state of ϕ to new state of ϕ;

R_{τ_i} is a function that calculates the value of the integral resilience indicator for τ_i disturbance implementation over model parameters ω during its adaptation and the test sample $D_{\tau_i}^{val}$.

$$R_{\tau_i} = \frac{1}{P_0 T}\sum_{t=1}^{T} P_{\tau_i}\left(\theta, \omega, \phi_t, W_t, D_{\tau_i}^{val}\right) \quad (2)$$

where P_{τ_i} is a performance metric for current state of model parameters and evaluation data.

If we use the SGD stochastic gradient descent algorithm with T steps in the U operator and use gradient meta-update in the outer loop, we will get the algorithm shown in Figure 3. The meta-gradient estimation can be performed over the Gaussian-smoothed version of the outer loop objective, which is calculated by the formula (38).

$$\nabla \underset{g \sim N(0,I)}{E}\left[F\left(\Xi + \sigma g\right)\right] = \frac{1}{2\sigma} E\left[R\left(\Xi + \sigma g\right) - R\left(\Xi - \sigma g\right)\right] \quad (3)$$

A perturbation vector g is formed for the meta-optimized parameters at the beginning of each meta-optimization iteration; the resulting algorithm will be as shown in Figure 4.

The analysis of Figure 4 shows that the type of disruptive influence does not change within a single meta-adaptation step. However, each meta-adaptation step begins with the selection of a disruptive influence type, followed by the generation of n implementations of the disruptive influence with a subsequent nested adaptation loop for each of them. Simultaneously combining disturbances may be ineffective.

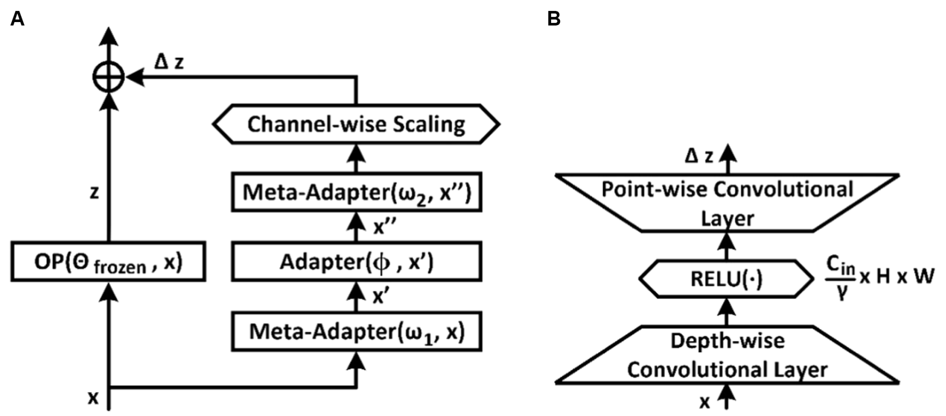


FIGURE 3
Parallel tuning scheme and tuner architectures. **(A)** Parallel tuning scheme for the frozen block; **(B)** Adapter or meta-adapter based on two-layer convolutional network with channel dimension down-sampling bottleneck.

Require: Distribution over disturbances $p(\tau)$; Step size hyperparameters α, β; Precision parameter σ; Number of adaptation steps T.

```
Pretrain ϕ, ω on original data D_base
While not done do:
    Select type of disturbance from set { fault injection, evasion adversarial attack, task change}
    Sample disturbance implementations τ_i ~ p(τ), i = 1,n
    Sample perturbation vectors g_{ϕ,τ_i} ~ N(0, I), i = 1,n; g_{ω,τ_i} ~ N(0, I), i = 1,n
    For i=1,2,...,n do:
        Clone the current parameters: θ́_τi, ώ_τi, ϕ́_τi, Ẃ_τi ← copy(θ, ω, ϕ, W_base)
        ϕ́_τi+ ← ϕ́_τi + σ g_{ϕ,τi};  ϕ́_τi− ← ϕ́_τi − σ g_{ϕ,τi}
        ώ_τi+ ← ώ_τi + σ g_{ω,τi} ; ώ_τi− ← ώ_τi − σ g_{ω,τi}
        If disturbance type is a task change:
            Sample the training and validation data D^tr_τi, D^val_τi from new task
        else:
            Sample the training and validation data D^tr_τi, D^val_τi from D_base
        If disturbance type is a fault injection:
            θ́_τi, ώ_τi+, ώ_τi−, ϕ́_τi+, ϕ́_τi− ← Fault_injection(θ́_τi, ώ_τi+, ώ_τi−, ϕ́_τi+, ϕ́_τi−)
        If disturbance type is an evasion adversarial attack:
            D^tr_τi, D^val_τi ← Adversarial_perturbation(D^tr_τi, D^val_τi)
        {ϕ́_τi+,t | t = 1,T} ← SGD_{ϕ,W}(L_τi(θ́_τi, ώ_τi+, ϕ́_τi+, Ẃ_τi, D^tr_τi), T, α)
        {ϕ́_τi−,t | t = 1,T} ← SGD_{ϕ,W}(L_τi(θ́_τi, ώ_τi−, ϕ́_τi−, Ẃ_τi, D^tr_τi), T, α)
        R_{+,τi} ← R({P_t(θ́_τi, ώ_τi+, ϕ́_τi+,t, D^val_τi)})
        R_{−,τi} ← R({P_t(θ́_τi, ώ_τi−, ϕ́_τi−,t, D^val_τi)})
        R_τi ← 1/2 (R_{+,τi} − R_{−,τi})
    ω ← ω + β (1/(σn)) Σ_{i=1}^{n} R_τi g_{ω,τi}
    ϕ ← ϕ + β (1/(σn)) Σ_{i=1}^{n} R_τi g_{ϕ,τi}
```

FIGURE 4
Pseudocode of model-agnostic meta-learning with evolution strategies for AI-system resilience optimization.

For example, after adding fault injection to the weights, we will have an outdated model, and applying adversarial attacks to it may be irrelevant.

The formation of adversarial samples is based on the *Adv_perturbation*( )Adv_perturbation( ) function. For differentiable models, FGSM attacks or PGD attacks can be used (36, 41). It is proposed to use adversarial attacks based on the search algorithm of the covariance matrix adaptation evolution strategy for non-differentiable models (39). The level of perturbation is limited by the $L_{\circ}$ -norm or L_0L$_0$-norm. In this case, if the image is normalized by dividing pixel brightness by 255, then the specified disturbance level is also divided by 255.

The formation of fault injections is performed by the $Fault_inj(\;)$Fault_inj( ) (42). It is suggested to choose the most difficult fault type to absorb, which involves generating an inversion of a randomly selected bit (bit-flip injection) in the model weight. For differentiable models, it is suggested to pass the test dataset through the network and calculate the gradients, which can then be sorted by their absolute values. In the top-*k* weights with the highest gradient, one bit is inverted in a random position. The proportion of weights for which one random bit is inverted can be denoted as the fault rate.

Task change is needed to simulate concept drift and out-of-distribution. Forming a sample of other tasks can be done by randomizing the domain of the same task or by selecting tasks from relevant domains but sampling truly different tasks. These two approaches can also be combined.

Augmented versions of training samples can be used for improved calibration on in-distribution data. Out-of-distribution data can be generated from other datasets that do not share similar labels. Out-of-distribution data can be generated from other datasets that do not have semantically similar labels. One of the effective methods of generating Out-of-distribution is the use of Soft or Hard Brownian Offset Sampling with Autoencoders (43). Software libraries and examples of application to various data modalities are available for the Soft Brownian Offset algorithm.

The *post-hoc* calibration algorithm requires adding certain add-ons to the frozen model that are adjusted on the calibration data to reduce the discrepancy between the prediction confidence and the actual probability. Calibration Add-Ons for classification model based on Temperature Scaling, Platt Scaling, Isotonic Regression, Histogram Binning, Bayesian neural networks, ensembles, etc.

Active Learning is a widespread practice in the medical industry and in our MLOps diagram, it is part of the feedback loop. In the case of conventional MLOps, the base model is tuned on the new data, while in the case of resilience-aware MLOps, adapters are tuned. Re-training of the base model and resilience optimization can be performed after a certain predefined amount of new data has been accumulated since the last resilience optimization.

3 Experiments and results

3.1 Experimental setup

MedMNIST datasets contain annotated medical data samples for testing machine learning and artificial intelligence techniques in healthcare (44). Experimental research is proposed to be performed on these datasets. DermaMNIST dataset, which contains seven classes, will be considered as the main task dataset. The BloodMNIST and PathMNIST datasets will be used as data sources for few-shot learning tasks in the resilience optimization (meta-learning) process. In this case, a set of five classes will be used for few-shot learning tasks, which are randomly selected from the set of available classes (n_{way} = 5). It is proposed to use 16 images per class (k_shot = 16), which are provided in mini-batches by four images (mini_batch_size = 4) during adaptation. Thus, the number of adaptation steps is T = (k_shot*n_{way})/mini_batch_size = 20 iterations. The learning rate of the inner and outer loop of meta-learning are $\alpha = 0.001$ and $\beta = 0.0001$, respectively. The maximum number of meta-iterations is 300. However, the Early Stopping algorithm is used to stop meta-learning, which terminates the execution if the criterion does not change for more than 10 consecutive iterations by more than 0.001. In this case, the convolutional network ResNet-18 and the visual transformer MedViT-T will be used as representatives of two main approaches to building a neural network architecture in the field of computer vision (45). In the case of ResNet-18, adapters and meta-adapters are connected in parallel to each ResBlock. In the case of MedViT-T, adapters and meta-adapters are connected in parallel to each Local Feed-Forward Network and Multi-Head Self-Attention Module.

Several configurations will be considered to illustrate the impact of additional stages of resilience-aware MLOps on the accuracy and speed of its recovery:

- Config 0—conventional MLOps with Fine-Tuning stage and Active Learning Feedback Loop;
- Config 1—upgraded Config 0 with Resilience Optimization stage; and
- Config 2—upgraded Config 1 with Predictive Uncertainty Calibration stage.

MedMNIST datasets contain training, validation, and test subsamples. To simplify the experiment and analyze the results, we will combine the validation and test samples into one test set and divide it into four parts. Each part of the test data is needed to simulate a part of the AI model's life cycle. Let us consider four consecutive parts of the life cycle:

- Test 0—training (parameter optimization) of the AI model on the training set and testing the model on the first part of the test set, and selecting 10% of the test data points with the highest predictive uncertainty;
- Test 1—fine-tuning of the AI model on the selected data points from the previous test and testing the model on the second part of the test dataset under the disturbance, and selecting 10% of the test data points with the highest predictive uncertainty;
- Test 2—fine-tuning on the selected data points from the previous test and testing the model on the third part of the test dataset under the disturbance, and selecting 10% of the test data points with the highest predictive uncertainty; and
- Test 3—fine-tuning on selected test data points from the previous test and testing the model on the fourth part of the test data set under conditions of increased disturbance intensity.

In order to keep things simple, we assume that the graceful degradation is implemented as a decision rejection in case of uncertainty detection, i.e., if the entropy exceeds a threshold. We assume that control is transferred to a human, a more efficient model, or a preconfigured default procedure. Therefore, we will consider model accuracies calculated in two ways:

- ACC1 is the accuracy which counts rejected examples as false decisions; and

- ACC2 is the accuracy that does not take into account rejected examples.

Conventional MLOps reject decisions based on predictive confidence, while resource-aware MLOps reject decisions based on uncertainty.

For training adapters with meta-adapters, fault injection is carried out by selecting weights with the largest absolute gradient values. The proportion of modified weights is fault_rate = 0.1. For testing the resulting model, fault injection will be performed by random bit-flips in randomly selected weights, the proportion of which (fault_rate) are equals to 0.1 or 0.15.

The training of the tuners and meta-tuners involves generating adversarial samples using the FGSM algorithm with perturbation_level according to $L\infty$ up to 3. However, to test the resulting model against adversarial attacks, the adversarial samples are generated using the CMA-ES algorithm with perturbation_level according to $L\infty$-norm are equals to 3 or 5. The number of solution generation in the CMA-ES algorithm is set to 10 to reduce the computational cost of conducting experiments.

Instead of directly modeling different types of concept drift or novelty in the data, it is proposed to model the ability to quickly adapt to task changes, as this can be interpreted as the most difficult case of real concept drift. The preparation for the experiment involved adding adapters and meta-adapters to the network, which had been trained on the DermaMNIST dataset. During meta-training, these adapters performed adaptations to either attacks or a five-class classification task, which was randomly generated from a selection of the nine-class PathMNIST set, the eight-class BloodMNIST set, or the seven-class DermaMNIST set. Subsequently, to verify the capability for rapid adaptation to a new task change, the new task was considered either as a classification task with the full set of PathMNIST classes or as a classification task with the full set of BloodMNIST classes. The resilience curve is constructed over 20 mini-batch fine-tunings, from which the resilience criterion (2) is calculated.

Taking into account the elements of randomization, it is proposed to use their average values when assessing the accuracy of the model. To this end, 100 implementations of a certain type of disturbance are generated and applied to the same model or data.

Uncertainty calibration will be performed on a dataset containing augmented test samples and out-of-distribution samples generated by Soft Brownian Offset Sampling. 300 images per class are generated for in-distribution test set to calibrate the uncertainty. The total number of out-of-distribution images is the same as the in-distribution calibration set. In this case, the Soft Brownian Offset Sampling algorithm is used with the following parameter values: minimum distance to in-distribution data is equal to 25; offset distance is equal to 20; and softness is equal to 0. Bayesian Binning into Quantiles with 10 bins was chosen as the calibration algorithm.

3.2 Results

Table 2 illustrates the change in accuracy with (ACC1) and without (ACC2) accounting for rejected decisions at different successive parts of the ResNet-18 model life cycle with resilience-aware add-ons under fault injections, depending on the selected MLOps configuration. Test 0, Test 1, Test 2, and Test 3 were repeated 100 times each, and Table 2 shows the average accuracy to account for the randomization effect.

Table 2 shows that after the first test with fault injection (Test1) and the last test with increasing fault injection intensity (Test3), config 1 (with resilience optimization) and config 2 (with uncertainty calibration) provide an increase in fault tolerance, which is fully consistent with the goals of resilience-aware MLOps. In addition, the dynamics of accuracy growth during adaptation (Tes 1-Test 2) is higher in the latter two configurations. Even with an increase in the fraction of damaged weight tensors, the accuracy of the system almost does not drop, unlike the configuration of conventional MLOps. Also, comparing ACC2 with ACC1, we can conclude that ACC2 is always larger than ACC1, but this difference is greater in the case of resilience optimization, and especially in the case of uncertainty calibration. Note that the averaged values of ACC1 and ACC2 for the ResNet-18-based model on Test 0-Test 3 test data with the corresponding fault injection rate without fine-tuning on 10% of human-labeled examples are 0.627 and 0.638, respectively. It proves the importance of using an active feedback loop for adaptation. For the average accuracy values in Table 2, the margin of error does not exceed 1% at a 95% confidence level.

Table 3 illustrates the change in the average accuracy with and without rejected samples, i.e., ACC1 and ACC2, at different successive parts of the lifecycle of the ResNet-18 model with resilience-aware add-ons under adversarial evasion attacks, depending on the selected MLOps configuration. Test 0, Test 1, Test 2, and Test 3 are repeated 100 times each, and Table 3 shows the average accuracy to account for the randomization effect.

The results of Test 1 and Test 3 in Table 3 show that config 1 (with resilience optimization) and config 2 (with uncertainty calibration) provide an increase in robustness. In addition, the dynamics of accuracy growth during adaptation (Tes 1, Test 2) is higher in the latter two configurations. However, the traditional MLOps (config 0) also showed the ability to adapt quickly during fine-tuning (comparing the results of Test 1 and Test 2), although it was not successful in restoring performance. Config 1 and Config 2 show a noticeable recovery in accuracy, and config 2 on Test 2 even showed an improvement in accuracy compared to its pre-disturbance value. Increasing the magnitude of the perturbation (Test 3) leads to a decrease in accuracy in all cases, but config 1 and config 2 show greater resilience compared to config 0. Also, across all experiments, ACC2 is larger than ACC1, which indicates the ability to recognize disturbances that cannot be absorbed. Note that the averaged values of ACC1 and ACC2 for the ResNet-18-based model on perturbed test data from Test 0-Test 3 stages without fine-tuning on 10% of human-labeled examples are 0.671 and 0.682, respectively. It also proves the importance of using an active feedback loop for adaptation. For the average accuracy values in Table 3, the margin of error does not exceed 1.1% at a 95% confidence level.

Tables 4, 5 illustrate the changes in the accuracy values of the MedViT-T-based model under the influence of fault-injection and adversarial attack during the life cycle, depending on the MLOps configuration. In the MedViT-T experiments, Test 0, Test 1, Test 2, and Test 3 are repeated 100 times each, and the average accuracy is reported in Tables 4, 5.

An analysis of Tables 4, 5 shows that MedViT-T exhibits similar behavior to ResNet-18 on the same tests and MLOps configurations.

TABLE 2 Accuracy of the ResNet-18-based model under the influence of fault injection during the life cycle depending on the MLOps configuration.

MLOps configuration	Test 0		Test 1		Test 2		Test 3	
	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2
Config 0	0.751	0.781	0.652	0.679	0.659	0.681	0.623	0.634
Config 1	0.750	0.801	0.722	0.765	0.737	0.773	0.739	0.774
Config 2	0.768	0.822	0.734	0.810	0.749	0.822	0.747	0.798

TABLE 3 Accuracy values of the ResNet-18-based model under adversarial attack during the life cycle depending on the MLOps configuration.

MLOps configuration	Test 0		Test 1		Test 2		Test 3	
	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2
Config 0	0.751	0.781	0.683	0.689	0.708	0.721	0.663	0.674
Config 1	0.750	0.801	0.735	0.785	0.749	0.797	0.742	0.753
Config 2	0.768	0.822	0.753	0.810	0.770	0.847	0.768	0.802

TABLE 4 Accuracy values of the MedViT-T-based model under the influence of fault injection during the life cycle depending on the MLOps configuration.

MLOps configuration	Test 0		Test 1		Test 2		Test 3	
	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2
Config 0	0.769	0.791	0.672	0.698	0.671	0.751	0.633	0.694
Config 1	0.772	0.835	0.742	0.775	0.750	0.781	0.731	0.784
Config 2	0.777	0.902	0.752	0.820	0.759	0.842	0.750	0.808

TABLE 5 Accuracy values of the MedViT-T-based model under the influence of adversarial attack during the life cycle depending on the MLOps configuration.

MLOps configuration	Test 0		Test 1		Test 2		Test 3	
	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2	ACC1	ACC2
Config 0	0.769	0.791	0.705	0.748	0.710	0.751	0.697	0.750
Config 1	0.772	0.835	0.742	0.788	0.750	0.781	0.748	0.780
Config 2	0.777	0.902	0.760	0.833	0.767	0.842	0.759	0.800

However, MedViT-T is characterized by a lower adaptation rate (comparison of Test1, Test2 results) compared to ResNet-18. Note that the averaged values of ACC1 and ACC2 for the MedViT-T-based model on Test 0-Test 3 test data with the corresponding fault injection rate without fine-tuning on 10% of human-labeled examples are 0.653 and 0.694, respectively. It proves the importance of using an active feedback loop for adaptation. Averaged values of ACC1 and ACC2 for the MedViT-T-based model on perturbed test data from Test 0 to Test 3 stages without fine-tuning on 10% of human-labeled examples are 0.687 and 0.695, respectively. It also proves the importance of using an active feedback loop for adaptation. For the average accuracy values presented in Table 4, the margin of error does not exceed 0.9% at a 95% confidence level. Similarly, for the average accuracy values in Table 5, the margin of error does not exceed 1% at the same confidence level.

To evaluate the robustness and speed of adaptation of a pre-configured medical AI system to concept drift, it is proposed to calculate the integral resilience criterion (2) in fine-tuning mode (few-shot learning) on BloodMNIST dataset or PathMNIST dataset (Table 6). In this case, the AI system was pre-trained and optimized on the DermaMNIST set. It is proposed to use 16 images per class (k_shot = 16), which are provided in mini-batches of four images (mini_batch_size = 4) during adaptation.

Analysis of Table 6 shows that adding a resilience optimization stage to MLOps increases resilience to concept drift, i.e., robustness and speed of adaptation. However, the architecture of visual transformers in our experiments shows itself to be less resilient. For the average accuracy values in Table 6, the margin of error does not exceed 1% at a 95% confidence level.

4 Discussion

The proposed structure of resilience-aware MLOps makes it possible to implement various specific solutions for the implementation of its separate stages and mechanisms. The main idea is to divide the labor of developers of the basic AI model that functions efficiently under normal conditions and specialists in ensuring the

TABLE 6 The value of the integral resilience criterion (2) to the change of the medical image analysis task depending on the MLOps configuration.

MLOps configuration	Fine-tuning of ResNet-18-based AI model		Fine-tuning of MedViT-T-based AI model	
	On 20 mini-batches of BloodMNIST (complete set of classes)	On 20 mini-batches of PathMNIST (complete set of classes)	On 20 mini-batches of BloodMNIST (complete set of classes)	On 20 mini-batches of PathMNIST (complete set of classes)
Config 0	0.68	0.74	0.66	0.71
Config 1	0.80	0.82	0.71	0.74

resilience of the intelligent system to disturbances and changes. The healthcare industry is extremely complex and requires deep knowledge in various fields. Developers of the basic AI model are usually overloaded with taking into account the specifics of data, methods of data collection and the application itself to solve the applied data analysis task. Solving problems related to ensuring AI resilience, i.e., security, trustworthiness, robustness and rapid adaptation to changes, relies on specific expertise not related to a particular application area (13). The main difficulty in separating the tasks is the lack of universality of methods for ensuring resilience and the lack of a complete understanding of the compatibility of methods that provide different aspects of resilience and protection against different types of disturbances (16). Determining the compatibility of methods and combining them can increase flexibility and resilience depending on requirements and constraints.

The proposed implementation of *Post-hoc* Resilience Optimization is just one of the possible solutions that shows the fundamental possibility of separating the stage of developing a basic model for normal conditions and add-ons to ensure resilience to disturbances and changes. Moreover, the importance of using the proposed *Post-hoc* Uncertainty Calibration stage was experimentally confirmed. This stage allows, firstly, to detect disturbances and, secondly, to correctly assess uncertainty and tolerate it, i.e., to ensure trustworthiness.

Unexpectedly, the worse resilience indicators for the visual transformer compared to the convolutional network were found. This indicates that there is a need to improve architectures and connection methods for adapters and meta-adapters. Similarly, uncertainty calibration methods did not ensure 100% accuracy of the models. This is partly due to the fact that there are two types of uncertainty—aleatory and epistemic uncertainty. In this case, the aleatory uncertainty cannot be eliminated, because it is an inherent part of the observed process or object (46). On the other hand, there are many calibration algorithms, each of which depends on a large number of hyperparameters. Understanding their impact on epistemic uncertainty in the context of Resilient-aware MLOps is an important research area. Progress in AI architectures and their hybridization also requires the development of a methodology to increase the flexibility of Resilience-aware MLOps tools.

Besides, it is needed to underline the following. The resilience of AI systems should, on the one hand, take into account traditional models and principles (47, 48) of its provision, which are based on evolutionary mechanisms for taking into account and adapting to changes in requirements, parameters of the information and physical environments, tolerance to uncertain failures, taking into account cyberattacks (49, 50). On the other hand, as it was mentioned in (16) the actual models and means of artificial intelligence have the potential of natural resilience, which should be used and which is used. This study is the next step in improving both of these approaches. From the point of view of the general principles of resilience, the idea of a certain resilient wrapper is proposed, which can be adapted for different applications. With regard to the developing the methodology of naturally resilient AI, the proposed solutions can be further improved through a more flexible setting, taking into account the features of the functional part of AI.

5 Conclusion

5.1 Summary

The structure of resilience-aware MLOps for medical diagnostic systems has been proposed. The main novelty lies in the separate work on creating a basic model for normal operating conditions and work on ensuring its resilience and trustworthiness. This is significant for the medical industry, as the developer of the basic model should devote more time to comprehending medical field and the diagnostic task at hand, rather than specializing in a specific area of resilient systems. Therefore, *Post-hoc* Resilience Optimization, *Post-hoc* Predictive Uncertainty Calibration, Uncertainty Monitoring and Graceful Degradation are used as additional stages of MLOps.

Resilience optimization increases robustness to disturbances and speed of adaptation. Fault injection attack, adversarial evasion attack, and concept drift are considered as disturbances. In order to optimize the resilience of the AI-based disease recognition system, additional add-ons are used in the form of adapters and meta-adapters. Meta-adapters are fine-tuned during meta-training based on the results of adaptation to synthetic disturbances. An additional model is added for *Post-hoc* Calibration of Predictive Uncertainty, which is tuned on in-distribution data and out-of-distribution data, to correct predictive confidence in inference mode. Calibrated confidences ensures recognition of a part of unabsorbed disturbances to mitigate their influence, and it improves the trustworthiness.

Experimentally confirmed increase of robustness and speed of adaptation for medical image recognition system during several intervals of the system's life cycle due to the use of resilience optimization and uncertainty calibration stages. The experiments were performed on the DermaMNIST dataset, BloodMNIST and PathMNIST. ResNet-18 as a representative of convolutional networks and MedViT-T as a representative of visual transformers are considered (51). It is shown that transformers are less resilient than a convolutional network, but this may be due to the imperfect architecture of adapters and meta-adapters.

5.2 Limitation

This study is demonstrated on the example of a medical image classification system and does not describe the specifics of using resilience-aware MLOps for self-supervised or reinforcement learning systems. Nevertheless, the general framework of resilience-aware MLOps can be applied to every type of intelligent system. Another limitation may be related to attempts to generalize the information found, which may affect the completeness of the literature review.

Moreover, well-known approaches to Explainable artificial intelligence, as well as Graceful Degradation, are excluded from detailed analysis of their impact on resilience. The paper focuses on the analysis of the general structure of resilience-aware MLOps and the stages of resilience optimization and predictive uncertainty calibration.

5.3 Future research direction

Future research should focus on the development new flexible adapter and meta-adapter architectures as addons for AI system resilience. Special attention should also be paid to the question of providing resilience for self-supervised and reinforcement learning systems. Another important area of research should be the investigation of methods to ensure resilience to new types of attacks on AI systems in the healthcare industry.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://zenodo.org/record/6496656#.ZGMJ--xByys.

Author contributions

VM: Writing – original draft. VK: Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors appreciate the scientific society of the consortium and, in particular, the staff of the Department of Computer Systems, Networks and Cybersecurity (DCSNCS) at the National Aerospace University "KhAI" and the Laboratory of Intellectual Systems (LIS) of the Computer Science Department at the Sumy State University for invaluable inspiration, hard work, and creative analysis during the preparation of this paper. In addition, the authors thanks Ministry of Education and Science of Ukraine for the support to the LIS in the framework of research project No. 0122U000782 "Information technology for providing resilience of artificial intelligence systems to protect cyber-physical systems" (2022–2024) and the support of project No. 0122U001065 "Dependability assurance methods and technologies for intellectual industrial IoT systems" (2022–2023) implemented by the DCSNCS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Testi M, Ballabio M, Frontoni E, Iannello G, Moccia S, Soda P, et al. MLOps: a taxonomy and a methodology. *IEEE Access*. (2022) 10:63606–18. doi: 10.1109/access.2022.3181730

2. Stirbu V, Granlund T, Mikkonen T. Continuous design control for machine learning in certified medical systems. *Softw Qual J*. (2022) 31:307–33. doi: 10.1007/s11219-022-09601-5

3. Subramanya R, Sierla S, Vyatkin V. From dev ops to MLOps: overview and application to electricity market forecasting. *Appl Sci*. (2022) 12:9851. doi: 10.3390/app12199851

4. Gupta R.K., Venkatachalapathy M., Jeberla F.K. (2019). "Challenges in adopting continuous delivery and dev ops in a globally distributed product team: a case study of a healthcare organization" in *2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE)*.

5. Niemelä P., Silverajan B., Nurminen M., Hukkanen J., Järvinen H.-M. (2022) "LAOps: Learning analytics with privacy-aware MLOps" in *Proceedings of the 14th International Conference on Computer Supported Education*.

6. Khattak F.K., Subasri V., Krishnan A., Dolatabadi E., Pandya D., Seyyed-Kalantari L., et al. (2023) MLHOps: Machine learning for healthcare operations. arXiv [Preprint]. doi: 10.48550/ARXIV.2305.02474

7. Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J. Privacy-preserving artificial intelligence in healthcare: techniques and applications. *Comput Biol Med*. (2023) 158:106848. doi: 10.1016/j.compbiomed.2023.106848

8. Fang M. A survey on adversarial attack and defense of deep learning models for medical image recognition. *Meta*. (2023) 4:17. doi: 10.54517/m.v4i1.2156

9. Kalin J. (2022) Defense against the adversarial arts: applying green team evaluations to harden machine learning algorithms from adversarial attacks. Dissertation. Auburn University, Auburn, Alabama.

10. Xu J., Li Z., Du B., Zhang M., Liu J. (2020). "Reluplex made more practical: Leaky ReLU" in *2020 IEEE Symposium on Computers and Communications (ISCC)*.

11. Yang S, Fevens T. Uncertainty quantification and estimation in medical image classification. *Lect Notes Comput Sci*. (2021) 12893:671–83. doi: 10.1007/978-3-030-86365-4_54

12. Zhang X., Jaskolka J. (2022). "Conceptualizing the secure machine learning operations (sec MLOps) paradigm" in *2022 IEEE 22nd International Conference on Software Quality, Reliability and Security (QRS)*. IEEE.

13. Olowononi FO, Rawat DB, Liu C. Resilient machine learning for networked cyber physical systems: a survey for machine learning security to securing machine learning for CPS. *IEEE Commun Surv Tutor*. (2021) 23:524–52. doi: 10.1109/comst.2020.3036778

14. Duddu V, Rajesh Pillai N, Rao DV, Balas VE. Fault tolerance of neural networks in adversarial settings. *IFS*. (2020) 38:5897–907. doi: 10.3233/jifs-179677

15. Duy PT, Tien LK, Khoa NH, Hien DTT, Nguyen AG-T, Pham V-H. DIGFuPAS: deceive IDS with GAN and function-preserving on adversarial samples in SDN-enabled networks. *Comput Secur*. (2021) 109:102367. doi: 10.1016/j.cose.2021.102367

16. Moskalenko V, Kharchenko V, Moskalenko A, Kuzikov B. Resilience and resilient systems of artificial intelligence: taxonomy, models and methods. *Algorithms*. (2023) 16:165. doi: 10.3390/a16030165

17. M. S., A.RNirmala CR, Aljohani M, Sreenivasa BR. A novel technique for detecting sudden concept drift in healthcare data using multi-linear artificial intelligence techniques. *Front Artific Intellig*. (2022) 5:950659. doi: 10.3389/frai.2022.950659

18. Bortsova G, González-Gonzalo C, Wetstein SC, Dubost F, Katramados I, Hogeweg L, et al. Adversarial attack vulnerability of medical image analysis systems: unexplored factors. *Med Image Anal*. (2021) 73:102141. doi: 10.1016/j.media.2021.102141

19. Awais M., Zhou F., Xu H., Hong L., Luo P., Bae S.-H., et al. (2021) "Adversarial robustness for unsupervised domain adaptation" in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*.

20. Gongye C., Li H., Zhang X., Sabbagh M., Yuan G., Lin X., et al. (2020) "New passive and active attacks on deep neural networks in medical applications" in *Proceedings of the 39th International Conference on Computer-Aided Design*.

21. Wang J, Lan C, Liu C, Ouyang Y, Qin T, Lu W, et al. Generalizing to unseen domains: a survey on domain generalization. *IEEE Trans Knowl Data Eng*. (2022) 35:1. doi: 10.1109/tkde.2022.3178128

22. Abusnaina A., Wu Y., Arora S., Wang Y., Wang F., Yang H., et al. (2021) "Adversarial example detection using latent neighborhood graph" in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*.

23. Karimi D, Gholipour A. Improving calibration and out-of-distribution detection in deep models for medical image segmentation. *IEEE Trans Artific Intellig*. (2023) 4:383–97. doi: 10.1109/tai.2022.3159510

24. Huang K, Siegel PH, Jiang A. Functional error correction for robust neural networks. *IEEE J Select Areas Info Theory*. (2020) 1:267–76. doi: 10.1109/jsait.2020.2991430

25. Islam MR, Ahmed MU, Barua S, Begum S. A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Appl Sci*. (2022) 12:1353. doi: 10.3390/app12031353

26. Pourpanah F, Abdar M, Luo Y, Zhou X, Wang R, Lim CP, et al. A review of generalized zero-shot learning methods. *IEEE Trans Pattern Anal Mach Intell*. (2022) 45:1–20. doi: 10.1109/tpami.2022.3191696

27. Baier L, Kühl N, Satzger G, Hofmann M, Mohr M. Handling concept drifts in regression problems—the error intersection approach In: Norbert G, Moreen H, Hanna K, Poustcchi K, editors. *WI2020 Zentrale Tracks*. Berlin: GITO Verlag (2020). p. 210–24.

28. Achddou R., Di Martino J.M., Sapiro G. (2021) "Nested learning for multi-level classification" in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

29. Shapeev A, Gubaev K, Tsymbalov E, Podryabinkin E. Active learning and uncertainty estimation. *Mach Learn Meets Quant Phys*. (2020) 968:309–29. doi: 10.1007/978-3-030-40245-7_15

30. Pianykh OS, Langs G, Dewey M, Enzmann DR, Herold CJ, Schoenberg SO, et al. Continuous learning AI in radiology: implementation principles and early applications. *Radiology*. (2020) 297:6–14. doi: 10.1148/radiol.2020200038

31. Liu Q, Tian Y, Zhou T, Lyu K, Xin R, Shang Y, et al. A few-shot disease diagnosis decision making model based on meta-learning for general practice. *Artif Intell Med*. (2023) 147:102718. doi: 10.1016/j.artmed.2023.102718

32. Jiao B, Guo Y, Gong D, Chen Q. Dynamic ensemble selection for imbalanced data streams with concept drift. *IEEE Trans Neural Netw Learn Syst*. (2024) 35:1278–91. doi: 10.1109/tnnls.2022.3183120

33. Qiu P., Wang Q., Wang D., Lyu Y., Lu Z., Qu G. (2020) "Mitigating adversarial attacks for deep neural networks by input deformation and augmentation" in *2020 25th Asia and South Pacific design automation conference (ASP-DAC)*.

34. Li W., Ning X., Ge G., Chen X., Wang Y., Yang H. (2020). "FTT-NAS: discovering fault-tolerant neural architecture" in *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*.

35. Xu H., Chen Z., Wu W., Jin Z., Kuo S., Lyu M. (2019). "NV-DNN: towards fault-tolerant DNN systems with N-version programming" in *2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*.

36. Vassilev A. (2023). Adversarial machine learning.

37. Ding N, Qin Y, Yang G, Wei F, Yang Z, Su Y, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nat Mach Intellig*. (2023) 5:220–35. doi: 10.1038/s42256-023-00626-4

38. Chen P-Y, Hsieh C-J. Adversarial robustness in meta-learning and contrastive learning In: Chen P-Y, Hsieh C-J, editors. *Adversarial Robustness for Machine Learning* Elsevier, Academic Press (2023) p. 183–98.

39. Shen M. W. (2022) Trust in AI: interpretability is not necessary or sufficient, while black-box interaction is necessary and sufficient. arXiv [Preprint]. doi: 10.48550/ARXIV.2202.05302

40. Hou W., Wang Y., Gao S., Shinozaki T. (2021) "Meta-adapter: efficient cross-lingual adaptation with meta-learning" in *ICASSP 2021-2021 IEEE international conference on acoustics, Speech and Signal Processing (ICASSP)*.

41. Moskalenko VV. Model-agnostic Meta-learning for resilience optimization of artificial intelligence system. *Radio Electron Comput Sci Control*. (2023) 2:79. doi: 10.15588/1607-3274-2023-2-9

42. Li G., Pattabiraman K., DeBardeleben N. (2018). "Tensor FI: a configurable fault injector for tensor flow applications" in *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*.

43. Peng P, Lu J, Xie T, Tao S, Wang H, Zhang H. Open-set fault diagnosis via supervised contrastive learning with negative out-of-distribution data augmentation. *IEEE Trans Industr Inform*. (2023) 19:2463–73. doi: 10.1109/tii.2022.3149935

44. Silva Filho T, Song H, Perello-Nieto M, Santos-Rodriguez R, Kull M, Flach P. Classifier calibration: a survey on how to assess and improve predicted class probabilities. *Mach Learn*. (2023) 112:3211–60. doi: 10.1007/s10994-023-06336-7

45. Kotyan S, Vargas DV. Adversarial robustness assessment: why in evaluation both L0 and L∞ attacks are necessary. *PLoS One*. (2022) 17:e0265723. doi: 10.1371/journal.pone.0265723

46. Dymond J. (2021). Graceful degradation and related fields. arXiv [Preprint]. doi: 10.48550/arXiv.2106.11119

47. Guelfi N. A formal framework for dependability and resilience from a software engineering perspective. *Open Comput Sci*. (2011) 1:294–328. doi: 10.2478/s13537-011-0025-x

48. Ponochovnyi Y, Kharchenko V. Dependability assurance methodology of information and control systems using multipurpose service strategies. *Radioelectron Comput Syst*. (2020) 3:43–58. doi: 10.32620/reks.2020.3.05

49. Lusenko S, Kharchenko V, Bobrovnikova K, Shchuka R. Computer systems resilience in the presence of cyber threats: Taxonomy and ontology. *Radioelectron Comput Syst*. (2020) 1:17–28. doi: 10.32620/reks.2020.1.02

50. Song X., Yang Y., Choromanski K., Caluwaerts K., Gao W., Finn C., et al. (2020). "Rapidly adaptable legged robots via evolutionary meta-learning" in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

51. Manzari ON, Ahmadabadi H, Kashiani H, Shokouhi SB, Ayatollahi A. Med ViT: a robust vision transformer for generalized medical image classification. *Comput Biol Med*. (2023) 157:106791. doi: 10.1016/j.compbiomed.2023.106791

TYPE Original Research
PUBLISHED 20 May 2024
DOI 10.3389/fpubh.2024.1397260

Check for updates

OPEN ACCESS

EDITED BY
Dmytro Chumachenko,
National Aerospace University – Kharkiv Aviation Institute, Ukraine

REVIEWED BY
Luigi Di Biasi,
University of Salerno, Italy
Xinqiang Chen,
Shanghai Maritime University, China
Ricardo Valentim,
Federal University of Rio Grande do Norte, Brazil

*CORRESPONDENCE
Ning Feng
✉ fengning@chinacdc.cn
Xiao-ping Dong
✉ Dongxp@chinacdc.cn

RECEIVED 07 March 2024
ACCEPTED 04 April 2024
PUBLISHED 20 May 2024

CITATION
Liu X-d, Hou B-h, Xie Z-j, Feng N and Dong X-p (2024) Integrating gated recurrent unit in graph neural network to improve infectious disease prediction: an attempt.
Front. Public Health 12:1397260.
doi: 10.3389/fpubh.2024.1397260

Integrating gated recurrent unit in graph neural network to improve infectious disease prediction: an attempt

Xu-dong Liu[1,2], Bo-han Hou[1,2], Zhong-jun Xie[1], Ning Feng[3]* and Xiao-ping Dong[4,5]*

[1]Faculty of Information Technology, Beijing University of Technology, Chaoyang District, Beijing, China, [2]Key Laboratory of Computational Intelligence and Intelligent Systems, Beijing University of Technology, Chaoyang District, Beijing, China, [3]Office of International Cooperation, Chinese Center for Disease Control and Prevention, Chaoyang District, Beijing, China, [4]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Chaoyang District, Beijing, China, [5]National Key-Laboratory of Intelligent Tracking and Forecasting for Infectious Disease, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Chang-Bai, Beijing, China

Objective: This study focuses on enhancing the precision of epidemic time series data prediction by integrating Gated Recurrent Unit (GRU) into a Graph Neural Network (GNN), forming the GRGNN. The accuracy of the GNN (Graph Neural Network) network with introduced GRU (Gated Recurrent Units) is validated by comparing it with seven commonly used prediction methods.

Method: The GRGNN methodology involves multivariate time series prediction using a GNN (Graph Neural Network) network improved by the integration of GRU (Gated Recurrent Units). Additionally, Graphical Fourier Transform (GFT) and Discrete Fourier Transform (DFT) are introduced. GFT captures inter-sequence correlations in the spectral domain, while DFT transforms data from the time domain to the frequency domain, revealing temporal node correlations. Following GFT and DFT, outbreak data are predicted through one-dimensional convolution and gated linear regression in the frequency domain, graph convolution in the spectral domain, and GRU (Gated Recurrent Units) in the time domain. The inverse transformation of GFT and DFT is employed, and final predictions are obtained after passing through a fully connected layer. Evaluation is conducted on three datasets: the COVID-19 datasets of 38 African countries and 42 European countries from worldometers, and the chickenpox dataset of 20 Hungarian regions from Kaggle. Metrics include Average Root Mean Square Error (ARMSE) and Average Mean Absolute Error (AMAE).

Result: For African COVID-19 dataset and Hungarian Chickenpox dataset, GRGNN consistently outperforms other methods in ARMSE and AMAE across various prediction step lengths. Optimal results are achieved even at extended prediction steps, highlighting the model's robustness.

Conclusion: GRGNN proves effective in predicting epidemic time series data with high accuracy, demonstrating its potential in epidemic surveillance and early warning applications. However, further discussions and studies are warranted to refine its application and judgment methods, emphasizing the ongoing need for exploration and research in this domain.

KEYWORDS
artificial intelligence technology, graph neural network, gated recurrent unit, infectious disease, time series prediction

1 Introduction

Multivariate time series forecasting plays a crucial role in various real-world scenarios such as transportation forecasting (1, 2), supply chain management (3), energy allocation (4, 5) and financial investment (6). The time series prediction is involves forecasting future values based on historical data points in a sequential order. This makes the statical method and supervised learning method, comparing with reinforcement learning (7, 8) or unsupervised learning methods, are more suitable for this task. In the field of public health, the problem of acute epidemic forecasting is of great relevance as a typical multivariate time series forecasting: if the future evolution of acute epidemic data can be estimated quickly and accurately for each geographic region, the forecasting results can be used as a reference to help governmental agencies make decisions on policy formulation and material deployment, and thus prevent the development and spread of epidemics.

The field of epidemiology and public health research has witnessed a large number of studies on time series prediction of infectious diseases which revealed the requirement of prediction method in the field of epidemiology and public health research. A selection of notable works has contributed to this progress, showcasing innovative approaches and methodologies for forecasting and managing disease outbreaks. For instance, Pinto et al. (9) applied a regressive model to estimate intervention effects over time by comparing rates of congenital syphilis. Cori et al. (10) presents a novel tool for tracking the spread of diseases by estimating time-varying reproduction numbers. Du et al. (11) focus on the research of serial interval of COVID-19 which contribute to the foundation of transmission dynamics of COVID-19 and is essential for effective prediction and control measures. However, when facing the outbreak of acute epidemic, the traditional transmission dynamics may be uncapable to prediction task. For example, in 2020, Ioannidis et al. (12) found that traditional transmission models failed in forecasting of COVID-19. And many research attempt to apply machine learning method to handle the problem. Dairi et al. (13) compared 7 kinds of neural network in the prediction of the number of COVID-19 cases. In fact, the neural networks were also applied to the prediction problem of other epidemics. Sanchez-Gendriz et al. (14) applied Long Short-Term Memory (LSTM) network in the prediction of dengue outbreak in Natal, demonstrates the potential of neural network in disease surveillance at a local scale. And It is worthwhile to research the potential of neural network in epidemic time series data prediction.

Early time series forecasting research mainly relied on traditional statistical models, including historical average (HA), autoregressive (AR), autoregressive integrated moving average (ARIMA) (15), VAR (16), and fuzzy methods (17). All of these statistical models rely on inherent *a priori* assumptions and require an artificial analysis of the characteristics of the study population to determine the applicability of the forecasting method.

Accurate prediction of multivariate time series data is a challenging type of time series forecasting problem, because both the correlation between the time nodes within each single time series and the correlation between the time series need to be considered comprehensively. During the outbreak of an infectious disease in a certain area, the changes in the number of active cases, on one hand, is related to the number of existing active cases in the locality or previous epidemic data. For instance, the outbreak of some infectious diseases has obvious seasonality, and by referring to the changes in active cases in previous years, one can roughly predict the current trend of active case changes. The data from a certain point or period in the time series is related to the data from the current or future time points, which reflects the correlation between the time nodes within each single time series. On the other hand, the number of active cases in a certain area may be related to the case numbers in neighboring areas or areas with frequent personnel movement. These time series may exhibit leading, lagging, or even synchronous trends, which demonstrates the correlation between different points within the time series. Deep learning models provide new ideas for this problem: on the one hand, Temporal Convolutional Network (TCN) (18) has excellent results in single time series prediction. Recurrent Neural Network (RNN) based methods (19–21) such as LSTM (Long Short-Term Memory) (22), Gated Recurrent Unit (23), Gated Linear Unit (GLU) (24) have good results in single time series prediction. GLU can effectively capture and learn the correlation and nonlinear features among time nodes within a time series (24). Han et al. (25) compared the prediction effects of ARIMA, deep neural network (DNN), and LSTM (Long Short-Term Memory) network for occupational pneumoconiosis data in Tianjin, China, and proved that LSTM (Long Short-Term Memory) can effectively predict occupational pneumoconiosis data, and at the same time has an advantage in prediction accuracy comparing to DNN and ARIMA. There is an advantage in prediction accuracy. However, most of these models ignore the dependencies between multiple variables and can only capture and learn the features within a single time series in isolation, which makes them perform poorly in practical multivariate time series prediction problems.

Meanwhile, in the problem of mining relationships between sequences, Yu et al. used matrix decomposition to model the relationship between multiple time series (26). Discrete Fourier Transform (DFT) is also useful in time series analysis by introducing it. For example, State Frequency Memory Network (27) combines the advantages of DFT and LSTM (Long Short-Term Memory) for stock price prediction; Spectral Residual model (28) utilized DFT to achieve desirable results in time series anomaly detection. Another important aspect of multivariate time series forecasting is modeling the correlation between multiple time series. For example, in traffic prediction tasks, neighboring roads naturally interact with each other. The state-of-the-art models rely heavily on graph convolution networks (GCNs) derived from graph Fourier transform (GFT) theory (29). These models (1, 2) directly stack GCNs and temporal modules (e.g., LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit)), which require predefined graph-structured relationships between sequences. By simultaneously capturing the dependencies between time nodes within each single sequence and between different time series to improve the learning of features of the time series and thus improve the prediction accuracy. Convolutional Neural Network (CNN) has a good performance in learning local features (30). There have been several methods to model spatial features using CNNs (31–35). Ma et al. (34) used deep CNN for traffic speed prediction. Huang et al. (36) tried to use transformer to predict multiple time series variables and obtained good prediction results.

The introduction of GRU (Gated Recurrent Unit) units provides better learning and fitting capabilities in the time domain compared to the linear units used in general GNN (Graph Neural Network) research methods. In addition, the above processes are modularized when implemented. Individual modules can be connected in series by shortcut connection to further improve the prediction accuracy of the neural network by constructing a deep network. Due to the advantages of RNN methods, such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit), comparing with normal feed-forward neural networks, exist clear advantages in time series prediction, there have been a large number of attempts to use RNNs combined with GNNs (Graph Neural Networks), CNNs, or other neural network architectures to predict multivariate time series: Lv et al. (33) combined RNN with CNN, where the RNN are responsible for mining and learning intra-sequence time series within single sequence features, and CNN captures the relationships between sequences. Luo et al. (37) introduced GRU (Gated Recurrent Unit) into GCN to predict the change of gas composition in transformer oil during transformer operation. Zhang et al. (38) proposed ST-ResNet based on residual convolutional network for crowd flow prediction. Shi et al. (20) combines convolutional network with LSTM (Long Short-Term Memory) network to extract spatio-temporal information separately.

Graph neural networks have also yielded many results in capturing dependencies among unstructured data (1, 2, 7, 29, 39–43). DCRNN (1) and STGCN (2) are two of the first studies to introduce graph convolutional networks into spatio-temporal data prediction for better modeling of spatial dependencies. ASTGCN (40) adds an additional attention layer to capture the dynamic change of spatio-temporal dependencies. Adaptive learning of adjacency matrices can also be introduced to solve problems that require predefined graphs for adjacency matrices (35, 39, 41, 42).

However, the previous studies have never processed the time series data from three domains and they have hardly ever been applied in dealing with epidemic time series data predicting problems. But they provide the fundamental framework of the GNN (Graph Neural Network) and GRU (Gated Recurrent Unit) methods and prove the effectiveness of the methods so that we can reform the methods to cater the requirement that introducing GRU (Gated Recurrent Unit) units into GNN (Graph Neural Network) to achieve better results in time series data prediction problems.

The goal of this study is to try to introduce a GRU (Gated Recurrent Unit) layer in the graph neural network to enable the network to better capture and learn the relationship of each single time node within a sequence and the correlation between individual time series. Specifically, after this change, the neural network is able to learn features and make predictions from multivariate time series data in the frequency, spectral, and time domains: after GFT and DFT, it is easier to perform convolution and graphical convolution operations on the time series in the frequency and spectral domains respectively, which in turn allows for more effective predictions. The introduction of GRU (Gated Recurrent Unit) units provides better learning in the time domain compared to linear units used in the general GNN (Graph Neural Network) research methods.

2 Methods

The overall structure of the improved GNN (Graph Neural Network) network (later referred to as GRGNN) with the introduction of GRU (Gated Recurrent Unit) consists of three parts: the preprocessing layer, the GRGNN module layer, and the output layer, and the overall structure is shown in Figure 1.

The input is a multivariate time series data $X = \{x_{it}\} \in \mathbb{R}^{N \times T}$ containing T time nodes in N columns, and before being processed layer by layer by the deep neural network, a graph structure $G = \{X, W\}$ describing the relationship between the input data is first obtained through the smoothing module and the graph building module, where X is the data of each node in the input, and $W_{N \times N}$ is the connection weight matrix between each node. $G = \{X, W\}$ is fed into the GRGNN module layer and the output layer after several rounds of training and learning to obtain the final prediction result $\hat{X} = \left\{ \hat{X}_{T+1}, \hat{X}_{T+2}, \ldots, \hat{X}_{T+H} \right\}$. Where T is the number of time nodes of the input time series data and H is the prediction step size. A mathematical description of the above process can be expressed in Equations 1, 2:

$$G = graphstruct(X) \tag{1}$$

$$\left\{ \hat{X}_{T+1}, \hat{X}_{T+2}, \ldots\ldots, \hat{X}_{T+H} \right\} = F(X, G) \tag{2}$$

2.1 Preprocessing layer

2.1.1 Smoothing processing module

The input data received by the smoothing module are multivariate time series data $\tilde{X} = \{\tilde{x}_{it}\}, i \in \mathbb{R}^N, t \in \mathbb{R}^T$. Due to the different statistical rules of the health statistics departments in each country, some countries will postpone the epidemic data from the weekend to Monday of the following week, which is reflected in the data as a line graph with a weekly cycle showing an obvious "sawtooth waveform." In order to eliminate the negative impact of this problem on the neural network prediction, but also to a certain extent to eliminate some of the noise of the input data, the neural network will be used after the input of a moving window average smoothing processing for a data preprocessing.

The principle of sliding window average smoothing processing is shown in Equation 3, Finally, we will get the smoothed data X after processing the data on day t of the time series will be equal to the average of its data on that day and the data on the n days before it and the n days after it, and $(2n+1)$ is called the window size. Considering the characteristics of the data in this experiment, n is set to 3, that is, the window size is 7.

$$x_t = \frac{1}{2n+1} \sum_{i=t-n}^{t+n} \tilde{x}_i, t = [n, N-n] \tag{3}$$

2.1.2 Graph building blocks

GNN (Graph Neural Network)-based methods need to construct a graph structure $G = \{N, E\}$ before forecasting multivariate time series. In this study, the number of active cases in a certain geographical area is taken as the object of the study, and the data of each subregion in the geographical area is taken as the node N of the

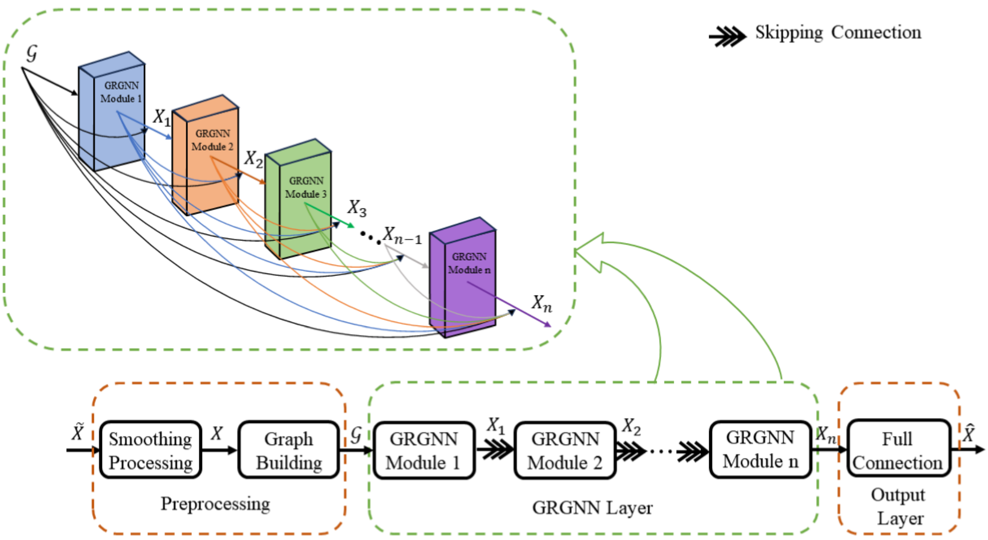


FIGURE 1
The overall structure of the improved GRGNN network.

graph, and the edges E of the graph denote the correlation and the magnitude of the influence of each node on each other. In this study, E is represented by the weight matrix $W \in \mathbb{R}^{N \times N}$. The element $w_{ij}, i \in [0, N-1], j \in [0, N-1]$ in W represents the magnitude of the influence weight of the ith node on the jth node. The graph structure in this study is denoted by $G = \{X, W\}$.

Part of the graph structure can be constructed by humans for observation or through experience or knowledge (e.g., road networks in traffic forecasting, grid systems in electrical energy forecasting). However, in general, there is usually no worthwhile sufficient *a priori* experience to accomplish graph construction artificially. For example, in this study, when dealing with data related to epidemics, there may be a situation where the transmission pathways and characteristics of the epidemics under study have yet to be studied, and the existing research and knowledge about them cannot support the construction of the graph. In order to cope with this situation, the correlation between multiple time series is captured in the preprocessing stage through the self-attention mechanism with the GRU (Gated Recurrent Unit) layer before the data is input into the neural network, and the correlation of each time series is determined in a data-driven manner, which then completes the construction of the required graph structure for the neural network (42).

A specific description of the self-attention mechanism approach for the composition layer is given below:

First of all, the multivariate time series $X \in \mathbb{R}^{N \times N}$ will be fed into the GRU (Gated Recurrent Unit) layer, which calculates the hidden state corresponding to each time node sequentially. The hidden states corresponding to each time nodes are computed sequentially. Then, we use the last hidden state to calculate the weight matrix through the self-attention mechanism. The mathematical description is as Equation 4–6:

$$W^Q = xaviernormal(H) \tag{4}$$

$$W^K = xaviernormal(H) \tag{5}$$

$$\begin{cases} Q = RW^Q \\ K = RW^H \\ W = soft\max\left(\dfrac{QK^T}{\sqrt{d}}\right) \end{cases} \tag{6}$$

where Q and K denote the query and key hiding matrices, respectively, and the magnitude of their values are computed by two learnable parameter matrices W^Q and W^K, respectively, whose initial values are obtained by xavier initialization of the input H (44); d is the size of the dimensions of the two matrices Q and K. The final output adjacency weight square matrix $W \in \mathbb{R}^{N \times N}$ will be used with the input multidimensional time series $X \in \mathbb{R}^{N \times T}$, which forms the final graph structure $G = \{X, W\}$.

2.2 GRGNN layer

The GRGNN layer consists of multiple GRGNN modules stacked in a shortcut connection manner, and the data will be captured and extracted features in the GRGNN modules from the three dimensions of the spectral domain, the frequency domain, and the time domain, respectively. The specific structure of the GRGNN block module, as shown in Figure 2. The features in data will be captured and extracted in three domains of the spectral domain, the frequency domain, and the time domain respectively, in the GRGNN modules. The following is a description of each part of GRGNN block and its functions:

Spectral domain graph convolution is a method that has been widely used in time series forecasting problems. The method has been widely used in time series forecasting problems due to its excellent results in learning potential representations of multiple time series in the spectral domain. The key to the method is the application of the Graph Fourier Transform (GFT) to capture the relationships between time series in the spectral domain. Its output is also a multivariate time series, and the GFT does not explicitly learn the relationship between the data at each time node within a given time series. Therefore, it is necessary to introduce the Discrete Fourier Transform (DFT) to learn the characterization of the input time series in the frequency domain, for example, to capture repetitive features in periodic data.

2.2.1 Frequency domain convolution part

The function of the frequency domain convolution part aims to transfer each individual time series into the frequency domain representation after processing it by DFT, and to learn its features by 1DConv layer in the frequency domain. It consists of four sub-parts in order: discrete Fourier transform (DFT), one-dimensional convolution (1DConv), gated linear unit (GLU), and inverse discrete Fourier transform (IDFT), where DFT and IDFT are used to transform the time series data between time and frequency domains, and 1DConv and GLU are used to learn the features of the time series in the frequency domain. The DFT processing of time sequence usually results in a complex sequence, and the frequency domain convolution is performed on the real part (X_u) and imaginary part (X_u) respectively, and the processing can be expressed by Equation 7 as:

$$M^*\left(\hat{X}_u^*\right) = GLU\left(\theta_\tau^*\left(\hat{X}_\tau^*\right), \theta_\tau^*\left(\hat{X}_u^*\right)\right) = \theta_\tau^*\left(X_u^*\right) \odot \sigma^*\left(\theta_\tau^*\left(X_u^*\right)\right), * \in \{r, i\} \quad (7)$$

Where θ_τ^* denotes the size of the convolution kernel for 1D convolution, $\odot$ denotes the Hadamard product operation, and σ^* denotes the *sigmoid* activation function. The final result $M^r\left(\hat{x}_u^r\right) + iM^i\left(\hat{x}_u^i\right)$ is converted back to the time domain after IDFT processing to participate in the subsequent part of the processing.

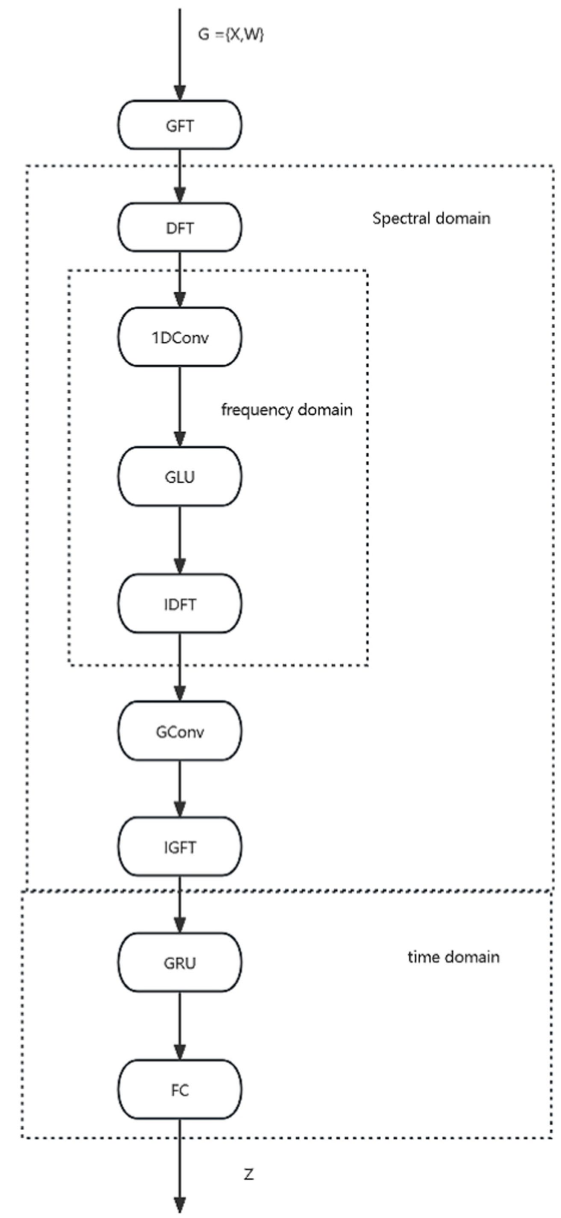


FIGURE 2
The overall structure of GRGNN module.

2.2.2 Spectral domain graph convolution part

Graph Convolution (29) consists of three parts.

First, Transformation of multivariate time series inputs to the spectral domain via GFT. Second, performing a graph convolution operation on the spectral domain graph structure using a graph convolution operator with a convolution kernel to learn. Third, performing the inverse graph Fourier transform (IGFT) on the spectral domain convolution result to generate the final output.

The graph Fourier transform (GFT) (22) is the basic operator for the convolution of spectral domain graphs. It projects the input graph into a standard orthogonal space where the basis is constructed from the eigenvectors of the normalized graph Laplacian. The normalized graph Laplacian matrix (15) can be computed as follows: $L = I_N - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}, I_N \in \mathbb{R}^{N\times N}$ where $I_N \in \mathbb{R}^{N\times N}$ is the unit matrix and D is the degree matrix with diagonal element $D_{ii} = \sum W_{ij}$. Then, the eigenvalue decomposition of the Laplace matrix is performed to obtain $L = U\Lambda U^T$, where $U \in \mathbb{R}^{N\times N}$ is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. After, the GFT, time series will be transformed into complex numbers, for example, three datasets after DFT are shown in Figure 3. For a detailed introduction to the dataset, see section 2.4.1. Given a multivariate time series $X \in \mathbb{R}^{N\times T}$, the GFT and IGFT operators and specific operations are, respectively, denoted as $\mathcal{GF}(X) = U^T X = \hat{X}$ and $\mathcal{GF}^{-1}\left(\hat{X}\right) = U\hat{X}$. The graph convolution operator is realized as a function $g_\Theta(\Lambda)$ of the eigenvalue matrix Λ, where Θ is the convolution kernel parameter.

10.3389/fpubh.2024.1397260

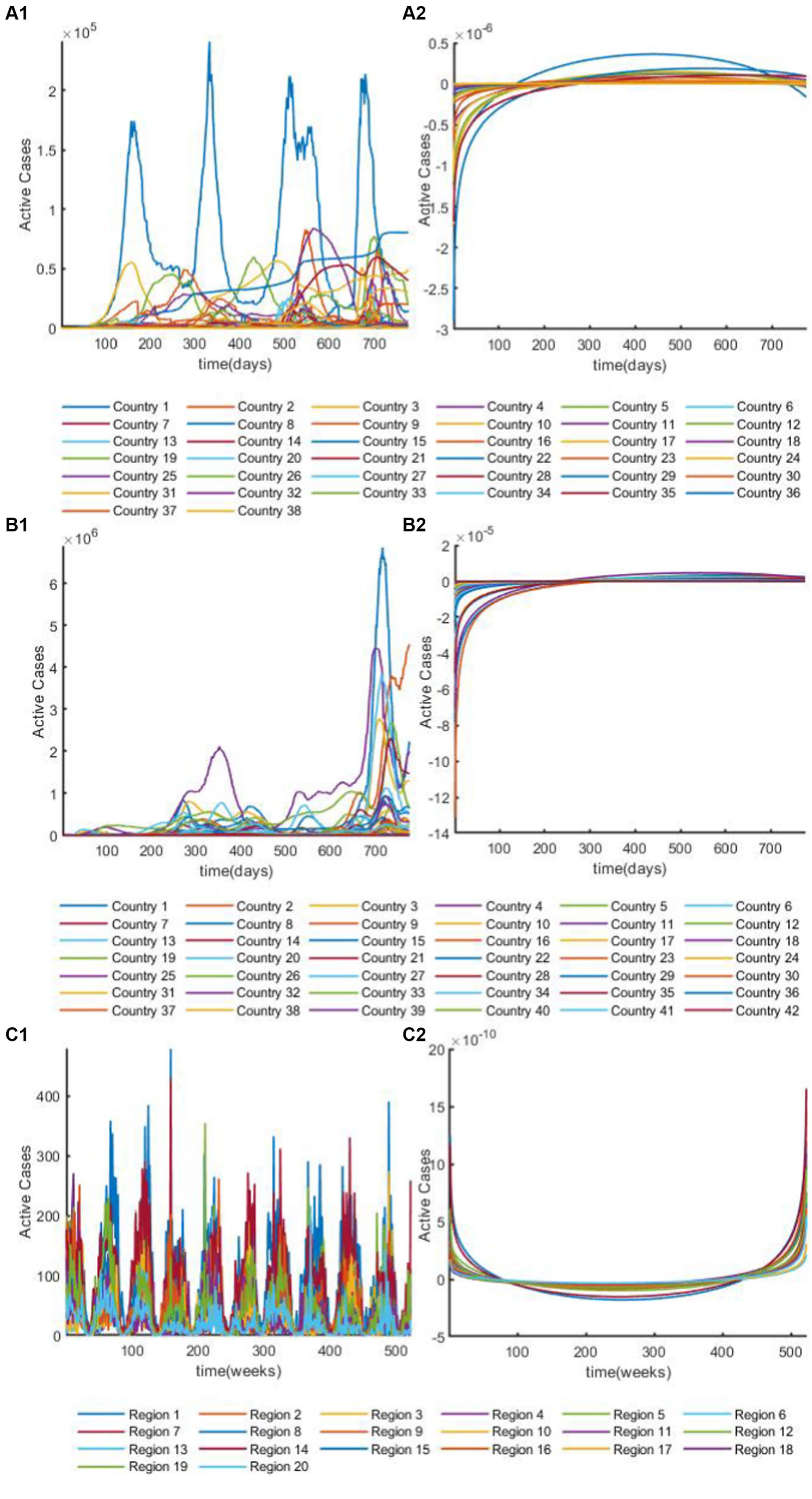


FIGURE 3
The overview plot of time series after discrete Fourier transform. **(A1)** The overview plot of real parts in time series for African dataset after discrete Fourier transform. **(A2)** The overview plot of image parts in time series for African dataset after discrete Fourier transform. **(B1)** The overview plot of real parts in time series for European dataset after discrete Fourier transform. **(B2)** The overview plot of image parts in time series for European dataset after discrete Fourier transform. **(C1)** The overview plot of real parts in time series for Hungarian dataset after discrete Fourier transform. **(C2)** The overview plot of image parts in time series for Hungarian dataset after discrete Fourier transform.

2.2.3 Time domain GRU (gated recurrent units) layer

Recurrent Neural Networks (RNN) are a type of neural networks with an inner recurrent loop structure (23). The reformed GRGNN with its introduction and GRGNN's application on the epidemic field is an important innovation in this study. GRU (Gated Recurrent Unit) processes sequences by traversing the sequence elements and generating a hidden state that contains pattern information related to the historical data, which contains the before-and-after relationships of the sequences. GRUs (Gated Recurrent Units) (23) are a type of recurrent neural networks in which each loop unit adaptively captures dependencies at different time scales. Similar to LSTM (Long Short-Term Memory) units, GRUs (Gated Recurrent Units) have a gating unit that regulates the information within the unit, but do not have a separate storage unit like LSTM (Long Short-Term Memory).

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{8}$$

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{9}$$

$$\tilde{h}_t = \tanh\left(W\left[r_t \cdot h_{t-1}, x_t\right]\right) \tag{10}$$

$$h_t = \left(1 - z_t\right) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{11}$$

The specific mathematical description of GRU (Gated Recurrent Unit) is shown in Equation 8–11, there are only two gate units in GRU (Gated Recurrent Unit), one is reset gate and the other is update gate, and the role of reset gate is similar to that of input gate and forgetting gate in LSTM (Long Short-Term Memory), $(1 - z)$ is equivalent to the input gate, and z is equivalent to the forgetting gate. The GRU (Gated Recurrent Unit) method uses fewer threshold units to accomplish a similar task as the LSTM (Long Short-Term Memory) method, so the GRU (Gated Recurrent Unit) method is usually considered when there is a lack of computational power or a desire to improve the training speed and efficiency of neural network learning. The GRU (Gated Recurrent Unit) method uses fewer gate units than the LSTM (Long Short-Term Memory) method and accomplishes a similar task.

2.3 Implementation and parameter design

The GRGNN method was developed using the Python language based on Pytorch and MATLAB language, the experiments of GRGNN were performed on a deep-learning server with NVIDIA Quadro GV100L GPU *1, Intel Xeon Gold 6,138 CPU *1 and DDR4 32G RAM *8, the operation system of Ubuntu 18.04.6 LTS. The baseline methods were all implemented using MATLAB language. on clearance version.

Hyperparameters such as input length, learning rate, batch size, training time and number of hidden units needed to be set in the GRGNN. Empirically, normalization method was set to z-score, input length to 15, learning rate to 4.7e-4, batch size to 15 and training epoch to 150 and the number of layers to 7. Additionally, the ADAM optimizer was used in the training process.

2.4 Dataset, baseline methods and evaluation indicators

2.4.1 Datasets

In this study, the prediction effect of GRGNN was tested using the 42 European countries' COVID-19 dataset, the 38 African countries' COVID-19 dataset and the 20 Hungarian regions' chickenpox dataset, the overview plots of the datasets are shown in Figure 4 both COVID-19 datasets in this study were collected from publicly available data provided by the Worldometers website (45). Worldometer is run by an international team of developers, researchers, and volunteers with the goal of making world statistics available in a thought-provoking and time relevant format to a wide audience around the world Government's communication channels which makes the data from it more reliable and realistic. The 42 European countries' COVID-19 dataset contains 42 time series, and the length of each time series in the dataset is 776. The 38 African countries' COVID-19 dataset contains 38 time series, and the length of each time series in the dataset is 776. The 20 Hungarian regions' chickenpox dataset contains 20 time series, and the length of each time series in the dataset is 523. Two COVID-19 datasets analyzed during the current study are available in the [Worldometers] repository.[1] The daily active case count data of each country were collected for a total of 776 days from February 15, 2020 to April 1, 2022, and the data were cleaned to exclude from the data that existed for more than 20 days without updating the data, and the data that had a negative number of active cases or other statistical errors, finally we classify the data that met the above requirements to obtain the continental active case dataset. The 20 Hungarian regions' chickenpox dataset was chosen to collect weekly chickenpox diagnosis data from 20 regions in Hungary for 523 weeks from January 3, 2005 to December 29, 2014. The 20 Hungarian regions' chickenpox dataset are available,[2] the dataset was downloaded from Kaggle (46), a website that focuses on providing developers and data scientists with a platform to hold machine learning competitions, host databases, and write and share code. The Hungarian chickenpox dataset, as a typical multivariate time series prediction problem dataset was consisted by the time series collected from the Hungarian Epidemiological Info, a weekly bulletin of morbidity and mortality of infectious disease in Hungary. This dataset was tested on the Kaggle platform with many time series prediction methods and data visualization methods.

2.4.2 Baseline methods

Three widely used neural network architectures; LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), CNN-LSTM and a statistical method, were chosen as the control group in this study, the statistical methods include, weighted moving average method(WMA)

1 https://www.worldometers.info/coronavirus

2 https://www.kaggle.com/datasets/die9origephit/chickenpox-cases-hungary

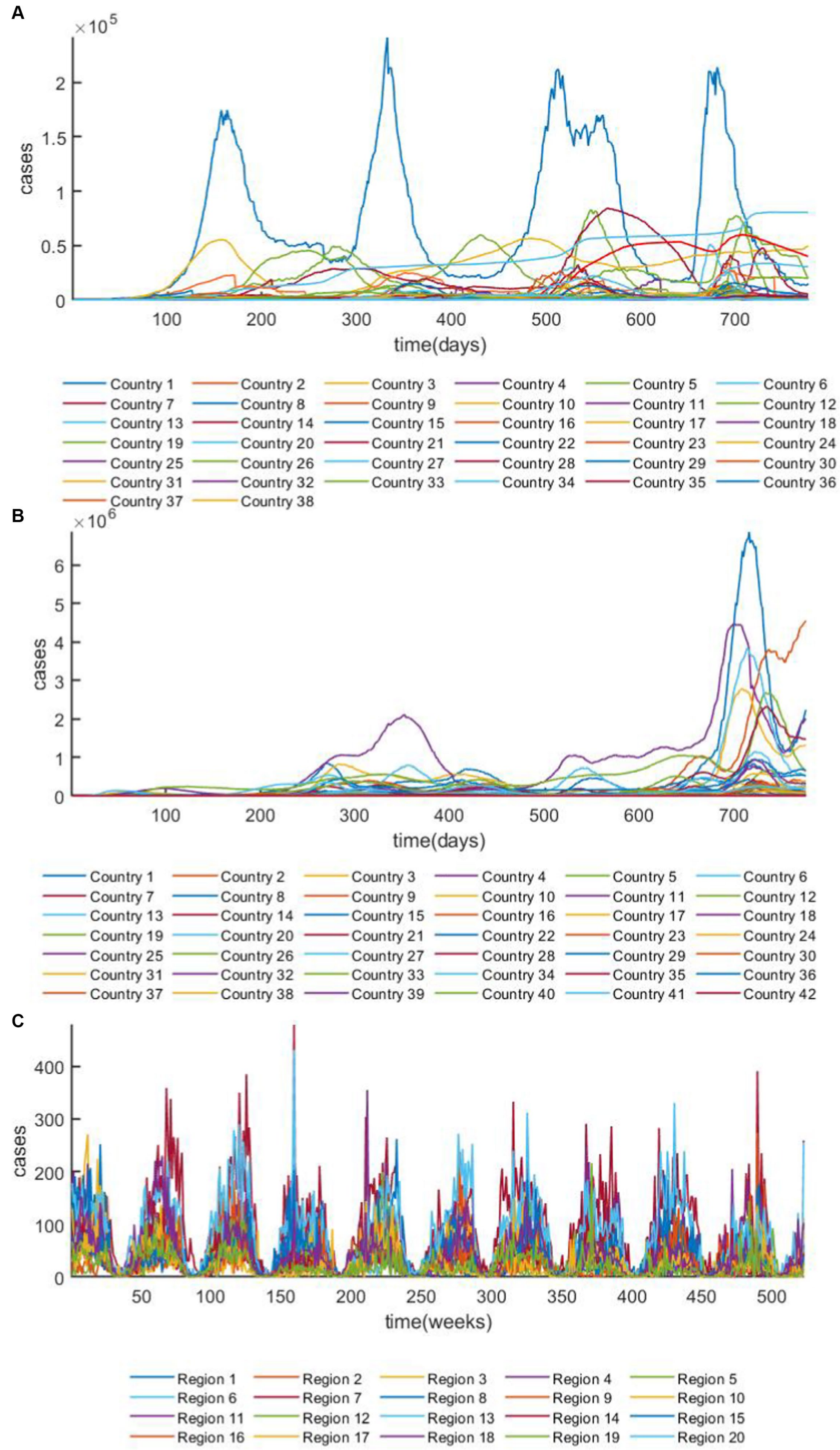


FIGURE 4
The overview plot of the datasets. **(A)** The overview plot of ARMSE of the 38 African countries' COVID-19 dataset. **(B)** The overview plot of the 42 European countries' COVID-19 dataset. **(C)** The overview plot of 20 Hungarian regions' chickenpox dataset.

(47), Gaussian function method (48) and polynomial functions method (48):

The following 7 baseline methods were used to compare the performance with the GRGNN:

ARIMA (15): ARIMA (Autoregressive Integrated Moving Average Model) is a widely applied time series forecasting method, extensively used across various fields. This paper adopts it as a classical statistical prediction method to compare with machine learning approaches for forecasting COVID-19 data in Africa. Its specific definition is given in Equation 12.

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d X_i=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{12}$$

Herein, L represents the lag operator, with $d>0, d\in Z$. The main steps of this method are as follows:

The prediction will finish in 4 steps: step 1, Time series preprocessing. The primary purpose here is to make the input to the ARIMA model a stationary time series. If the data series is non-stationary and exhibits certain growth or decline trends, it is necessary to differentiate the data. Step 2, Establishing the model based on identification rules for time series models. If the partial autocorrelation function of the stationary series is truncated while the autocorrelation function is tailed, the series is suitable for an AR model; if the partial autocorrelation function is tailed while the autocorrelation function is truncated, the series is suitable for an MA model; if both the partial autocorrelation and autocorrelation functions are tailed, the series fits an ARIMA model. Step 3, Determining the order of AR and MA. Utilize the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the orders p and q of AR and MA, respectively. Step 4, ARIMA fitting and forecasting. Fit the ARIMA model, then use the fitted results to forecast the test set. It's worth mentioning that these results are after one differentiation, and the forecasted values need to be restored through inverse differentiation.

weighted moving average method (WMA) (47): the weighted moving average (WMA) method is a time series analysis technique that assigns different weights to historical observations based on their relative importance. Unlike the simple moving average (SMA) method, which assigns equal weight to all observations, the WMA method seeks to accentuate the impact of more recent data and reduce the impact of older data points. The WMA method calculates the weighted average of a sequence of observations, with the most recent values carrying the highest weightings. The weightings assigned to each observation are typically determined by a predefined set of coefficients or by subjective judgment based on the characteristics of the data being analyzed. The WMA method is frequently used in financial market analysis to identify trends and forecast future prices. The specific definition of WMA is given in Equation 13.

$$\hat{X}_{t+1}=\omega_0 X_t+\omega_1 X_{t-1}+\cdots+\omega_N X_{t-N+1} \tag{13}$$

Where $\hat{X}_{t+1}$ denotes the prediction for the time point $t+1$, X_* stands for the observation value, and ω_* stands for the weight of X_*.

Gaussian function fitting method (48): one of the most popular curve fitting algorithms for fitting the time series with a n-order Gaussian function $G(x)$, which has been widely applied in prediction. The specific definition of Gaussian function fitting method is given in Equation 14. In this research we applied 3-order Gaussian function to fitting each time series.

$$G(x)=a_1\cdot e^{-\left(\frac{x-b_1}{c_1}\right)^2}+a_2\cdot e^{-\left(\frac{x-b_2}{c_2}\right)^2}+a_3\cdot e^{-\left(\frac{x-b_3}{c_3}\right)^2} \tag{14}$$

Polynomial function fitting method (48): one of the most popular curve fitting algorithms for fitting the time series with a n-order polynomial function, which has been widely applied in prediction. The specific definition of polynomial function fitting method is given in Equation 15. in this research we applied 5-order polynomial function $G(x)$ to fitting each time series.

$$G(x)=p_1\cdot x^5+p_2\cdot x^4+p_3\cdot x^3+p_4\cdot x^2+p_5\cdot x+p_6 \tag{15}$$

LSTM (Long Short-Term Memory): Long Short-Term Memory networks were first introduced by Hochreiter in 1997 (22). They are a specific form of RNN (Recurrent Neural Network), which is a general term for a series of neural networks that can process sequential data.

Generally, RNNs possess three characteristics: first, they can generate an output at each time step, with connections between hidden units being cyclic; second, they produce an output at each time step, where the output at a given time step is only cyclically connected to the hidden unit of the next time step; third, RNNs contain hidden units with cyclic connections and can process sequential data to produce a single prediction.

LSTM (Long Short-Term Memory) is such a gated RNN. The ingenuity of LSTM (Long Short-Term Memory) lies in the addition of input, forget, and output gates, allowing the self-recurrent weights to vary. Thus, the integration scale at different moments can dynamically change even when the model parameters are fixed, thereby avoiding problems of gradient vanishing or exploding.

Each LSTM (Long Short-Term Memory) unit is composed of a memory cell and three gating units: the input gate, the output gate, and the forget gate. Within this architecture, LSTM (Long Short-Term Memory) attempts to create a controlled flow of information by deciding what information to "forget" and what to "remember," thereby learning long-term dependencies.

$$z_t=\sigma\left(W_z\cdot\left[h_{t-1},x_t\right]\right) \tag{16}$$

$$f_t=\sigma\left(U_g x_t+W_g h_{t-1}+b_g\right) \tag{17}$$

$$\tilde{c}_t=\tanh\left(U_c x_t+W_c h_{t-1}+b_c\right) \tag{18}$$

$$c_t=g_t\odot c_{t-1}+i_t\odot\tilde{c}_t \tag{19}$$

$$o_t=\sigma\left(U_o x_t+W_o h_{t-1}+b_o\right) \tag{20}$$

$$h_t=o_t\odot\tanh\left(c_t\right) \tag{21}$$

More specifically, the input gate i_t alongside the second gate $\tilde{c}_t$ control the new information stored in the memory state c_t at a certain time t. The forget gate f_t controls the disappearance or retention of information from time $t-1$ in the storage unit, while the output gate o_t controls which information can be outputted by the storage unit. Equations 16–21 succinctly describe the operations performed by an LSTM (Long Short-Term Memory) unit.

Herein, x_t represents the input at a certain moment, W_* and U_* represent weight matrices, b_* denotes the bias vector, σ is the sigmoid function, and the operator $\odot$ represents element-wise multiplication. Finally, the hidden state unit h_t, which forms part of the memory cell's output, is calculated as shown in Equation 21.

It is noteworthy that if multiple LSTM (Long Short-Term Memory) layers are stacked together, the memory state c_t and hidden state h_t of each LSTM (Long Short-Term Memory) layer will serve as inputs to the next LSTM (Long Short-Term Memory) layer.

In this paper, the main hyperparameters for the LSTM (Long Short-Term Memory) method are set as follows: the number of iterations is 150, the number of hidden units is 400, the initial learning rate is 0.001, and the optimizer used is ADAM.

GRU (Gated Recurrent Unit): The GRU (Gated Recurrent Unit) is also a type of recurrent neural network. Like LSTM (Long Short-Term Memory), it was developed to address issues related to long-term memory and gradients in backpropagation. Compared to LSTM (Long Short-Term Memory), using GRU (Gated Recurrent Unit) can achieve comparable results and is easier to train, significantly enhancing training efficiency. Therefore, GRU (Gated Recurrent Unit) is often preferred, especially in scenarios with limited computational power or when there is a need to conserve computational resources.

GRU (Gated Recurrent Unit) has only two gating units: a reset gate and an update gate, as shown in Equations 8–11, where x_t represents the input at a given time, W_* represents a weight matrix, σ denotes the tanh function, z_t is the state of the update gate, and r_t is the reset gate. The function of the reset gate is similar to the input and forget gates in LSTM (Long Short-Term Memory), where $1-z_t$ acts like the input gate, and z_t functions as the forget gate. Given that GRU (Gated Recurrent Unit) uses fewer gating units to accomplish tasks similar to those of LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit) is typically considered in situations where computational capacity is limited.

In this paper, the hyper parameters for the GRU (Gated Recurrent Unit) method are set as follows: the number of maximum training epoch is 150, the batch size is 12, the number of hidden units is 400, the initial learning rate is 0.001, and the optimizer used is ADAM.

CNN-LSTM: CNN-LSTM is an advanced neural network architecture that combines Convolutional Neural Networks (CNNs) and LSTMs (Long Short-Term Memory networks) to harness the strengths of both in processing sequential data. This hybrid model is particularly effective for tasks where the input data involves both spatial and temporal dimensions, making it popular in areas such as video analysis, natural language processing, and time series forecasting.

Crucially, to adapt the time series data for the CNN-LSTM architecture, we employ lag features transformation. This involves creating new datasets where each feature corresponds to the original data shifted by values within a specified lag range, effectively capturing temporal dependencies across multiple time steps. These transformed datasets are then organized into matrices, with each column representing a different lagged version of the data, making it suitable for sequential processing by the model.

For the LSTM (Long Short-Term Memory) component, it is the same like the LSTM (Long Short-Term Memory) methods we introduced above. And for the CNN component, the data is initially processed through a sequence folding layer, transforming the sequential input into a format amenable to convolutional operations. This step is pivotal for extracting spatial features from the lagged inputs, which are then unfolded and flattened to preserve the temporal sequence structure, allowing the subsequent LSTM (Long Short-Term Memory) layers to learn long-term dependencies from these extracted features effectively. By meticulously mapping our datasets through these preparatory stages, we ensure that the CNN-LSTM architecture leverages both spatial and temporal dimensions of the data, thereby enhancing the model's forecasting accuracy.

In this paper, the hyper parameters for the CNN-LSTM method are set as follows: the number of maximum training epoch is 150, the batch size is 12, the lag is 8, the number of hidden units [LSTM (Long Short-Term Memory) component] is 150, the initial learning rate is 0.001, and the optimizer used is ADAM.

2.4.3 Evaluation indicators

Average RMSE and average MAE are used as evaluation metrics to measure the magnitude of error in the prediction results:

The average RMSE is calculated by sequentially calculating the RMSE for each of the N countries in the prediction result of the sequence prediction step H. The specific mathematical description is as following Equation 22:

$$\begin{cases} RMSE_i = \sqrt{\dfrac{\sum_{i=1}^{H} y_{pred,i} - y_{obs,i}}{H}} \\ RMSE_{ave} = \dfrac{\sum_{i=1}^{N} RMSE_i}{N} \end{cases} \tag{22}$$

The average MAE is calculated by sequentially calculating the MAE for each of the N countries in the prediction result of the sequence prediction step H, and then calculating the average value, which is mathematically described as following Equation 23:

$$\begin{cases} MAE_i = \dfrac{\sum_{i=1}^{H} \left| y_{pred,i} - y_{obs,i} \right|}{H} \\ MAE_{ave} = \dfrac{\sum_{i=1}^{N} MAE_i}{N} \end{cases} \tag{23}$$

3 Results

Predictions were made using GRGNN, LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), CNN-LSTM, and ARIMA for 42 countries in Europe, 38 countries in Africa, two continents' COVID-19 active case datasets, and Hungary's 20 regions' varicella datasets, respectively. The last 2 weeks (14 days), 3 weeks (21 days), 4 weeks (28 days), 5 weeks (35 days), and 6 weeks (42 days) data were taken as the test set in the prediction, and after dividing the test set, all the data prior to the test set data were divided into the training set and validation set in the ratio of 10:1.

The prediction results of each method for each dataset at different step sizes are shown in Tables 1–5.

As can be seen from Table 1, with a prediction step of 2 weeks (14 days), GRGNN achieves optimal results for both the African and Hungarian datasets, and slightly underperforms the CNN-LSTM method and the ARIMA method for the European dataset. The LSTM (Long Short-Term Memory) method and the GRU (Gated Recurrent Unit) method underperform in all datasets. The CNN-LSTM method performs best in the prediction of the European dataset, and underperforms GRGNN and ARIMA in the African dataset, and performs worse in the Hungarian dataset. The ARIMA method has the best prediction accuracy of the eight methods. The CNN-LSTM method performs best in the prediction of the European dataset, while it does not perform as well as GRGNN and ARIMA on the African dataset, and performs even worse on the Hungarian dataset. The prediction accuracy of the ARIMA method is in the middle of the range of the eight methods. The WMA method can achieve predictions with an accuracy approximately equal to that of ARIMA. Conversely, the Gaussian function method and the polynomial function method produce predictions significantly deviating from the real data, obtaining the lowest accuracies among all eight methods across all three datasets.

As can be seen from Table 2, the comparison of the overall prediction results when extending the prediction step to 3 weeks (21 days) is not much different from that of the prediction step of 2 weeks. The GRGNN method still achieves the best results in the prediction of both the African and Hungarian datasets, and is slightly less accurate in the prediction of the European dataset than the CNN-LSTM and the ARIMA methods. The prediction accuracy of the LSTM (Long Short-Term Memory) method and the GRU (Gated Recurrent Unit) method is the worst two of the eight methods in the African and European datasets. The prediction errors of LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) methods in the African and European datasets are the worst two out of the eight methods. The CNN-LSTM method still performs the best in the prediction of the European dataset. The ARIMA method does not achieve the optimal prediction accuracy but outperforms LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) in the African and European datasets, and outperforms CNN-LSTM in the Hungarian dataset in terms of prediction error. The WMA method still yields slightly inferior results compared to ARIMA and marginally better outcomes than the LSTM (Long Short-Term Memory) method. However, the Gaussian function method and the polynomial function method continue to exhibit the poorest two results.

As can be seen from Table 3, with a prediction step of 4 weeks (28 days), GRGNN still maintains the optimal prediction in the prediction of the African and Hungarian datasets, and the prediction

TABLE 1 Prediction results for each prediction method for each dataset for 2 weeks (14 days).

	African dataset		European dataset		Hungarian dataset	
	ARMSE	AMAE	ARMSE	AMAE	ARMSE	AMAE
GRGNN	683.27	621.38	54568.57	49345.78	28.82	23.64
LSTM	1288.20	1071.58	78093.59	64940.05	29.69	24.57
CNN-LSTM	812.45	790.14	38421.52	31634.68	32.85	26.29
GRU	1115.73	907.52	56406.04	47197.14	32.21	27.66
ARIMA	783.04	657.50	40086.60	42310.69	29.61	23.83
Poly	4620.15	4480.89	301141.17	298245.73	44.11	36.52
Gauss	2289.51	2214.87	109168.62	103422.19	41.55	34.48
WMA	820.68	691.10	70424.89	62469.22	35.13	29.24

TABLE 2 Prediction results for each prediction method for each dataset for 3 weeks (21 days).

	African dataset		European dataset		Hungarian dataset	
	ARMSE	AMAE	ARMSE	AMAE	ARMSE	AMAE
GRGNN	836.26	770.61	75623.18	83044.94	31.47	28.47
LSTM	1375.33	1116.70	113619.62	135365.11	33.62	28.75
CNN-LSTM	915.06	892.35	48978.62	55363.46	35.65	31.46
GRU	1608.06	1260.72	115653.55	144957.77	34.41	28.90
ARIMA	997.03	848.51	68989.77	82938.58	35.22	29.77
Poly	5428.18	5195.21	409270.15	401718.39	29.61	28.93
Gauss	2641.67	2531.15	188754.28	181754.93	36.31	28.60
WMA	1007.55	831.20	119667.10	104058.45	29.70	24.21

 10.3389/fpubh.2024.1397260

results in the European dataset are only slightly inferior to those of the CNN-LSTM method. The prediction errors of the LSTM (Long Short-Term Memory) method and the GRU (Gated Recurrent Unit) method are still poor in the African and European datasets. The CNN-LSTM method still performs optimally in the prediction of the European dataset, but poorly in the European dataset. The ARIMA method is still in the mid-range of the eight prediction mid-range levels. Still performs the best in prediction, but has poor prediction in the Hungarian dataset. The prediction accuracy of the ARIMA method is still in the middle of the range of the 5 prediction mid-range. The performance of the WMA method is slightly inferior to the ARIMA method but slightly superior to the GRU (Gated Recurrent Unit) and LSTM (Long Short-Term Memory) methods. However, the Gaussian method and the polynomial method remain the least effective, exhibiting significant errors in their prediction results.

As can be seen from Table 4, when the prediction step size is set to 5 weeks (35 days), the ranking of the prediction results of each method is not much different from that of the case with a step size of 4 weeks, and it is worth noting that: the main change occurs in the prediction results for European data, and the average index of GRGNN exceeds that of CNN-LSTM as the smallest among the prediction methods. The performance of the WMA method deteriorates rapidly, reaching a point where it only outperforms two other methods. The Gaussian function method and the polynomial function method still remain the poorest performers, with their accuracy indices worsening even further as the prediction steps increase.

As can be seen from Table 5, when the prediction step size is 6 weeks (42 days), the average of the prediction results of GRGNN in the prediction of the European dataset exceeds that of the CNN-LSTM (Long Short-Term Memory) method to become the smallest among the results of each prediction method, and realizes the prediction accuracy of the prediction of each data to be the highest among all eight prediction methods. The prediction error of WMA only slightly exceeds that of LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit), placing its results ahead of both LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit). However, it falls short compared to GRGNN, CNN-LSTM, and ARIMA methods. The polynomial method and Gaussian function method persist as the least effective, exhibiting the highest ARMSE and AMAE values.

The average indictors of the prediction results of each method in each dataset are plotted at different step sizes, as shown in Figure 5.

To enhance the clarity and simplicity of conveying the prediction results, we have selected 5 time series from each dataset, focusing on a prediction step set to 6 weeks (42 days) for visualization. Specifically, we depict the time series data of 5 countries from the 38 African countries' COVID-19 dataset in Figure 6, and the time series of 5 countries from the 42 European countries' COVID-19 dataset in Figure 7, and illustrate the time series of 5 regions from the 20 Hungarian regions' chickenpox dataset in Figure 8. Through these figures, it becomes evident that GRGNN generally captures and mirrors the trends observed in the majority of the time series from the original real-world data.

TABLE 3 Prediction results for each prediction method for each dataset for 4 weeks (28 days).

	African dataset		European dataset		Hungarian dataset	
	ARMSE	AMAE	ARMSE	AMAE	ARMSE	AMAE
GRGNN	748.42	858.05	111743.17	123580.61	27.48	21.75
LSTM	2296.05	2775.97	125159.58	151888.46	28.02	22.36
CNN-LSTM	882.52	921.98	88773.80	97859.95	29.10	22.21
GRU	1718.08	2188.22	188863.87	161955.56	28.88	23.04
ARIMA	921.32	1082.60	136034.82	112387.49	27.55	20.98
Poly	6628.94	6351.37	534460.56	520915.42	35.01	26.59
Gauss	3254.86	3078.51	214257.05	194298.74	33.07	25.51
WMA	1437.04	1228.77	152546.44	132098.87	28.13	22.02

TABLE 4 Prediction results for each prediction method for each dataset for 5 weeks (35 days).

	African dataset		European dataset		Hungarian dataset	
	ARMSE	AMAE	ARMSE	AMAE	ARMSE	AMAE
GRGNN	820.70	1004.61	120230.14	127749.64	27.27	21.45
LSTM	2507.72	3072.21	255698.55	219314.79	27.40	21.55
CNN-LSTM	1536.70	1593.65	128824.08	111020.42	28.14	21.91
GRU	2234.65	2667.76	250900.80	213550.93	28.26	22.00
ARIMA	1537.67	1731.40	150250.91	125333.12	29.96	22.37
Poly	8436.73	8093.56	652543.68	625536.10	34.23	26.48
Gauss	4526.36	4227.41	212263.48	193738.61	32.04	25.00
WMA	2525.29	2301.08	238699.26	209711.46	30.51	22.38

TABLE 5 Prediction results for each prediction method for each dataset for 6 weeks (42 days).

	African dataset		European dataset		Hungarian dataset	
	ARMSE	AMAE	ARMSE	AMAE	ARMSE	AMAE
GRGNN	1545.62	1763.28	124665.83	133453.75	25.51	19.17
LSTM	3418.20	4090.81	308407.33	367230.08	27.58	22.07
CNN-LSTM	1657.79	1810.84	124829.94	153435.48	26.18	20.08
GRU	4648.19	5709.85	232157.67	269820.41	25.72	20.48
ARIMA	2673.66	3035.86	188922.10	229932.70	27.45	20.27
Poly	10843.29	10305.87	739501.24	697691.22	33.02	24.93
Gauss	4735.57	4382.75	251950.31	218247.07	30.33	23.09
WMA	3435.84	3093.05	426603.38	363924.52	27.98	20.14

4 Discussion

Observing Tables 1–5, it can be found that for the prediction results of the data of the 38 African countries' COVID-19 dataset and the 20 Hungarian regions' chickenpox dataset, GRGNN is able to achieve better prediction results compared with other prediction methods at different prediction steps, and the average RMSE and average MAE of its prediction results are the smallest among the prediction methods at different steps, which indicates that GRGNN is able to capture and learn the features in the data better than the three neural network methods and statistical methods in the baseline methods, and make accurate predictions.

Observing Figures 6, 8, it becomes apparent that for African dataset and Hungarian dataset, the prediction results of GRGNN consistently align with the developmental trend of the original time series, albeit with varying degrees of error. This observation suggests that GRGNN, to a certain extent, can predict the developmental trends within the datasets.

The prediction errors at different step lengths are compared with the step lengths on each dataset, as shown in Figure 5 and it can be found that the prediction errors for the African data generally increase with the extension of the prediction step lengths, and the errors of the GRGNN method increase relatively less with the extension of the prediction step lengths compared with the others, which indicates that the GRGNN compared with the three neural network in the baseline methods and statistical methods to capture and learn more adequately the relationships and features among the temporal nodes of the time series. This also indicates that GRGNN learns the data in three dimensions: time domain, frequency domain and spectral domain, compared to the seven comparative forecasting methods that only learn and capture the data in the time domain, which proves that GRGNN can capture more features in the data, better grasp the overall trend of the data, and realize more accurate medium- and long-term forecasting results for the two datasets, namely, the data of the 38 countries in Africa and the data of the 20 regions in Hungary. The results demonstrate that this allows GRGNN to explore more features in the data, better grasp the general trend of the data, and thus achieve more accurate medium-term and long-term predictions for the 38 African countries' COVID-19 dataset and the 20 Hungarian regions' chickenpox dataset.

For the 20 Hungarian regions' chickenpox dataset, it should be separately stated that since the data in this dataset are weekly collected, the actual predictions obtained at the same prediction step size are less than other two dataset. Therefore, as shown in Figures 5E,F, when the prediction step length is extended from 2 weeks to 3 weeks, each prediction method shows an increase in prediction error, whereas the error of each prediction method except ARIMA method shows a decreasing trend when the step length is extended from 4 weeks to 6 weeks. Meanwhile, GRGNN was able to achieve better results than the other seven comparison methods in both average RMSE and average MAE. This indicates that GRGNN and the neural network prediction methods in the baseline methods can realize the capture of the overall trend characteristics of the data, which in turn shows that the prediction accuracy will be improved when the data prediction step length is extended to a certain length, and compared with the seven comparative methods, GRGNN achieves more accurate prediction results, which indicates that GRGNN is more adequate than the other seven methods for the capture and learning of the overall trend characteristics of the data. This indicates that GRGNN is more adequate than the other seven methods for capturing and learning the general trend features of the data.

Finally, the GRGNN do not always make the most accurate prediction, as can be seen from Figures 5C,D, for the prediction experiments of 42 European countries, the errors of each prediction method are much larger than the errors of the prediction results for the African data, and the indicators of each prediction result under the same hyper-parameters mostly reaches 10,000 counts or even 100,000 counts, in which case the CNN-LSTM method has the best prediction results in the experiments with the prediction step lengths of 14, 21, and 28 days, and its indicators are the smallest values among the eight prediction methods, but these two metrics of CNN-LSTM become larger with the increase of the prediction step. When the prediction step is extended to 35 days, the average of CNN-LSTM is still the smallest among the eight methods, but the mean becomes sub-optimal, and the optimal value is obtained from the prediction results of GRGNN. When the prediction step size is increased to 42 days, the prediction result of GRGNN becomes optimal in both indicators. The prediction results of each prediction method in the experiment are not satisfactory in the European dataset, which may be caused by the inadequacy of the type of data collected and the insufficient amount of data collected for this phenomenon. Data inapplicability is an insurmountable problem for data-driven methods, and if the applicability of the prediction methods to the data cannot be assessed, this will greatly limit the application prospects of the

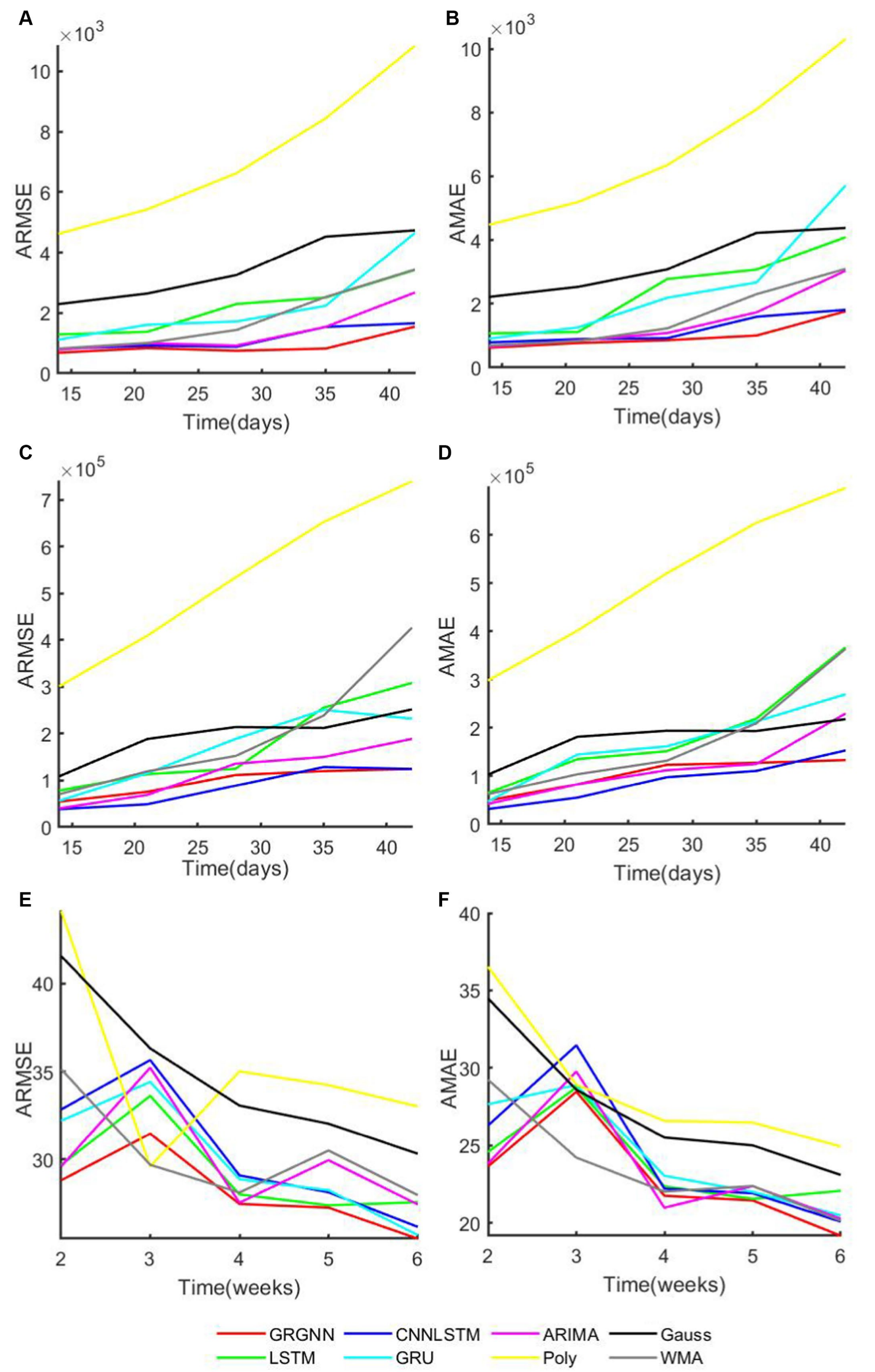


FIGURE 5
The overview plot of evaluation indicator of datasets **(A)** the overview plot of ARMSE of the 38 African countries' COVID-19 dataset. **(B)** The overview plot of AMAE of the 38 African countries' COVID-19 dataset. **(C)** the overview plot of ARMSE of the 42 European countries' COVID-19 dataset. **(D)** The overview plot of AMAE of the 42 European countries' COVID-19 dataset. **(E)** the overview plot of ARMSE of the20 Hungarian regions' Chickenpox dataset. **(F)** The overview plot of AMAE of the 20 Hungarian regions' Chickenpox dataset.

prediction methods. Therefore, there is a need to discuss the applicability of GRGNN to different data:

Plotting the heatmap of the weight matrix (W) for each dataset in Figure 9, where the blocks in the plot represent the correlation between the time series marked by the x-axis and y-axis the lighter the color of the block is the related closer the time series are. it can be observed that the accuracy of GRGNN is linked to the correlation among time series in the datasets. In cases such as the African and

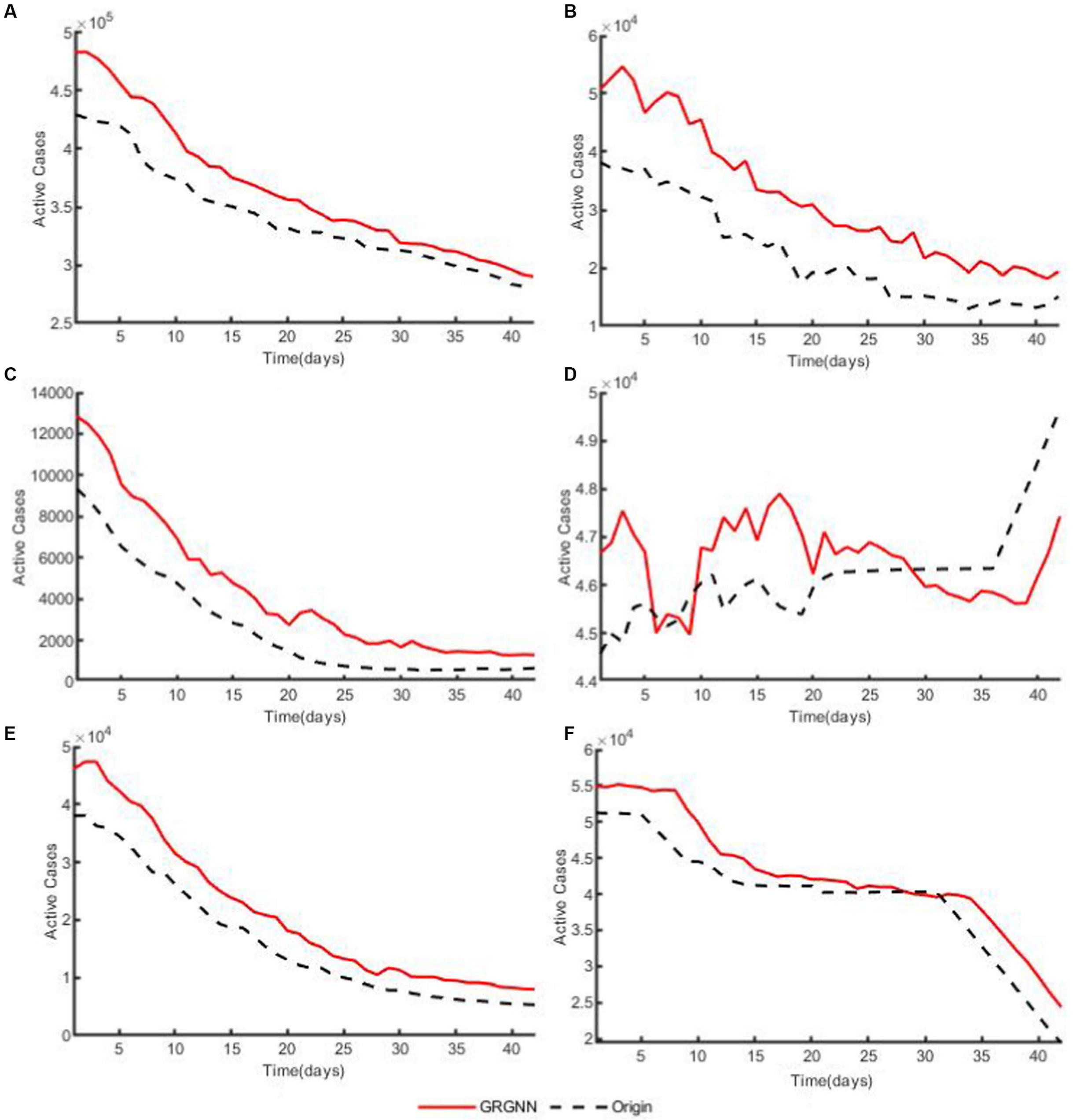


FIGURE 6
The plots of original data and prediction result for countries from the 38 African countries' COVID-19 dataset of GRGNN. **(A)** The plot of original data and prediction result for the total cases of the 38 African countries' COVID-19 dataset of GRGNN. **(B)** The plot of original data and prediction result for Country1 from the 38 African countries' COVID-19 dataset of GRGNN. **(C)** The plot of original data and prediction result for County2 from the 38 African countries' COVID-19 dataset of GRGNN. **(D)** The plot of original data and prediction result for Country3 from the 38 African countries' COVID-19 dataset of GRGNN. **(E)** The plot of original data and prediction result for Country4 from the 38 African countries' COVID-19 dataset of GRGNN. **(F)** The plot of original data and prediction result for Country5 from the 38 African countries' COVID-19 dataset of GRGNN.

Hungarian datasets in this research, where the correlation between time series is relatively close, GRGNN exhibits accurate predictions and the ability to forecast the developmental trend of the time series. However, when facing datasets like the European dataset in this research, where the correlation among time series is less pronounced, GRGNN struggles to achieve a more accurate prediction compared to other neural network methods.

We find that for the weight matrix W obtained after preprocessing of the dataset, the average of the sum of the weights of each node over the other nodes is calculated, as shown in Table 6, and it can be found that when the average value tends to 1 then the dataset yields better prediction results by GRGNN.

Therefore, we hypothesize that if the average value of the sum of the weights of each node in the weight matrix over the other nodes converges

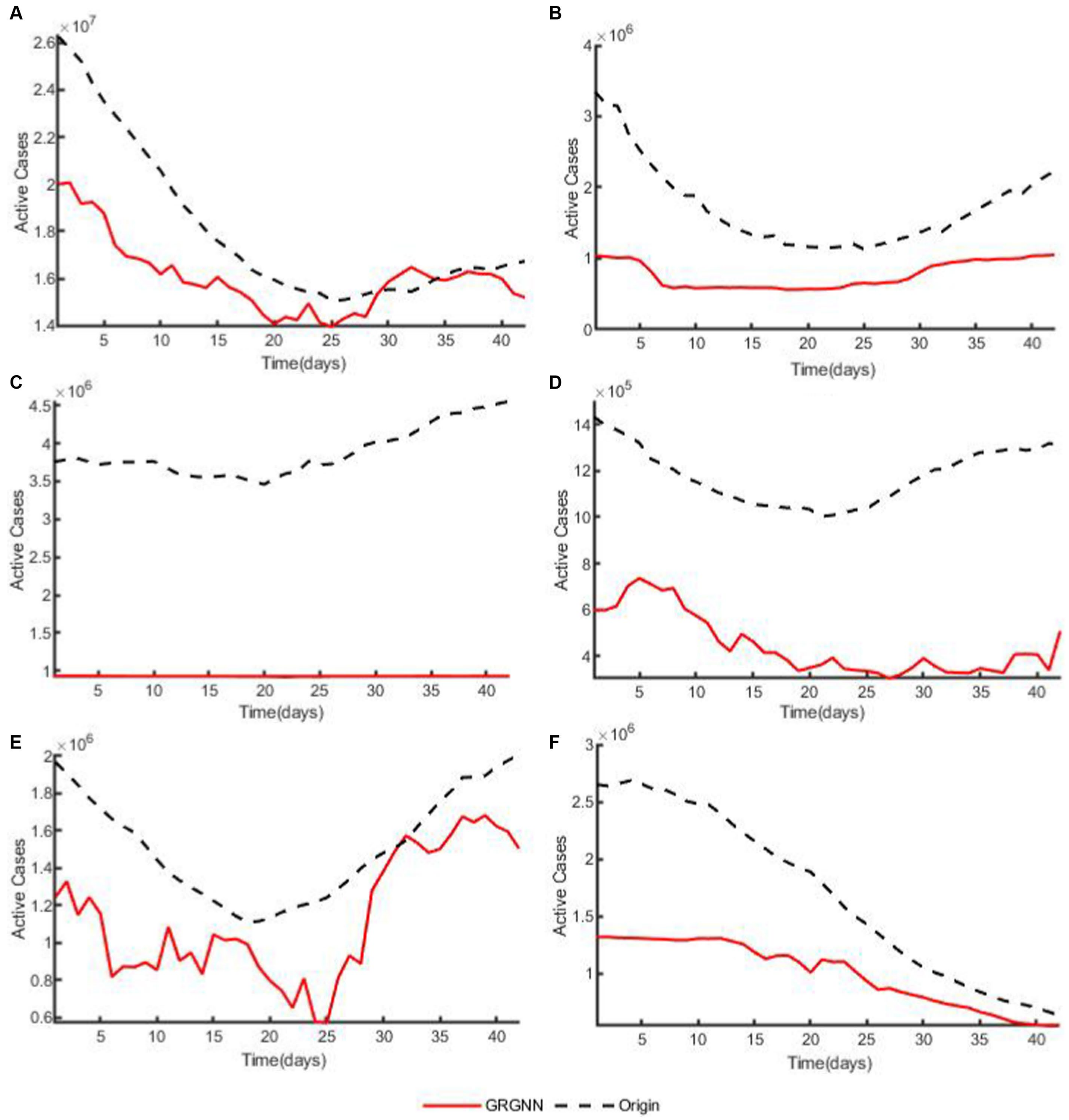


FIGURE 7
The plots of original data and prediction result for countries from the 42 European countries' COVID-19 dataset of GRGNN. **(A)** The plot of original data and prediction result for the total cases of the 42 European countries' COVID-19 dataset of GRGNN. **(B)** The plot of original data and prediction result for Country1 from the 42 European countries' COVID-19 dataset of GRGNN. **(C)** The plot of original data and prediction result for County2 from the 42 European countries' COVID-19 dataset of GRGNN. **(D)** The plot of original data and prediction result for Country3 from the 42 European countries' COVID-19 dataset of GRGNN. **(E)** The plot of original data and prediction result for Country4 from the 42 European countries' COVID-19 dataset of GRGNN. **(F)** The plot of original data and prediction result for Country5 from the 42 European countries' COVID-19 dataset of GRGNN.

to 1, then the dataset will yield better prediction results by GRGNN. As a matter of fact, there are some researches to construct the graph by SoftMax and other methods to make the average value of the sum of the weights of each node in the weight matrix of each node to other nodes converge to 1 (40), but this hypothesis is only based on the observation of the phenomenon shown in the experimental results, and the mathematical proofs and the verification of the actual additional experiments are still need to be further supplemented.

This paper is significantly innovative: the main focus of this study is to realize the ability of the network to analyze datasets in multiple dimensions in the time, spectral, and frequency domains by introducing a GRU (Gated Recurrent Unit) layer in the GNN (Graph Neural Network) network. This gives the following advantages to the neural network used in this study: Firstly, the multiple-input multiple-output temporal prediction of multiple time series variables is more efficient compared to the single-input single-output prediction method of a single time series variable; Secondly, due to the introduction of the GRU (Gated Recurrent Unit) layer, it yields a more accurate prediction in terms of prediction accuracy; and Thirdly, as a data-driven method, it does not require human *a priori* knowledge as

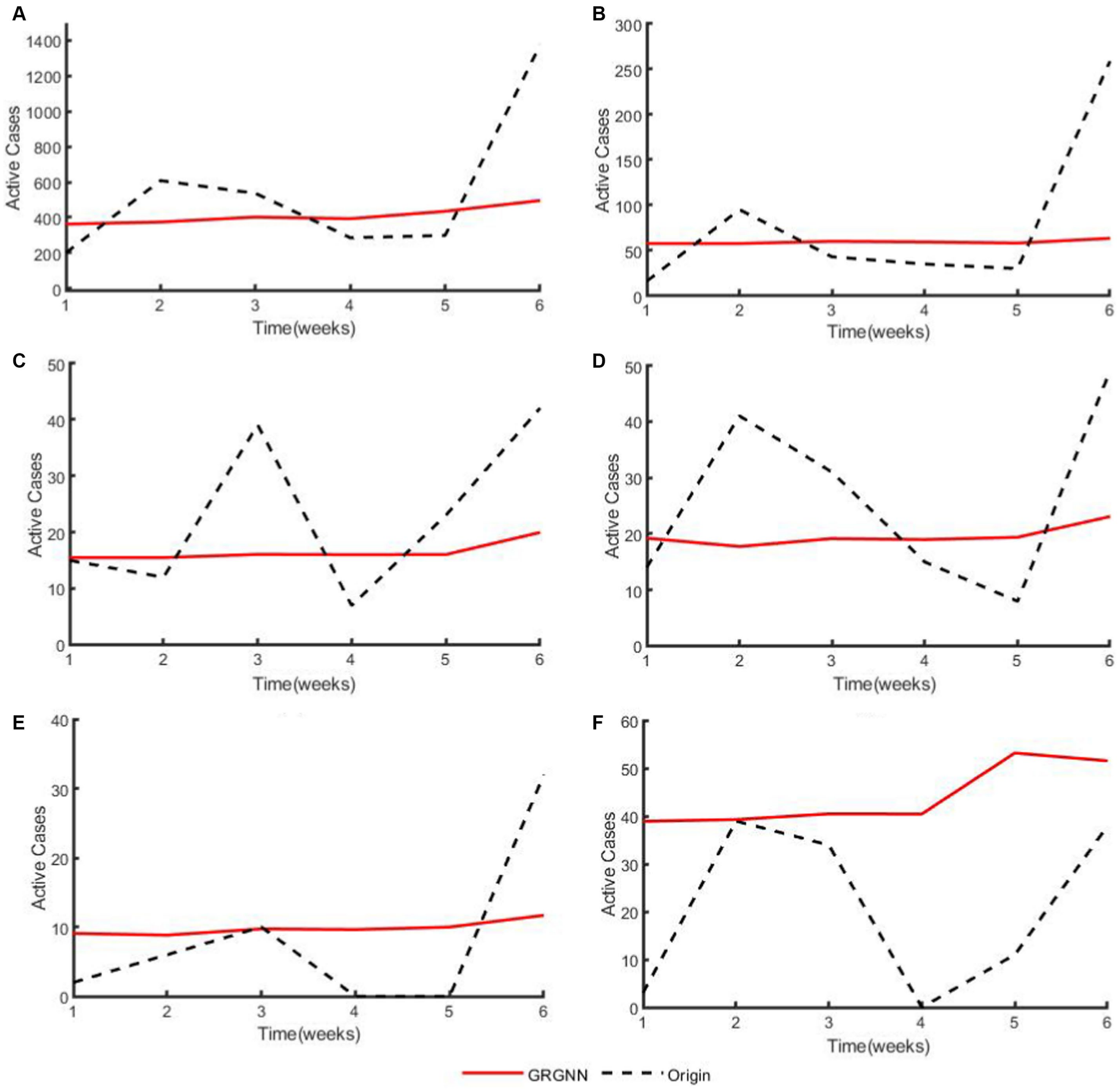


FIGURE 8
The plots of original data and prediction result for regions from the 20 Hungarian regions' Chickenpox dataset of GRGNN. **(A)** The plot of original data and prediction result for the total cases of the 20 Hungarian regions' Chickenpox dataset of GRGNN. **(B)** The plot of original data and prediction result for Region1 from the 20 Hungarian regions' Chickenpox dataset of GRGNN. **(C)** The plot of original data and prediction result for Region2 from the 20 Hungarian regions' Chickenpox dataset of GRGNN. **(D)** The plot of original data and prediction result for Region3 from the 20 Hungarian regions' Chickenpox dataset of GRGNN. **(E)** The plot of original data and prediction result for Region4 from the 20 Hungarian regions' Chickenpox dataset of GRGNN. **(F)** The plot of original data and prediction result for Region5 from the 20 Hungarian regions' Chickenpox dataset of GRGNN.

a basis, which makes it easy to migrate the application to the other data processing.

5 Conclusion

In this paper, gated recurrent units are attempted to be introduced into graph neural network, enabling graph neural networks to capture and learn features from data in three dimensions, namely, null, frequency, and time domains, which is utilized to produce notable results in the epidemic data prediction problem, which is a typical multivariate time series prediction problem. Compared with classical prediction methods, graph neural networks, as an multiple-input-multiple-output method, can quickly and easily construct graphs for multiple time series and realize effective prediction in a data-driven manner. In terms of prediction accuracy, when the predicted multivariate correlation reaches a certain level (specifically, the phenomenon observed in this study is that the closer the average of the sum of the connection weights of each node to the other nodes tends to be 1, the better the prediction results obtained from the GRGNN for the dataset), the graph neural network with the introduction of gated recurrent units can achieve more accurate predictions in medium-term or long-term forecasting.

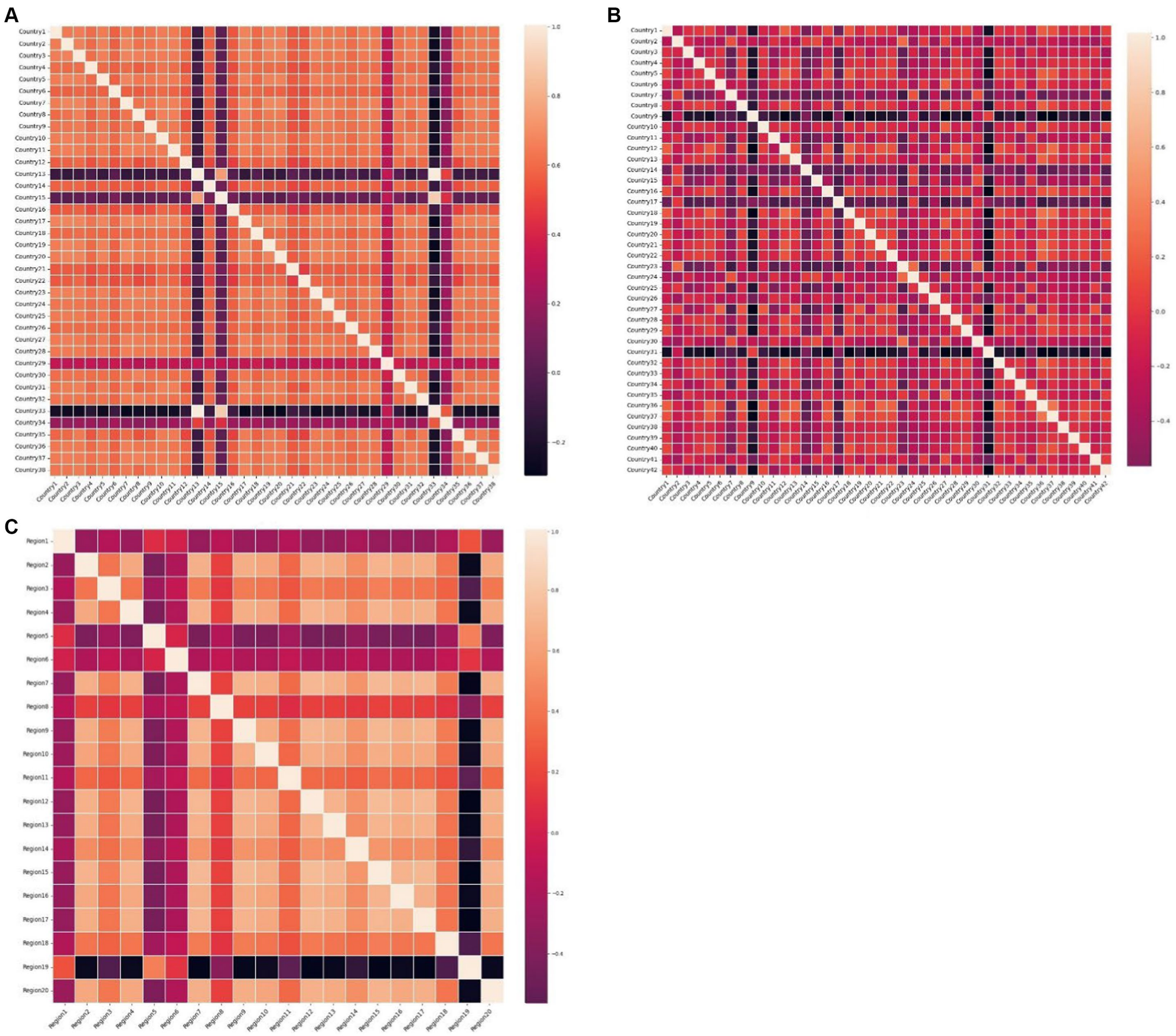


FIGURE 9
The heat maps of the weight matrices of datasets **(A)** The heat map of the weight matric of the 38 African countries' COVID-19 dataset. **(B)** The heat map of the weight matric of the 42 European countries' COVID-19 dataset. **(C)** The heat map of the weight matric of the 20 Hungarian regions' Chickenpox dataset.

TABLE 6 The average node sum weights of each dataset.

	African dataset	European dataset	Hungarian dataset
average node sum weights	1.10	0.78	0.96

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

X-dL: Writing – original draft, Writing – review & editing. B-hH: Writing – original draft, Writing – review & editing. Z-jX: Writing – original draft. NF: Writing – original draft, Writing – review & editing. X-pD: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by National Natural Science Foundation of China (82341035), the Grant (2019SKLID603) from the State Key Laboratory for Infectious Disease Prevention and Control (Current name is National Key-Laboratory of Intelligent Tracking and Forecasting for Infectious Disease), China CDC.

Acknowledgments

The authors gratefully acknowledge Tao Hong, Mingkuan Feng, Yi Yang, for their assistance with data collection and inspiration of the idea in this study. And sincerely acknowledge Olasehinde Toba Stephen, for helping improve the manuscript from language perspective.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. *arXiv preprint arXiv*. (2017) 1707:01926. doi: 10.48550/arXiv.1707.01926

2. Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv*. (2017) 1709:04875. doi: 10.48550/arXiv.1709.04875

3. Yang S, Zhang Z, Zhou J, Wang Y, Sun W, Zhong X, et al. Financial risk analysis for SMEs with graph-based supply chain mining. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (2021) (pp. 4661–4667).

4. Khodayar M, Wang J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. *IEEE Transactions on Sustain Energy*. (2018) 10:670–81. doi: 10.1109/TSTE.2018.2844102

5. Wu Q, Zheng H, Guo X, Liu G. Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks. *Renew Energy*. (2022) 199:977–92. doi: 10.1016/j.renene.2022.09.036

6. Wang J, Zhang S, Xiao Y, Song R. A review on graph neural network methods in financial applications. *arXiv preprint arXiv*. (2021) 2111:15367. doi: 10.48550/arXiv.2111.15367

7. Chen W, Chen L, Xie Y, Cao W, Gao Y, Feng X. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (2020) (Vol. 34, pp. 3529–3536).

8. Chen X, Liu S, Zhao J, Wu H, Xian J, Montewka J. Autonomous port management based AGV path planning and optimization via an ensemble reinforcement learning framework. *Ocean Coast Manag*. (2024) 251:107087. doi: 10.1016/j.ocecoaman.2024.107087

9. Pinto R, Valentim R, da Silva LF, de Souza GF, de Moura Santos TG, de Oliveira CA, et al. Use of interrupted time series analysis in understanding the course of the congenital syphilis epidemic in Brazil. *Lancet Regional Health–Americas*. (2022) 7:100163. doi: 10.1016/j.lana.2021.100163

10. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*. (2013) 178:1505–12. doi: 10.1093/aje/kwt133

11. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg Infect Dis*. (2020) 26:1341–3. doi: 10.3201/eid2606.200357

12. Ioannidis JP, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. *Int J Forecast*. (2022) 38:423–38. doi: 10.1016/j.ijforecast.2020.08.004

13. Dairi A, Harrou F, Zeroual A, Hittawe MM, Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. *J Biomed Inform*. (2021) 118:103791. doi: 10.1016/j.jbi.2021.103791

14. Sanchez-Gendriz I, de Souza GF, de Andrade IG, Neto AD, de Medeiros TA, Barros DM, et al. Data-driven computational intelligence applied to dengue outbreak forecasting: a case study at the scale of the city of Natal, RN-Brazil. *Sci Rep*. (2022) 12:6550. doi: 10.1038/s41598-022-10512-5

15. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. (2003) 50:159–75. doi: 10.1016/S0925-2312(01)00702-0

16. Zivot E, Wang J. Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®*. New York, NY: Springer (2006). doi: 10.1007/978-0-387-21763-5_11

17. Yang H, Jiang Z, Lu H. A hybrid wind speed forecasting system based on a 'decomposition and ensemble' strategy and fuzzy time series. *Energies*. (2017) 10:1422. doi: 10.3390/en10091422

18. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv*. (2018) 1803:01271. doi: 10.48550/arXiv.1803.01271

19. Yu R, Zheng S, Anandkumar A, Yue Y. Long-term forecasting using higher order tensor RNNs. *arXiv preprint arXiv*. (2017) 1711:00073. doi: 10.48550/arXiv.1711.00073

20. Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Adv Neural Inf Proces Syst*. Berlin, Heidelberg: Springer. (2015):28:802–810. doi: 10.5555/2969239.2969329

21. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell G. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv*. (2017) 1704:02971. doi: 10.48550/arXiv.1704.02971

22. Graves A. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*. Berlin, Heidelberg: Springer (2012).

23. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv*. (2014) 1406:1078. doi: 10.48550/arXiv.1406.1078

24. Dauphin YN, Fan A, Auli M, Grangier D. Language modeling with gated convolutional networks. In *International conference on machine learning*. (2017) (pp. 933–941). PMLR.

25. Lou HR, Wang X, Gao Y, Zeng Q. Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China. *BMC Public Health*. (2022) 22:2167. doi: 10.1186/s12889-022-14642-3

26. Yu HF, Rao N, Dhillon IS. Temporal regularized matrix factorization for high-dimensional time series prediction. *Adv Neural Inf Proces Syst*. (2016) 29:847–855. doi: 10.5555/3157096.3157191

27. Zhang L, Aggarwal C, Qi GJ. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (2017) (pp. 2141–2149).

28. Ren H, Xu B, Wang Y, Yi C, Huang C, Kou X, et al. Time-series anomaly detection service at microsoft. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (2019) (pp. 3009–3017).

29. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv*. (2016) 1609:02907. doi: 10.1145/3292500.3330680

30. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. (2017) 60:84–90. doi: 10.1145/3065386

31. Yang D, Li S, Peng Z, Wang P, Wang J, Yang H. MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion. *IEICE Trans Inf Syst*. (2019) E102.D:1526–36. doi: 10.1587/transinf.2018EDP7330

32. Yu H, Wu Z, Wang S, Wang Y, Ma X. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*. (2017) 17:1501. doi: 10.3390/s17071501

33. Lv Z, Xu J, Zheng K, Yin H, Zhao P, Zhou X. Lc-rnn: a deep learning model for traffic speed prediction. *IJCAI*. (2018) 2018:27. doi: 10.24963/ijcai.2018/482

34. Ma X, Dai Z, He Z, Ma J, Wang Y, Wang Y. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*. (2017) 17:818. doi: 10.3390/s17040818

35. Wu Z, Pan S, Long G, Jiang J, Zhang C. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv*. (2019) 1906:00121. doi: 10.48550/arXiv.1906.00121

36. Huang L, Mao F, Zhang K, Li Z. Spatial-temporal convolutional transformer network for multivariate time series forecasting. *Sensors*. (2022) 22:841. doi: 10.3390/s22030841

37. Luo D, Chen W, Fang J, Liu J, Yang J, Zhang K. GRU-AGCN model for the content prediction of gases in power transformer oil. *Front Energy Res*. (2023) 11:1135330. doi: 10.3389/fenrg.2023.1135330

38. Zhang J, Zheng Y, Qi D, Li R, Yi X, Li T. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artif Intell*. (2018) 259:147–66. doi: 10.1016/j.artint.2018.03.002

39. Song C, Lin Y, Guo S, Wan H. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (2020) (Vol. 34, pp. 914–921).

40. Guo S, Lin Y, Feng N, Song C, Wan H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (2019) (Vol. 33, pp. 922–929).

41. Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C. Connecting the dots: multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (2020) (pp. 753–763).

42. Cao D, Wang Y, Duan J, Zhang C, Zhu X, Huang C, et al. Spectral temporal graph neural network for multivariate time-series forecasting. *Adv Neural Inf Proces Syst*. (2020) 33:17766–78. doi: 10.48550/arXiv.2103.07719

43. Zheng C, Fan X, Wang C, Qi J. Gman: a graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence* (2020) (Vol. 34, pp. 1234–1241).

44. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (2010) (pp. 249–256). *JMLR Workshop and Conference Proceedings*.

45. COVID - Coronavirus Statistics. Worldometer. (2023). Available from: https://www.worldometers.info/coronavirus/

46. Kaggle: Your Machine Learning and Data Science Community. (2023). Available from: https://www.kaggle.com/

47. Anggraini P, Amin M, Marpaung N. Comparison of weighted moving average method with double exponential smoothing in estimating production of oil palm fruit. *Building of Informatics, Technology and Science (BITS)*. (2022) 4:705–22. doi: 10.47065/bits.v4i2.2066

48. Liu XD, Wang W, Yang Y, Hou BH, Olasehinde TS, Feng N, et al. Nesting the SIRV model with NAR, LSTM and statistical methods to fit and predict COVID-19 epidemic trend in Africa. *BMC Public Health*. (2023) 23:138. doi: 10.1186/s12889-023-14992-6

Glossary

AMAE	Average Mean Absolute Error
ARIMA	Autoregressive Integrated Moving Average
ARMSE	Average Root Mean Square Error
AR	Autoregressive
CNN	Convolutional Neural Network
DFT	Discrete Fourier Transform
DNN	Deep Neural Network
GFT	Graph Fourier Transform
GLU	Gated Linear Unit
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
HA	Historical Average
IDFT	Inverse Discrete Fourier Transform
IGFT	Inverse Discrete Fourier Transform
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SMA	Simple Moving Average
VAR	Vector Autoregressive Model
WMA	Weighted Moving Average

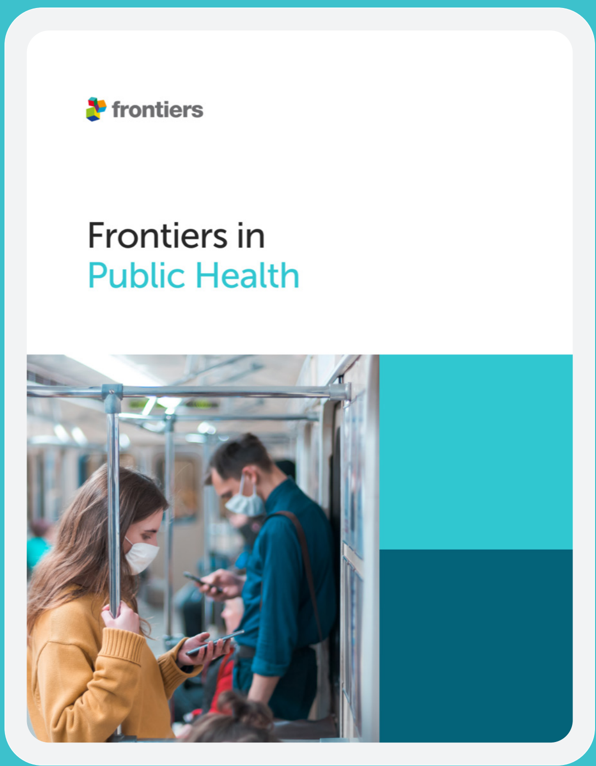
Frontiers in
Public Health

Explores and addresses today's fast-moving healthcare challenges

One of the most cited journals in its field, which promotes discussion around inter-sectoral public health challenges spanning health promotion to climate change, transportation, environmental change and even species diversity.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact