frontiers | Research Topics

Artificial intelligence in pathogenic microorganism research

Edited by
Chen Li, Yu-Dong Yao and Marcin Grzegorzek

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-6377-9
DOI 10.3389/978-2-8325-6377-9

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 in pathogenic microorganism research

Topic editors
Chen Li — Northeastern University, China
Yu-Dong Yao — Stevens Institute of Technology, United States
Marcin Grzegorzek — University of Lübeck, Germany

Citation
Li, C., Yao, Y.-D., Grzegorzek, M., eds. (2025). *Artificial intelligence in pathogenic microorganism research*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-6377-9

Table of contents

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 06 August 2024
DOI 10.3389/fmicb.2024.1449844

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Shasha Han,
Peking University, China
Wen Shen,
Second Affiliated Hospital of Nanchang University, China

*CORRESPONDENCE
Xin Zhang
✉ zhangxin800705@163.com

†These authors have contributed equally to this work

RECEIVED 16 June 2024
ACCEPTED 04 July 2024
PUBLISHED 06 August 2024

CITATION
Zhang X, Zhang D, Zhang X and Zhang X (2024) Artificial intelligence applications in the diagnosis and treatment of bacterial infections.
Front. Microbiol. 15:1449844.
doi: 10.3389/fmicb.2024.1449844

Artificial intelligence applications in the diagnosis and treatment of bacterial infections

Xiaoyu Zhang[1], Deng Zhang[2†], Xifan Zhang[1†] and Xin Zhang[1*]

[1]First Department of Infectious Diseases, The First Affiliated Hospital of China Medical University, Shenyang, China, [2]Department of Infectious Diseases, The First Affiliated Hospital of Xiamen University, Xiamen, China

The diagnosis and treatment of bacterial infections in the medical and public health field in the 21st century remain significantly challenging. Artificial Intelligence (AI) has emerged as a powerful new tool in diagnosing and treating bacterial infections. AI is rapidly revolutionizing epidemiological studies of infectious diseases, providing effective early warning, prevention, and control of outbreaks. Machine learning models provide a highly flexible way to simulate and predict the complex mechanisms of pathogen-host interactions, which is crucial for a comprehensive understanding of the nature of diseases. Machine learning-based pathogen identification technology and antimicrobial drug susceptibility testing break through the limitations of traditional methods, significantly shorten the time from sample collection to the determination of result, and greatly improve the speed and accuracy of laboratory testing. In addition, AI technology application in treating bacterial infections, particularly in the research and development of drugs and vaccines, and the application of innovative therapies such as bacteriophage, provides new strategies for improving therapy and curbing bacterial resistance. Although AI has a broad application prospect in diagnosing and treating bacterial infections, significant challenges remain in data quality and quantity, model interpretability, clinical integration, and patient privacy protection. To overcome these challenges and, realize widespread application in clinical practice, interdisciplinary cooperation, technology innovation, and policy support are essential components of the joint efforts required. In summary, with continuous advancements and in-depth application of AI technology, AI will enable doctors to more effectivelyaddress the challenge of bacterial infection, promoting the development of medical practice toward precision, efficiency, and personalization; optimizing the best nursing and treatment plans for patients; and providing strong support for public health safety.

KEYWORDS

bacterial infections, artificial intelligence, machine learning, diagnosis, treatment, epidemiologic surveillance

1 Introduction

Bacterial infections remain a major challenge in medical and public health in the 21st century, with millions of patient deaths annually. According to a study published in The Lancet on November 21, 2022, bacterial infections are one of the leading causes of global health loss and have become the second leading cause of death globally, after ischemic heart disease (GBD, 2019). Accurate and rapid identification of pathogens and their drug susceptibility profiles is essential for selecting the right treatment and reducing mortality. However, most current bacterial identification and drug susceptibility testing require culture times of several days, which not only delays the initiation of treatment, but also increases the risk of the development of resistant bacteria due to the long-term use of broad-spectrum antibiotics. At the same time, surveillance and management of bacterial infections are essential to prevent their spread and safeguard public health. In this context, the medical community urgently seeks new tools and strategies to better cope with bacterial infections. The rise of artificial intelligence (AI) technology, offers a new way to deal with bacterial infection (Mintz and Brodie, 2019; Larentzakis and Lygeros, 2021; Ting Sim et al., 2023).

Recently AI, as a powerful computational tool, has shown great potential in the diagnosis and treatment of bacterial infections (Goodswen et al., 2021; Jiang et al., 2022). AI is a science and technology that simulates human intelligence through computers, capable of mimicking human cognitive abilities and decision-making processes. In medicine, the main focus should be on the following terms: machine learning (particularly deep learning), natural language processing, computer vision, knowledge graph, and robotics, etc. (Mintz and Brodie, 2019) (Figure 1). The rapid expansion of AI technology spans from enhancing epidemiological surveillance to accelerating pathogen identification and predicting bacteria sensitivity to antimicrobial agents, Furthermore, AI supports the research and development of new drugs, vaccines, and innovative therapies, thereby promoting the development advancement of personalized medicine. tThe wide application of AI is expected to fundamentally transform the management, diagnosis, and treatment of bacterial infection (Wong et al., 2023).

Based on a comprehensive analysis of the existing literature and the latest research results, this study aimed to explore how AI technology can improve the efficiency and accuracy of medical diagnosis, as well as the level of personalized treatment, while focusing on the challenges that may hinder its practical clinical application. This will primarily provide medical workers with a comprehensive understanding of the application of AI technology in the diagnosis and treatment of bacterial infectious diseases, jointly promote the application of AI in the fight against bacterial infections, provide patients with more accurate and efficient medical services, and contribute to the development of global public health.

2 Application of AI in epidemiological surveillance of bacterial infectious diseases

AI and big data technologies are rapidly transforming the epidemiology of infectious diseases, particularly in the research and management of public health emergencies (PHEs). The modelsof infectious disease dynamics (IDD) and dynamic Bayesian networks (DBN)have not only promoted the spread of disease forecast accuracy, but also strengthened the ability analysis outbreakevolution (Gao and Wang, 2022). Through cloud computing platforms, AI can process massive data in real time and effectively monitor infectious disease outbreaks. Despite the challenge of long model training time, its practicability makes it an indispensable tool for early epidemic warning (Li et al., 2023). In addition, the development and application of geographic information systems (GIS), with its advanced data overlay capabilities, has greatly optimized the integration of public health data and has gained widespread acceptance (Wells et al., 2021). Similarly, the ToxPi*GIS Toolkit enables the visualization and analysis of geospatial data in the ArcGIS environment, a visualization framework that integrates multiple data sources and generates intuitive graphic files with through Python scripts, ArcGIS Pro methods, and custom toolkits (Fleming et al., 2022). In addition, the cloud data storage and use of Internet search data, such as Google Flu Trends, show the potential of disease surveillance systems based on large data to enhance real-time monitoring (Pfeiffer and Stevens, 2015).

Although these advanced tools and methods are currently used primarily in viral epidemiology, their potential for disease surveillance, data presentation and analysis, and public health decision-making continues to evolve. This suggests that their contribution to bacterial epidemiology is also expected to increase. For example, machine learning models can predict in advance the risk of *Clostridioides difficile* infection among patients in large hospitals, allowing healthcare teams to implement preventive measures proactively before infection occurs (Oh et al., 2018; Tilton and Johnson, 2019). Real-time locator systems can be used for contact tracing in the emergency department, which is not only more efficient and timely than tracing methods relying on electronic medical records, but also significantly increases the number of potentially exposed individuals identified while optimizing the use of time and resources (Hellmich et al., 2017). Maia Lesosky et al. revealed the impact of inter-hospital patient flow on methicillin-resistant *Staphylococcus aureus* (MRSA) transmission through Monte Carlo simulation (Lesosky et al., 2011). Further studies explored cross-hospital pathogen transmission using a susceptible infection model, demonstrating the important value of AI and big data in curbing hospital-acquired infections (Ciccolini et al., 2014).

AI is paving new ways to predict and prevent bacterial infections. AI technology integrates and analyzes vast amounts of complex data to achieve early recognition and accurate prediction of bacterial infection outbreaks. This optimizes prevention and control measures, guides public health decisions, and supports the global fight against infectious diseases and the new solution.

3 AI has revolutionized the study of bacterial infection mechanism

Further study of the pathogenesis of bacterial infectious diseases is crucial to fully understand the nature of these diseases. This process not only involves the complex process of how bacteria colonize, invade, and spread in the host but also involves the host's immune response and its interaction with pathogens. Among them, pathogen-host interaction is the key link, and animal models have been an indispensable tool in traditional research. They provide valuable data for observing the infection process of pathogens, host immune response, and disease development (Younes et al., 2020; Burkovski, 2022). While such approaches, although capable of providing accurate and rich biologic insights, are often costly, time-consuming, and associated with ethical concerns. With the rapid development of AI technology, especially the emergence of machine learning models, researchers can simulate and understand the complex interactions between pathogens and hosts without animal

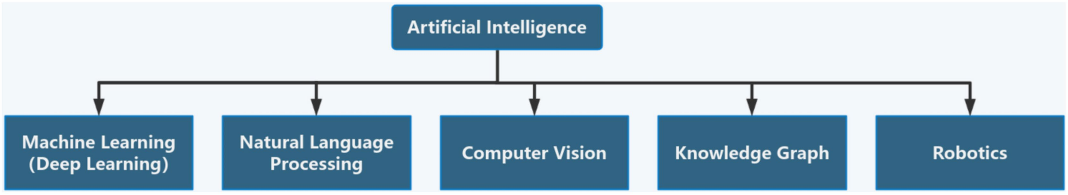


FIGURE 1
The relationship between machine learning (particularly deep learning), natural language processing, computer vision, knowledge graph, robotics, and artificial intelligence.

experiments. For example, the PHISTO tool promotes a deep understanding of infection mechanisms by synthesizing different databases and using text mining techniques, supplemented by graph theory analysis and BLAST search (Durmuş Tekir et al., 2013). A novel set of modular structural plasmids named pTBH (toolbox of Haemophilus) demonstrates coexistence and co-infection kinetics of fluorescently labeled strains by 3D microscopy combined with quantitative image analysis (Rapún-Araiz et al., 2023). Furthermore, AI models can effectively simulate the complex interactions between bacteria and hosts in different metabolic states (Dillard et al., 2023). Using advanced fluorescence microscopy detection and automated image analysis techniques, researchers have found that *Staphylococcus aureus* isolates from patients with bone/joint infection, bacteremia, and infective endocarditis show different infection characteristics in different host cell types (Rodrigues Lopes et al., 2022). These techniques not only provide a visual basis for understanding microbial behavior in specific host environments but also assist in the design of drugs and vaccines.

The application of machine learning models provides us with a highly flexible way to predict and simulate the complex mechanisms of pathogen-host interactions, which not only accelerates the research process but also reduces the research cost. Although AI models are not a complete replacement for all animal model studies, they provide new ways to explore uncharted territories.

4 AI application in the diagnosis of bacterial infections

In the traditional approach to diagnosing bacterial infectious diseases, laboratory technicians rely on microbiological and biochemical tests to identify pathogens. It includes bacterial culture, morphological observation, biochemical reaction tests, and serological techniques (Ernst et al., 2006; Váradi et al., 2017) (Table 1). In addition, molecular biology techniques are widely used for the identification of bacterial DNA sequences, of which the polymerase chain reaction (PCR) is a commonly used method (Wilson, 2015; Deusenbery et al., 2021). Although PCR technology is more advanced than traditional biochemical and microbiological methods, it requires a long time to complete the experimental process. Moreover, the integration and application of AI technology not only optimizes the traditional bacterial detection and management process, but also has the potential to bring about a complete revolution (Ho et al., 2019; Wang et al., 2020; Paquin et al., 2022; Howard et al., 2024) (Figure 2).

TABLE 1 Advantages and limitations of the traditional bacterial identification methods.

Method	Advantages	Limitations
Bacterial culture (Baron, 2019)	✓ The cost is low ✓ Effective for various bacteria ✓ Easy to operate	• Long time consumption • Some bacteria cannot develop • Susceptible to contamination • It is not suitable for highly specific tests
Morphological observation (Periasamy, 2014)	✓ No special equipment ✓ Intuitive is strong ✓ Accumulation of experience	• Subjectivity is strong • Limited information • The lack of specificity • Need to develop
Biochemical reaction tests (Ohkusu, 2000)	✓ Cost-effective ✓ Easy to operate	• Limited specificity • Does not apply to all bacteria
Serological technique (Eldin et al., 2019)	✓ High specificity ✓ Quick results ✓ Quantifiable analysis	• Greatly influenced by sampling time • There were false positive and false negative results • A variety of pathogens have cross-reacted

4.1 AI improves the efficiency and accuracy of pathogen identification

AI technology provides a new way to diagnose bacterial infections rapidly and accurately. For example, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) combined with ClinProTools software provided a method for the rapid identification of two *Staphylococcus aureus* subspecies, which achieved 100% identification and classification accuracy through genetic analysis and a fast classifier model (Pérez-Sancho et al., 2018). Findaureus, an open-source application based on Python, demonstrated the ability to automatically locate bacteria in the tissue section using immune fluorescent tags. It overcomes the challenges of the manual threshold-setting process and optimizes the analysis of the condition of complex tissue cell efficiency (Mandal et al., 2024). PhenoMatrix (PM) Colorimetric Detection Module (CDM) digital

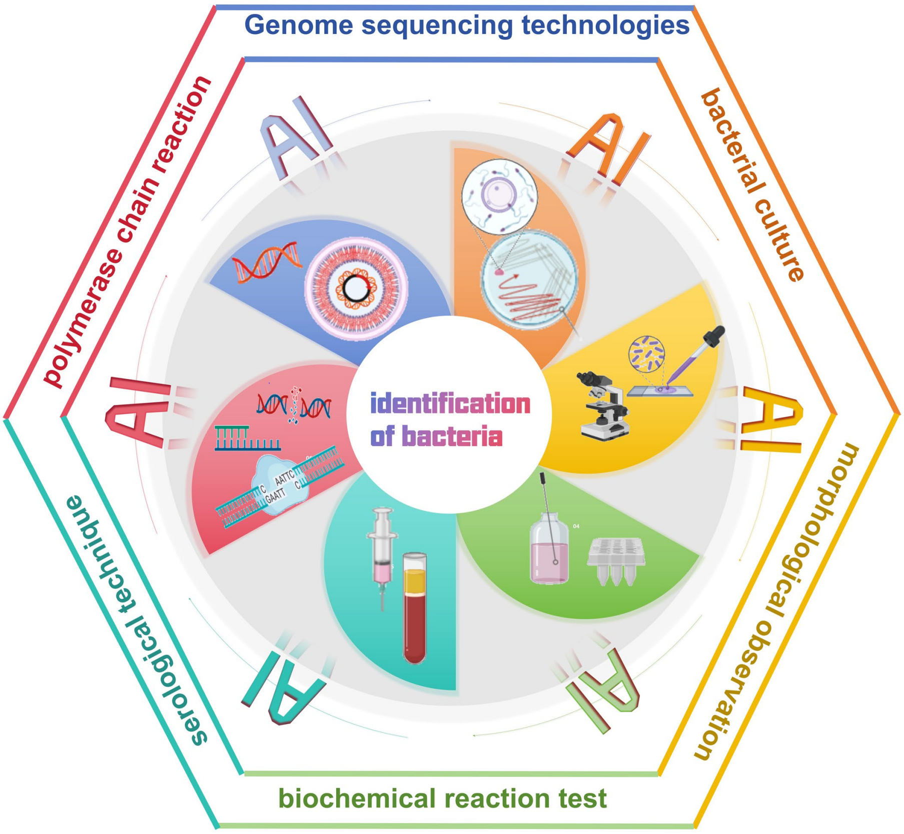


FIGURE 2
Artificial intelligence facilitates the diagnosis of bacterial infectious diseases.

imaging software uses the automated Walk Away Specimen Processor to detect Group B Streptococcus (with high sensitivity similar to that of molecular testing methods, increasing laboratory productivity and reducing the potential for human error (Baker et al., 2020). In addition, DNA microarray technology, using a machine learning decision-making algorithm (DendrisChips), identifies 11 types of bacteria associated with respiratory tract infections within 4 h. This technology combines PCR amplification of bacterial 16S rDNA and specific oligonucleotide hybridization on DendrisChips®, which are read with a laser scanner, thereby achieving quick and accurate detection and differentiation with over 95% accuracy (Senescau et al., 2018). Using neural networks to analyze response patterns, a researcher has designed a sensor capable of identifying 16 different bacterial species and their Gram-staining properties with >90% accuracy. The sensor is stable for up to 6 months after preparation and requires one-thirtieth the amount of dye and sample as traditional solution-based sensors, compared to conventional techniques (Laliwala et al., 2022). Thus, this method provides an innovative diagnostic tool that promises clinical applications in resource-limited settings.

In diagnosing diseases that pose a serious threat to human health, such as tuberculosis, although conventional microscopy methods are effective, they are slow and of limited sensitivity. The introduction of AI, specifically Metasystems' automated antifungal bacilli (AFB) smear microscopy scanning and deep learning-based image analysis module (Neon Metafer), has greatly improved the speed and accuracy of antifungal bacilli (AFB) smear-negative slide recognition speed and accuracy (Desruisseaux et al., 2024). A deep neural network (DNN) classifier combined with an automated slide scanning system reduces analysis time from several minutes to approximately 10 s per slide (Horvath et al., 2020). Further, a novel diagnostic system combining T-SPOT with DL-based computed tomography image analysis can significantly improve the classification accuracy of nontuberculous mycobacterial lung disease and pulmonary tuberculosis (Ying et al., 2022). AI tools, such as artificial neural networks, are becoming important in providing rapid and effective pathogen detection methods (Dande and Samant, 2018). AI technology brings unprecedented accuracy and speed to pathogen detection through efficient learning and analysis capabilities. It will not only promote the automation of pathogen detection but also substantially decrease error rates caused by human operation, thereby improving the reliability of the diagnostic process.

4.2 AI optimizes antimicrobial susceptibility testing

Identifying pathogens and performing Antimicrobial Susceptibility Testing (AST) in today's clinical laboratories often relies on culturing and isolating pathogens. Standard AST methods (CLSI, 2023 such as disk diffusion, microbroth dilution, and AGAR dilution methods, typically require 2–3 days or longer from sample collection to obtaining culture and drug susceptibility results (Abu-Aqil et al., 2023). To effectively control infections and prevent them from rapidly deteriorating or spreading to other parts of the body, clinicians often choose broad-spectrum antimicrobials for empirical treatment, given that many infectious diseases are often difficult to diagnose by symptoms in the early stages. However, this practice may increase the risk of drug-resistant strains arising due to inappropriate drug selection; therefore, there is an urgent need for rapid and accurate AST technologies to guide diagnosis and treatment.

With the rapid advancement of technology, AI has become an important tool in bacterial AST, providing various efficient and rapid methods to perform drug susceptibility testing. For example, Raman spectroscopy based on image stitching technology enables single-cell level detection, which can automatically, efficiently, and rapidly identify drug-resistant bacteria (Nakar et al., 2022; Dou et al., 2023). Combining machine learning and infrared spectroscopy enables rapid and definitive identification of urinary tract infection bacteria and their drug resistance, dramatically reducing the time from sample collection to results. This approach decreases the time of identification and sensitization of *Escherichia coli*, *Proteus mirabilis*, and *Pseudomonas aeruginosa* from 48 h to approximately 40 min (Ciccolini et al., 2014; Tilton and Johnson, 2019; Younes et al., 2020; Burkovski, 2022). Similarly, the SlipChip microfluidic device uses electrophoresis technology to extract and enrich bacteria directly from positive blood cultures. This device enables parallel inoculation of bacteria into nanoscale droplets of broth, facilitating simultaneous multiple AST. Results can be reported to clinicians within 3–8 h, ensuring reliable AST results and enabling earlier reporting and targeted antimicrobial treatment (Yi et al., 2019).

Automation technology has also demonstrated high efficiency in detecting certain special drug-resistant bacteria, such as MALDI-TOF MS, for the detection of MRSA and carbapenem-resistant *Klebsiella pneumoniae* (CRKP) (Wieser et al., 2012; Zhang et al., 2023). However, the novel ML-based MALDI-TOF MS method enables rapid identification of MRSA and CRKP from labeled blood cultures within 1 h (Yu et al., 2023a,b). Recent studies have shown that using computer science to analyze a large number of MALDI-TOF MS data can provide a comprehensive understanding of western blot mapping between resistant and sensitive isolates (Wang et al., 2021). WASPLab automation system can significantly shorten the vancomycin resistant enterococcus (VRE) recognition time (Cherkaoui et al., 2019). In addition, the automated plate evaluation system (APAS Independence) has significantly improved the productivity of high-throughput laboratories through its highly sensitive digital image analysis technology to accurately classify MRSA and sensitive *Staphylococcus aureus* (MSSA) cultures as negative or positive without human intervention (Gammel et al., 2021).

In conclusion, the application of AI technologies to antimicrobial susceptibility testing enables the rapid and accurate identification of drug-resistant bacteria, thereby dramatically shortening the time from sample collection to result confirmation, and can be accomplished without human intervention. These technologies provide laboratories with a rapid and automated means of drug resistance monitoring, which significantly improves diagnostic efficiency and helps clinicians make rational antimicrobial treatment decisions as early as possible (Table 2).

TABLE 2 Artificial intelligence in the bacteria identification and drug sensitivity analysis.

Technology	Application	References
MALDI-TOF MS + ClinProTools software	Rapidly identified *Staphylococcus aureus* subspecies	Pérez-Sancho et al. (2018)
Findaureus	Automatic localization of bacteria in immunofluorescently labeled tissue sections	Mandal et al. (2024)
PM + CDM + WASP	High sensitivity to identify group B streptococcus	Baker et al. (2020)
Machine learning-based DNA micro-matrix technology	More than 95% accuracy in identifying respiratory bacteria	Senescau et al. (2018)
Neural network-based sensors	90% accuracy in bacterial identification	Laliwala et al. (2022)
AFB + Neon Metafer	Significantly improved the speed and accuracy of identification of acid-fighting bacilli (AFB) on smear-negative slides	Desruisseaux et al. (2024)
DNN + an automated slide scanning system	Significantly reduced slide analysis time	Horvath et al. (2020)
T-SPOT + DL-based technology	Significantly improved the classification accuracy of NTM—PD and PTB	Ying et al. (2022)
Raman spectroscopy based on image stitching technology	Automatically, efficiently and rapidly identified drug-resistant bacteria	Dou et al. (2023) and Nakar et al. (2022)
SlipChip microfluidic device	Significant reduction in bacterial drug sensitivity test time	Yi et al. (2019)
A novel MALDI-TOF MS method based on ML	Rapidly identified MRSA and CRKP	Yu et al. (2023a,b)
WASPLab automation system	Significantly shorten the vancomycin-resistant enterococcus (VRE) recognition time	Cherkaoui et al. (2019)
APAS Independence	Accurately distinguish MRSA and MSSA	Gammel et al. (2021)

4.3 AI can improve bacterial genome sequencing

Genome sequencing technologies (including whole genome sequencing and next-generation sequencing) have significantly accelerated not only the identification of infectious agents, but also the tracking of transmission pathways in healthcare settings and the analysis of the impact of complex microbial communities on human health (d'Humières et al., 2021; Deusenbery et al., 2021). It also provides a powerful tool for monitoring and responding to antimicrobial resistance (AMR) globally (Waddington et al., 2022; Sherry et al., 2023).

Current genetic testing techniques mainly match based on sequence similarity; however, these tools are often unsuccessful in identifying new species without closely related genomes or related sequences in reference databases. In response to this challenge, the machine learning-based PaPrBaG method provides a reliable and consistent prediction method that maintains its reliability even with low genome coverage (Deneke et al., 2017). In addition, machine learning combined with metagenomic sequencing can significantly improve the diagnostic accuracy of diseases that are difficult to diagnose, such as tuberculous meningitis (Ramachandran et al., 2022).

Another challenge for genetic testing technologies is how to rapidly and accurately interpret high-dimensional genomic data as the cost of second-generation sequencing technology decreases and throughput increases. Machine learning techniques have shown their potential in processing large genomic data by analyzing and predicting the health impact of Shiga toxin-producing *Escherichia coli* infections, providing new methods and perspectives for microbial risk assessment (Njage et al., 2019). In addition, Bayesian neural networks using a nonparametric Bayesian algorithm excelled in accelerating the analysis of genetic association studies and efficiently and accurately identifying variant strains of infection (Beam et al., 2014).

Combining machine-learning models with genomics technology has shown excellent performance in predicting pathogen resistance, which is significantly better than existing methods. Some researchers have used machine learning to construct a knowledge map of antimicrobial resistance in *Escherichia coli*, which realizes the automatic discovery of knowledge of antimicrobial resistance in *Escherichia coli* and reveals unknown drug resistance genes (Youn et al., 2022). Based on the XGBoost and convolutional neural network approaches, the researchers not only accurately predicted the minimum inhibitory concentrations of *Klebsiella pneumoniae* clinical isolates against 20 antimicrobial drugs, but also successfully identified strains with high drug resistance or high virulence (Nguyen et al., 2018; Liu et al., 2021; Lu et al., 2022). Similarly, some researchers have innovated a decision tree method called Treesist-TB for identifying mutant strains and predicting drug resistance, which has a recognition ability beyond the existing TB-Profiler tools (Deelder et al., 2022), This technique not only demonstrates the value of decision trees in the tuberculosis field but also provides a reference template to identify other drug-resistant pathogens.

AI has shown great potential in genome sequencing technology. In response to the challenges of identifying new species and interpreting high-dimensional data, machine learning has surpassed the limitations of traditional genetic detection techniques and deepened our understanding of the microscopic world of pathogens. Furthermore, machine learning excels in predicting antimicrobial drug resistance, outperforming traditional methods, and strengthening global antibiotic resistance (AMR) monitoring efforts.

5 Application of AI in the treatment of bacterial infections

The challenges in the treatment of bacterial infections are diverse, and one of the most serious is the increasing resistance to antimicrobial agents. The importance of Antimicrobialresistance was formally declared at the United Nations General Assembly High-level Meeting on antimicrobial Resistance in 2016, and countries were called on to commit to developing their national action plans on antimicrobial resistance. Nearly 5 million people died globally due to resistant pathogens in 2019 (Antimicrobial Resistance Collaborators, 2022). Current projections suggest that by 2050, 10 million people globally could be burdened by antimicrobial drug resistance each year (Walsh et al., 2023). Over time, bacteria have acquired resistance to antimicrobial drugs through natural selection and genetic variation, thereby undermining the effectiveness of traditional treatments. In addition, the high diversity of bacteria and the complexity of bacterial-host interactions further increase the difficulty of treatment, making the development of vaccines and novel drugs difficult. Hence, developing new antimicrobial strategies and therapeutic approaches are urgently needed to address these issues (Stracy et al., 2022).

In this context, AI technology accurately simulates the complex interactions between pathogen, host, and drugs, revealing microbial infection features and optimizing drug and vaccine design (Figure 3). In addition, AI application in the field of phage therapy brings new hope for the fight against bacterial resistance.

5.1 AI revolutionizes drug discovery and development

In drug research and development, the application of AI is breaking the boundaries of traditional research, providing new strategies to overcome the problem of drug resistance. For example, by combining high-throughput biophysical analysis and machine learning, a framework was established to identify and predict bioactive targets of antimicrobial drugs, which successfully revealed the relationship between phenotype, target, and chemotype, providing an effective way to identify candidate therapeutic drugs (Santa Maria et al., 2017). Meanwhile, combining fragment-based drug design with quantitative structure–activity relationship modeling demonstrates the potential of artificial neural networks in the drug discovery process (Kleandrova and Speck-Planche, 2020). Using data-driven techniques, the study of bacterial minimal inhibitory concentration data using machine learning and matched molecular pair analysis has revealed key chemical features that affect bacterial biological activity, thus promising to expand the chemical space of broad-spectrum antimicrobial agents (Gurvic et al., 2022). In a study, a support vector machine learning approach was applied to analyze genomics, metabolomics, and transcriptomics data of *Pseudomonas aeruginosa*. This approach successfully identified a key molecular mechanism that distinguishes between pathogenic and non-pathogenic strains of *Pseudomonas aeruginosa*, which not only provides high-value targets for the development of novel antimicrobial therapeutics but also highlights the importance of dynamically integrating multidimensional data in modern drug discovery and development (Larsen et al., 2014).

Furthermore, significant breakthroughs have been made in the application of AI in specific disease areas, such as anti-tuberculosis drug development. The machine learning and artificial neural network method can be used to successfully find LeuRS for *Mycobacterium tuberculosis* and MetRS double targets of inhibitors (Volynets et al., 2022), and small-molecule inhibitors of the enzymes required for *M. tuberculosis* topoisomerase I have been successfully identified

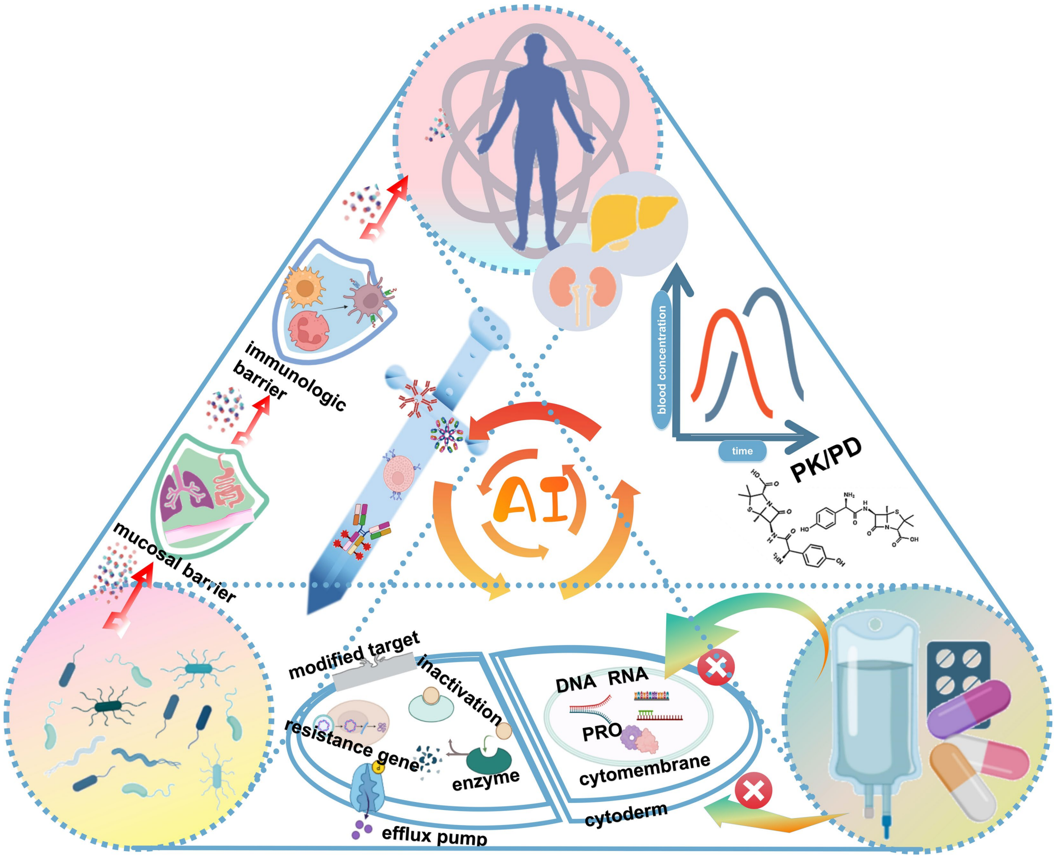


FIGURE 3
AI technology can model complex interactions between pathogens, hosts, and drugs.

(Ekins et al., 2017), providing a new strategy to overcome multidrug resistance in tuberculosis. In addition, combining public Mtb data with machine learning not only greatly improves the efficiency of drug discovery, but also accumulates valuable data resources for future anti-tuberculosis research and new drug development (Lane et al., 2022).

These advanced technologies not only accelerate the research and development process of new drugs, but also enhance the possibility of discovering potential therapeutic options, fundamentally changing how researchers understand and operate complex biological systems, and heralding a new era of smarter and more precise development in the pharmaceutical field.

5.2 AI brings breakthroughs in vaccine development

Currently rapid progress has been made in vaccine research and development against viral diseases. In particular, the speed and efficiency of response to emerging virus epidemics have been greatly improved, such as the application of computer-aided design of COVID-19 vaccine candidates in the global pandemic of COVID-19 in early 2020 (Abbasi et al., 2022). In contrast, bacteria in the field of vaccine research and development are faced with more complicated challenges. The high variability of bacteria, rapidly evolving drug resistance, and complexity of interactions between bacteria and their hosts all challenge the development of effective vaccines against bacterial infectious diseases. To address these challenges, leveraging emerging tools such as artificial intelligence, computer-aided design, and advanced immunological evaluation techniques has become pivotal to accelerating the development of safe and effective vaccines.

In the process of vaccine design, scientists are challenged not only to identify the key antigens that can trigger lasting immune memory, but also to ensure that the vaccine can elicit broad protective immune responses, including humoral and cellular immune responses, to achieve effective protection in the long term. Recently, reverse vaccinology (RV) technology has been widely used in vaccine research and development. As a calculation method, RV is mainly applied to bacterial pathogens. Bexsero, a *Neisseria meningitidis* B vaccine designed by RV, has been registered and widely used in many countries (Heinson et al., 2015). In addition, a key component of vaccine development—antigen identification—is strongly supported by computational tools such as deep learning, reverse vaccinology and immunoinformatics. In-depth analysis of vaccine targets derived from pathogen protein-coding genomes has led to the successful development of a multi-epitope subunit vaccine with potentially potent protection. Although the safety and immunogenicity of the vaccine need to be further verified (Rawal et al., 2021), this approach not only accelerates the vaccine design process and reduces the reliance on traditional trial methods, but also has important implications for addressing the threat of drug-resistant bacteria. Research shows that a new type of machine learning model, compared with traditional methods, achieves higher precision and sensitivity in predicting aspects of *mycobacterium tuberculosis* (Khanna and Rana, 2019).

The application of machine learning technology not only optimizes the vaccine development process and improves efficiency by reducing the reliance on traditional experiments and animal testing, but also provides strong scientific and technological support to cope with evolving epidemics of bacterial infections.

5.3 AI drives innovative applications of phage therapy

Phage therapy has attracted much attention from the scientific community for its potential advantages in combating drug-resistant bacterial infections (Viertel et al., 2014; Kulshrestha et al., 2024). However, accurate prediction of the complex interactions between phages and their target pathogens and hosts remains challenging (Cisek et al., 2017), and AI models become an important tool to overcome this challenge. For example, a machine learning-based local K-mer strategy is used to accurately predict phage-bacteria interactions (Qiu et al., 2024). Simultaneously, machine learning can assist in the design of clinical phage therapy, particularly for urinary tract infections caused by multidrug-resistant *E. coli* (Keith et al., 2024). In addition, a tool called HostPhinder predicted phage host genus and species with 81 and 74% accuracy, respectively, demonstrating the technology's ability to pinpoint therapeutic targets (Villarroel et al., 2016).

Consequently, the application of phages, either alone or in combination with antimicrobial agents, can be a viable alternative to treat infections with resistant pathogens (Tagliaferri et al., 2019). The rapid development of AI technology enhances the potential of phage therapy by accurately predicting complex interactions between pathogens and phages, thereby contributing to the design of personalized treatment. This not only accelerates the development of phage therapy but also enhances its treatment success.

5.4 AI-assisted clinical decision support systems

The timing of effective antimicrobials is a key determinant of morbidity and mortality in the management of infectious diseases, specifically in the case of septic shock (Evans et al., 2021). Early identification can not only reduce the poor prognosis caused by delayed treatment, but also help avoid unnecessary medical intervention and reduce treatment costs, thus significantly improving the survival rate and quality of life of patients.

Under the background of increasing emphasis on individualized treatment and precision medicine, AI progress not only promotes medical innovation, but also may overturn the existing diagnosis and treatment mode. In bacterial infectious disease diagnosis and treatment, AI and ML are used to simplify the clinicians' work process, improve the quality of decision-making, and promote the development of personalized treatment options (Langford et al., 2024). For example, ML models have been successfully applied to diagnose respiratory syncytial virus infection and pertussis in children by combining clinical symptoms with laboratory test results (Mc Cord-De Iaco et al., 2023). Based on statistically significant clinical indicators such as sex and age, LightGBM and other ML models have a good effect on predicting the etiology of classical Fever of Unknown Origin in patients (Yan et al., 2021). In addition, ML models can rapidly predict the risk of MRSA infection in patients with community-acquired pneumonia and facilitate the implementation of targeted antimicrobial treatment (Rhodes et al., 2023). Clinical decision trees generated based on recursive methods are valuable for determining the likelihood of infection with extended-spectrum beta-lactamase strains in patients with bacteremia (Goodman et al., 2016). A system for early warning of antimicrobial drug allergies, K-CDSTM, effectively warns of antimicrobial drug allergies and prevents patients from being prescribed antimicrobial drugs that may trigger allergic reactions (Han et al., 2024). The ontology-driven clinical decision support system uses big data to assist the treatment decision-making of infectious diseases and constructs a bridge between patients and medical workers (Shen et al., 2018). In the development of predictive disease models, tools such as multiple infectious disease diagnostic models are significantly more accurate than traditional prediction techniques based on large amounts of training data (Wang et al., 2022). In a 3-month case–control study using a computerized clinical decision support system in an experimental group, time was reduced by approximately 1 h and antimicrobial costs were saved by approximately US $84,000 (McGregor et al., 2006).

In summary, machine learning models have been successfully used to improve diagnostic accuracy and predict disease risk in clinical decision-making, showing better accuracy and efficiency than traditional approaches (Figure 4). AI and ML technologies are leading the wave of medical innovation and have the potential to change the traditional methods of diagnosis and treatment.

6 AI helps personalized medical development

Through deep study and the analysis of the complex algorithm, AI can process and interpret patients with huge amounts of data, including genetic information, living habits and historical health records, etc. This not only enables accurate diagnosis of the disease, but also facilitates personalized treatment plans for each patient. For example, in cancer treatment, AI can help doctors choose the most appropriate combination of drugs for patients, reduce side effects, and improve cure rates. Similarly, AI can also predict efficacy and possible complications and provide tailored health management plans for patients (Bilgin et al., 2024; Elemento, 2024).

In the field of bacterial infections, a novel method called CombiANT can rapidly quantify antimicrobial synergy through a single test and automated image analysis, enabling personalized clinical synergy testing to improve the anti-infection combination therapy (Fatsis-Kavalopoulos et al., 2020). Kuo-Wei Hsu et al. developed an automated portable antimicrobial susceptibility testing system for four common urinary tract infection bacterial strains, taking only 4.5–9 h to complete the test, which holds

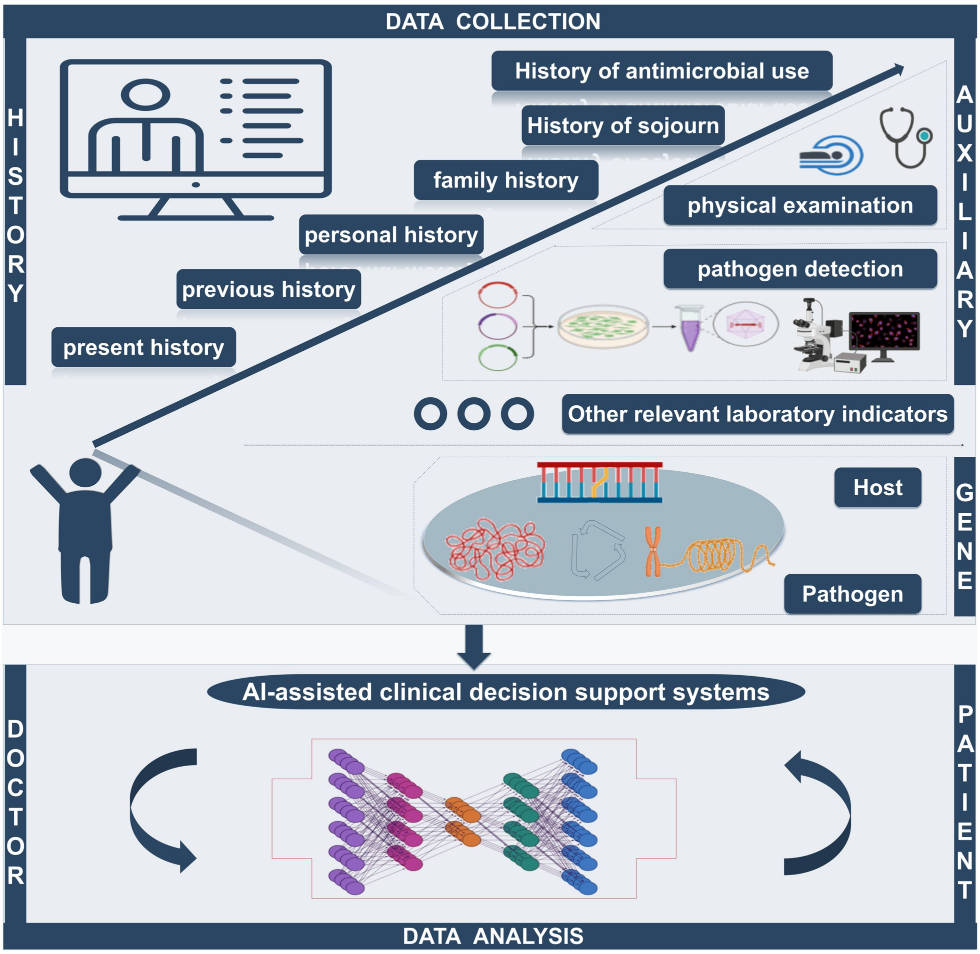


FIGURE 4
The AI-assisted clinical decision support system can quickly collect the patient's history of present disease, past history, personal history, family history, travel history, and antibiotic use history. Simultaneously, the system can integrate relevant auxiliary examination (including imaging examination and laboratory examination) and analysis of the genetic information of hosts and pathogens to provide the best treatment, becoming a bridge of effective communication between doctors and patients.

promise for future application in personalized medicine practice (Hsu et al., 2021). Connor Rees et al. showed an overall success rate of > 90% for correct diagnoses in the list of 10 differential diagnoses generated by ChatGP-3 (Hirosawa et al., 2023). In the future, more research is expected to focus on evaluating more complex cases and promote the development of fully trained artificial intelligence chatbots to improve the accuracy and completeness of diagnosis and further personalize patient treatment.

7 Challenges of AI in the medical field

Although the application of AI in the field of bacterial infections has great potential and prospects, it also faces numerous challenges. The first is the problem of data quantity and data quality. The collection, sorting and sharing of case data related to bacterial infectious diseases are restricted by privacy protection and standardization, which limits the training efficiency and application scope of AI models (Cath, 2018; Baowaly et al., 2019; Hummel and Braun, 2020). Second, deep-learning algorithms often lack the ability to provide a convincing explanation for their predictions—the so-called "black box" problem—which can affect prediction accuracy and public trust in AI systems (Schwartz et al., 2024). In addition, most healthcare AI research to date has been done in non-clinical Settings, with few instances of successful integration of AI into clinical care and most cases are still in the experimental stage (Alami et al., 2020). Therefore, generalizing the results of the study may be challenging. Moreover, complex and variable bacterial infection mechanisms and rapid mutation of bacterial genes make it more difficult to accurately predict pathogen behavior and drug sensitivity. Furthermore, the establishment of AI models requires interdisciplinary fields, including microbiology, biochemistry, genetics, mathematics and computer science, etc. (Figure 5), This requires a high level of knowledge and skills from the researchers and developers, posing a significant challenge for research teams with limited resources.

Currently in the field of artificial intelligence, a perfect legal system and authoritative standards have not been established. With the continuous progress of technology and the expansion of

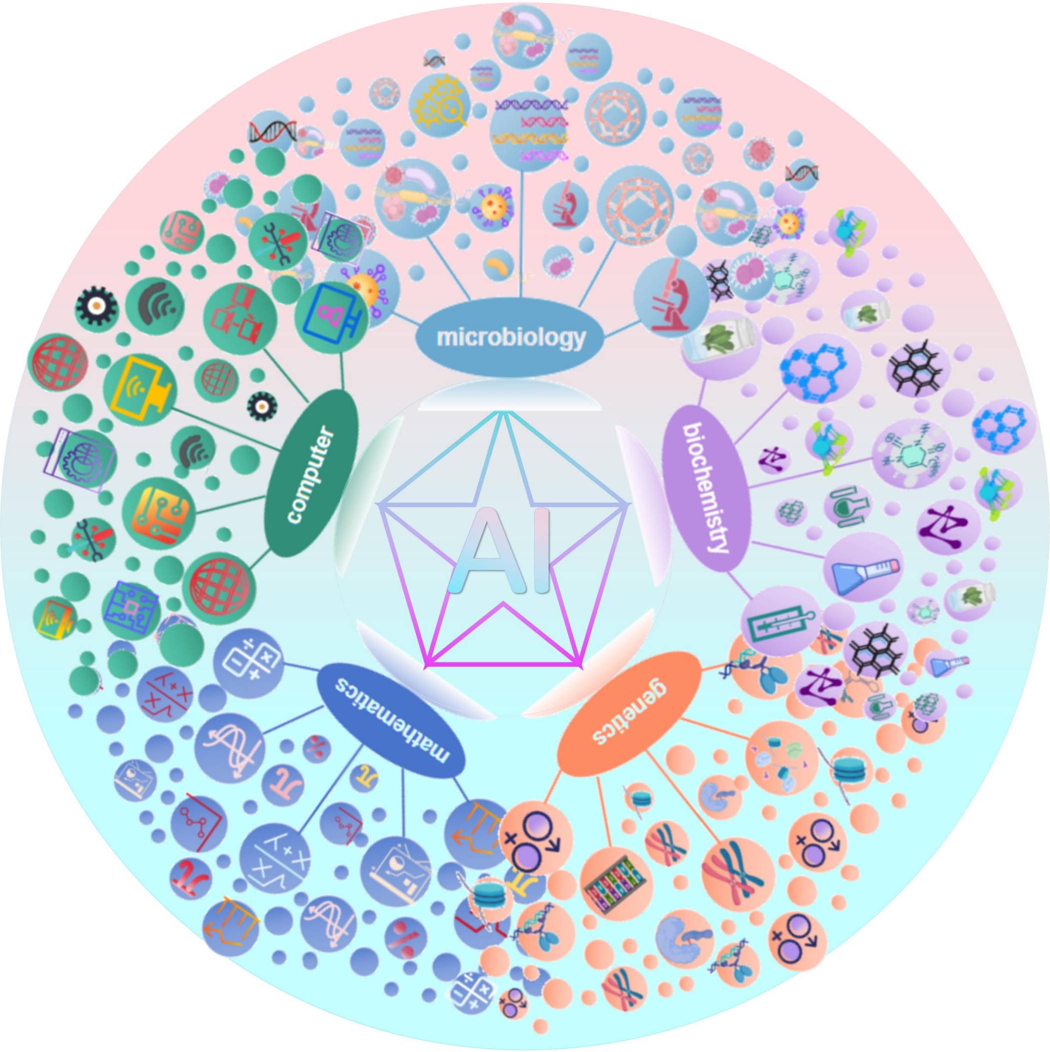


FIGURE 5
The successful application of AI models in medicine relies on multidisciplinary collaboration.

application fields, formulating and updating relevant regulations is essential,which will be a dynamic development process (Rees and Müller, 2022).

8 Conclusion

The rise of AI technology has opened up a new way to deal with bacterial infections. With the help of advanced technologies such as machine learning and deep learning, AI has been applied in many key areas, from rapid pathogen detection and antimicrobial susceptibility analysis to the interpretation of complex genomic data and the development of personalized treatment options. Through highly optimized algorithms, AI technology not only greatly improves the speed and accuracy of pathogen identification, but also accurately predicts the susceptibility of pathogens to specific antibiotics based on historical data, thus providing strong scientific decision support for doctors. Similarly, in the field of epidemiological surveillance, AI technology has strengthened the real-time monitoring and early warning ability of the spread of bacterial infectious diseases by analyzing and processing a large amount of epidemiological dataand providing a powerful analytical tool and basis for public health decision-making.

Although AI has a broad application prospect in the treatment of bacterial infectious diseases, there remain important issues to be solved, such as how to ensure the transparency and interpretability of AI decision-making and how to accelerate the diagnosis and treatment while strictly controlling the ethics and patient safety. To overcome these challenges and achieve its wide application in clinical practice, interdisciplinary cooperation, technological innovation and policy support are needed.

Prospectively, AI technology will bring a profound transformation in the field of diagnosis and treatment of bacterial infections. With the continuous strengthening and maturity of AI in pathogen identification, drug susceptibility testing and genomic analysis, it will become the right hand of clinicians. With the assistance of AI, medical workers will can better cope with the challenges brought by bacterial infections, continue to promote the development of medical practice in the direction of more precision, efficiency, and personalization, and ultimately achieve the goal of providing optimal care and treatment for patients.

Author contributions

XYZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. DZ: Data curation, Formal analysis, Software, Supervision, Validation, Visualization, Writing – review & editing. XFZ: Data curation, Investigation, Software, Supervision, Visualization, Writing – review & editing. XZ: Conceptualization, Formal analysis, Project administration, Resources, Software, Supervision, Visualization, 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 would like to thank Editage (www.editage.cn) for English language editing. The images in this article are drawn with the help of https://app.biorender.com/ and WPS office.

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

Abbasi, B. A., Saraf, D., Sharma, T., Sinha, R., Singh, S., Sood, S., et al. (2022). Identification of vaccine targets & design of vaccine against SARS-CoV-2 coronavirus using computational and deep learning-based approaches. *PeerJ* 10:e13380. doi: 10.7717/peerj.13380

Abu-Aqil, G., Lapidot, I., Salman, A., and Huleihel, M. (2023). Quick detection of Proteus and Pseudomonas in patients' urine and assessing their antibiotic susceptibility using infrared spectroscopy and machine learning. *Sensors (Basel)* 23:8132. doi: 10.3390/s23198132

Alami, H., Lehoux, P., Denis, J.-L., Motulsky, A., Petitgand, C., Savoldelli, M., et al. (2020). Organizational readiness for artificial intelligence in health care: insights for decision-making and practice. *J. Health Organ. Manag.* 35, 106–114. doi: 10.1108/JHOM-03-2020-0074

Antimicrobial Resistance Collaborators (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* 399, 629–655. doi: 10.1016/S0140-6736(21)02724-0

Baker, J., Timm, K., Faron, M., Ledeboer, N., and Culbreath, K. (2020). Digital image analysis for the detection of group B Streptococcus from ChromID Strepto B medium using PhenoMatrix algorithms. *J. Clin. Microbiol.* 59, e01902–e01919. doi: 10.1128/JCM.01902-19

Baowaly, M. K., Lin, C.-C., Liu, C.-L., and Chen, K.-T. (2019). Synthesizing electronic health records using improved generative adversarial networks. *J. Am. Med. Inform. Assoc.* 26, 228–241. doi: 10.1093/jamia/ocy142

Baron, E. J. (2019). Clinical microbiology in Underresourced settings. *Clin. Lab. Med.* 39, 359–369. doi: 10.1016/j.cll.2019.05.001

Beam, A. L., Motsinger-Reif, A., and Doyle, J. (2014). Bayesian neural networks for detecting epistasis in genetic association studies. *BMC Bioinformatics* 15:368. doi: 10.1186/s12859-014-0368-0

Bilgin, G. B., Bilgin, C., Burkett, B. J., Orme, J. J., Childs, D. S., Thorpe, M. P., et al. (2024). Theranostics and artificial intelligence: new frontiers in personalized medicine. *Theranostics* 14, 2367–2378. doi: 10.7150/thno.94788

Burkovski, A. (2022). Host–pathogen interaction 3.0. *Int. J. Mol. Sci.* 23:12811. doi: 10.3390/ijms232112811

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philos. Trans. A Math. Phys. Eng. Sci.* 376:20180080. doi: 10.1098/rsta.2018.0080

Cherkaoui, A., Renzi, G., Charretier, Y., Blanc, D. S., Vuilleumier, N., and Schrenzel, J. (2019). Automated incubation and digital image analysis of chromogenic media using Copan WASPLab enables rapid detection of vancomycin-resistant Enterococcus. *Front. Cell. Infect. Microbiol.* 9:379. doi: 10.3389/fcimb.2019.00379

Ciccolini, M., Donker, T., Grundmann, H., Bonten, M. J. M., and Woolhouse, M. E. J. (2014). Efficient surveillance for healthcare-associated infections spreading between hospitals. *Proc. Natl. Acad. Sci. USA* 111, 2271–2276. doi: 10.1073/pnas.1308062111

Cisek, A. A., Dąbrowska, I., Gregorczyk, K. P., and Wyżewski, Z. (2017). Phage therapy in bacterial infections treatment: one hundred years after the discovery of bacteriophages. *Curr. Microbiol.* 74, 277–283. doi: 10.1007/s00284-016-1166-x

d'Humières, C., Salmona, M., Dellière, S., Leo, S., Rodriguez, C., Angebault, C., et al. (2021). The potential role of clinical metagenomics in infectious diseases: therapeutic perspectives. *Drugs* 81, 1453–1466. doi: 10.1007/s40265-021-01572-4

Dande, P., and Samant, P. (2018). Acquaintance to artificial neural networks and use of artificial intelligence as a diagnostic tool for tuberculosis: a review. *Tuberculosis (Edinb.)* 108, 1–9. doi: 10.1016/j.tube.2017.09.006

Deelder, W., Napier, G., Campino, S., Palla, L., Phelan, J., and Clark, T. G. (2022). A modified decision tree approach to improve the prediction and mutation discovery for drug resistance in *Mycobacterium tuberculosis*. *BMC Genomics* 23:46. doi: 10.1186/s12864-022-08291-4

Deneke, C., Rentzsch, R., and Renard, B. Y. (2017). PaPrBaG: a machine learning approach for the detection of novel pathogens from NGS data. *Sci. Rep.* 7:39194. doi: 10.1038/srep39194

Desruisseaux, C., Broderick, C., Lavergne, V., Sy, K., Garcia, D.-J., Barot, G., et al. (2024). Retrospective validation of MetaSystems' deep-learning-based digital microscopy platform with assistance compared to manual fluorescence microscopy for detection of mycobacteria. *J. Clin. Microbiol.* 62:e0106923. doi: 10.1128/jcm.01069-23

Deusenbery, C., Wang, Y., and Shukla, A. (2021). Recent innovations in bacterial infection detection and treatment. *ACS Infect. Dis.* 7, 695–720. doi: 10.1021/acsinfecdis.0c00890

Dillard, L. R., Glass, E. M., Lewis, A. L., Thomas-White, K., and Papin, J. A. (2023). Metabolic network models of the Gardnerella Pangenome identify key interactions with the vaginal environment. *mSystems* 8:e0068922. doi: 10.1128/msystems.00689-22

Dou, X., Yang, F., Wang, N., Xue, Y., Hu, H., and Li, B. (2023). Rapid detection and analysis of Raman spectra of Bacteria in multiple fields of view based on image stitching technique. *FBL* 28:249. doi: 10.31083/j.fbl2810249

Durmuş Tekir, S., Çakır, T., Ardıç, E., Sayılırbaş, A. S., Konuk, G., Konuk, M., et al. (2013). PHISTO: pathogen–host interaction search tool. *Bioinformatics* 29, 1357–1358. doi: 10.1093/bioinformatics/btt137

Ekins, S., Godbole, A. A., Kéri, G., Orfi, L., Pato, J., Bhat, R. S., et al. (2017). Machine learning and docking models for *Mycobacterium tuberculosis* topoisomerase I. *Tuberculosis (Edinb.)* 103, 52–60. doi: 10.1016/j.tube.2017.01.005

Eldin, C., Parola, P., and Raoult, D. (2019). Limitations of diagnostic tests for bacterial infections. *Med. Mal. Infect.* 49, 98–101. doi: 10.1016/j.medmal.2018.12.004

Elemento, O. (2024). How artificial intelligence unravels the complex web of Cancer drug response. *Cancer Res.* 84, 1745–1746. doi: 10.1158/0008-5472.CAN-24-1123

Ernst, D., Bolton, G., Recktenwald, D., Cameron, M. J., Danesh, A., Persad, D., et al. (2006). "Bead-based flow Cytometric assays: a multiplex assay platform with applications in diagnostic microbiology" in Advanced techniques in diagnostic microbiology (Boston, MA: Springer US), 427–443.

Evans, L., Rhodes, A., Alhazzani, W., Antonelli, M., Coopersmith, C. M., French, C., et al. (2021). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med.* 47, 1181–1247. doi: 10.1007/s00134-021-06506-y

Fatsis-Kavalopoulos, N., Roemhild, R., Tang, P.-C., Kreuger, J., and Andersson, D. I. (2020). CombiANT: antibiotic interaction testing made easy. *PLoS Biol.* 18:e3000856. doi: 10.1371/journal.pbio.3000856

Fleming, J., Marvel, S. W., Supak, S., Motsinger-Reif, A. A., and Reif, D. M. (2022). ToxPi*GIS toolkit: creating, viewing, and sharing integrative visualizations for geospatial data using ArcGIS. *J. Expo. Sci. Environ. Epidemiol.* 32, 900–907. doi: 10.1038/s41370-022-00433-w

Gammel, N., Ross, T. L., Lewis, S., Olson, M., Henciak, S., Harris, R., et al. (2021). Comparison of an automated plate assessment system (APAS Independence) and artificial intelligence (AI) to manual plate Reading of methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* CHROMagar surveillance cultures. *J. Clin. Microbiol.* 59:e0097121. doi: 10.1128/JCM.00971-21

Gao, S., and Wang, H. (2022). Scenario prediction of public health emergencies using infectious disease dynamics model and dynamic Bayes. *Futur. Gener. Comput. Syst.* 127, 334–346. doi: 10.1016/j.future.2021.09.028

GBD (2019). Antimicrobial resistance collaborators. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 400, 2221–2248. doi: 10.1016/S0140-6736(22)02185-7

Goodman, K. E., Lessler, J., Cosgrove, S. E., Harris, A. D., Lautenbach, E., Han, J. H., et al. (2016). A clinical decision tree to predict whether a Bacteremic patient is infected with an extended-Spectrum β-lactamase-producing organism. *Clin. Infect. Dis.* 63, 896–903. doi: 10.1093/cid/ciw425

Goodswen, S. J., Barratt, J. L. N., Kennedy, P. J., Kaufer, A., Calarco, L., and Ellis, J. T. (2021). Machine learning and applications in microbiology. *FEMS Microbiol. Rev.* 45:fuab015. doi: 10.1093/femsre/fuab015

Gurvic, D., Leach, A. G., and Zachariae, U. (2022). Data-driven derivation of molecular substructures that enhance drug activity in gram-negative Bacteria. *J. Med. Chem.* 65, 6088–6099. doi: 10.1021/acs.jmedchem.1c01984

Han, N., Oh, O. H., Oh, J., Kim, Y., Lee, Y., Cha, W. C., et al. (2024). The application of knowledge-based clinical decision support systems to detect antibiotic allergy. *Antibiotics (Basel)* 13:244. doi: 10.3390/antibiotics13030244

Heinson, A. I., Woelk, C. H., and Newell, M.-L. (2015). The promise of reverse vaccinology. *Int. Health* 7, 85–89. doi: 10.1093/inthealth/ihv002

Hellmich, T. R., Clements, C. M., El-Sherif, N., Pasupathy, K. S., Nestler, D. M., Boggust, A., et al. (2017). Contact tracing with a real-time location system: a case study of increasing relative effectiveness in an emergency department. *Am. J. Infect. Control* 45, 1308–1311. doi: 10.1016/j.ajic.2017.08.014

Hirosawa, T., Harada, Y., Yokose, M., Sakamoto, T., Kawamura, R., and Shimizu, T. (2023). Diagnostic accuracy of differential-diagnosis lists generated by generative Pretrained transformer 3 Chatbot for clinical vignettes with common chief complaints: a pilot study. *Int. J. Environ. Res. Public Health* 20:3378. doi: 10.3390/ijerph20043378

Ho, C.-S., Jean, N., Hogan, C. A., Blackmon, L., Jeffrey, S. S., Holodniy, M., et al. (2019). Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* 10:4927. doi: 10.1038/s41467-019-12898-9

Horvath, L., Hänselmann, S., Mannsperger, H., Degenhardt, S., Last, K., Zimmermann, S., et al. (2020). Machine-assisted interpretation of auramine stains substantially increases through-put and sensitivity of microscopic tuberculosis diagnosis. *Tuberculosis (Edinb.)* 125:101993. doi: 10.1016/j.tube.2020.101993

Howard, A., Aston, S., Gerada, A., Reza, N., Bincalar, J., Mwandumba, H., et al. (2024). Antimicrobial learning systems: an implementation blueprint for artificial intelligence to tackle antimicrobial resistance. *Lancet Digit. Health* 6, e79–e86. doi: 10.1016/S2589-7500(23)00221-2

Hsu, K.-W., Lee, W.-B., You, H.-L., Lee, M. S., and Lee, G.-B. (2021). An automated and portable antimicrobial susceptibility testing system for urinary tract infections. *Lab Chip* 21, 755–763. doi: 10.1039/d0lc01315c

Hummel, P., and Braun, M. (2020). Just data? Solidarity and justice in data-driven medicine. *Life Sci. Soc. Policy* 16:8. doi: 10.1186/s40504-020-00101-7

CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 33rd ed.CLSI supplement M100. Clinical and Laboratory Standards Institute; 2023. https://iacld.com/UpFiles/Documents/672a1c7c-d4ad-404e-b10e-97c19e21cdce.pdf [Accessed April 7, 2024]

Jiang, Y., Luo, J., Huang, D., Liu, Y., and Li, D. (2022). Machine learning advances in microbiology: a review of methods and applications. *Front. Microbiol.* 13:925454. doi: 10.3389/fmicb.2022.925454

Keith, M., Park de la Torriente, A., Chalka, A., Vallejo-Trujillo, A., McAteer, S. P., Paterson, G. K., et al. (2024). Predictive phage therapy for *Escherichia coli* urinary tract infections: cocktail selection for therapy based on machine learning models. *Proc. Natl. Acad. Sci. USA* 121:e2313574121. doi: 10.1073/pnas.2313574121

Khanna, D., and Rana, P. S. (2019). Ensemble technique for prediction of T-cell *Mycobacterium tuberculosis* epitopes. *Interdiscip. Sci.* 11, 611–627. doi: 10.1007/s12539-018-0309-0

Kleandrova, V. V., and Speck-Planche, A. (2020). The QSAR paradigm in fragment-based drug discovery: from the virtual generation of target inhibitors to multi-scale modeling. *Mini Rev. Med. Chem.* 20, 1357–1374. doi: 10.2174/1389557520666200204123156

Kulshrestha, M., Tiwari, M., and Tiwari, V. (2024). Bacteriophage therapy against ESKAPE bacterial pathogens: current status, strategies, challenges, and future scope. *Microb. Pathog.* 186:106467. doi: 10.1016/j.micpath.2023.106467

Laliwala, A., Svechkarev, D., Sadykov, M. R., Endres, J., Bayles, K. W., and Mohs, A. M. (2022). Simpler procedure and improved performance for pathogenic Bacteria analysis with a paper-based Ratiometric fluorescent sensor Array. *Anal. Chem.* 94, 2615–2624. doi: 10.1021/acs.analchem.1c05021

Lane, T. R., Urbina, F., Rank, L., Gerlach, J., Riabova, O., Lepioshkin, A., et al. (2022). Machine learning models for *Mycobacterium tuberculosis in vitro* activity: prediction and target visualization. *Mol. Pharm.* 19, 674–689. doi: 10.1021/acs.molpharmaceut.1c00791

Langford, B. J., Branch-Elliman, W., Nori, P., Marra, A. R., and Bearman, G. (2024). Confronting the disruption of the infectious diseases workforce by artificial intelligence: what this means for us and what we can do about it. *Open Forum Infect. Dis.* 11:ofae053. doi: 10.1093/ofid/ofae053

Larentzakis, A., and Lygeros, N. (2021). Artificial intelligence (AI) in medicine as a strategic valuable tool. *Pan Afr. Med. J.* 38:184. doi: 10.11604/pamj.2021.38.184.28197

Larsen, P. E., Collart, F. R., and Dai, Y. (2014). Using metabolomic and transportomic modeling and machine learning to identify putative novel therapeutic targets for antibiotic resistant pseudomonad infections. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2014, 314–317. doi: 10.1109/EMBC.2014.6943592

Lesosky, M., McGeer, A., Simor, A., Green, K., Low, D. E., and Raboud, J. (2011). Effect of patterns of transferring patients among healthcare institutions on rates of nosocomial methicillin-resistant *Staphylococcus aureus* transmission: a Monte Carlo simulation. *Infect. Control Hosp. Epidemiol.* 32, 136–147. doi: 10.1086/657945

Li, R., Shen, M., Liu, H., Bai, L., and Zhang, L. (2023). Do infrared thermometers hold promise for an effective early warning system for emerging respiratory infectious diseases? *JMIR Form Res.* 7:e42548. doi: 10.2196/42548

Liu, W., Ying, N., Mo, Q., Li, S., Shao, M., Sun, L., et al. (2021). Machine learning for identifying resistance features of Klebsiella pneumoniae using whole-genome sequence single nucleotide polymorphisms. *J. Med. Microbiol.* 70. doi: 10.1099/jmm.0.001474

Lu, J., Chen, J., Liu, C., Zeng, Y., Sun, Q., Li, J., et al. (2022). Identification of antibiotic resistance and virulence-encoding factors in *Klebsiella pneumoniae* by Raman spectroscopy and deep learning. *Microb. Biotechnol.* 15, 1270–1280. doi: 10.1111/1751-7915.13960

Mandal, S., Tannert, A., Löffler, B., Neugebauer, U., and Silva, L. B. (2024). Findaureus: an open-source application for locating *Staphylococcus aureus* in fluorescence-labelled infected bone tissue slices. *PLoS One* 19:e0296854. doi: 10.1371/journal.pone.0296854

Mc Cord-De Iaco, K. A., Gesualdo, F., Pandolfi, E., Croci, I., and Tozzi, A. E. (2023). Machine learning clinical decision support systems for surveillance: a case study on pertussis and RSV in children. *Front. Pediatr.* 11:1112074. doi: 10.3389/fped.2023.1112074

McGregor, J. C., Weekes, E., Forrest, G. N., Standiford, H. C., Perencevich, E. N., Furuno, J. P., et al. (2006). Impact of a computerized clinical decision support system on reducing inappropriate antimicrobial use: a randomized controlled trial. *J. Am. Med. Inform. Assoc.* 13, 378–384. doi: 10.1197/jamia.M2049

Mintz, Y., and Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minim. Invasive Ther. Allied Technol.* 28, 73–81. doi: 10.1080/13645706.2019.1575882

Nakar, A., Pistiki, A., Ryabchykov, O., Bocklitz, T., Rösch, P., and Popp, J. (2022). Detection of multi-resistant clinical strains of *E. coli* with Raman spectroscopy. *Anal. Bioanal. Chem.* 414, 1481–1492. doi: 10.1007/s00216-021-03800-y

Nguyen, M., Brettin, T., Long, S. W., Musser, J. M., Olsen, R. J., Olson, R., et al. (2018). Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci. Rep.* 8:421. doi: 10.1038/s41598-017-18972-w

Njage, P. M. K., Leekitcharoenphon, P., and Hald, T. (2019). Improving hazard characterization in microbial risk assessment using next generation sequencing data and machine learning: predicting clinical outcomes in shigatoxigenic *Escherichia coli*. *Int. J. Food Microbiol.* 292, 72–82. doi: 10.1016/j.ijfoodmicro.2018.11.016

Oh, J., Makar, M., Fusco, C., McCaffrey, R., Rao, K., Ryan, E. E., et al. (2018). A generalizable, data-driven approach to predict daily risk of *Clostridium difficile* infection at two large academic health centers. *Infect. Control Hosp. Epidemiol.* 39, 425–433. doi: 10.1017/ice.2018.16

Ohkusu, K. (2000). Cost-effective and rapid presumptive identification of gram-negative bacilli in routine urine, pus, and stool cultures: evaluation of the use of CHROMagar orientation medium in conjunction with simple biochemical tests. *J. Clin. Microbiol.* 38, 4586–4592. doi: 10.1128/JCM.38.12.4586-4592.2000

Paquin, P., Durmort, C., Paulus, C., Vernet, T., Marcoux, P. R., and Morales, S. (2022). Spatio-temporal based deep learning for rapid detection and identification of bacterial colonies through lens-free microscopy time-lapses. *PLoS Digit. Health* 1:e0000122. doi: 10.1371/journal.pdig.0000122

Pérez-Sancho, M., Vela, A. I., Horcajo, P., Ugarte-Ruiz, M., Domínguez, L., Fernández-Garayzábal, J. F., et al. (2018). Rapid differentiation of *Staphylococcus aureus* subspecies based on MALDI-TOF MS profiles. *J. Vet. Diagn. Invest.* 30, 813–820. doi: 10.1177/1040638718805537

Periasamy, A. (2014). Advanced light microscopy. *Methods* 66, 121–123. doi: 10.1016/j.ymeth.2014.03.011

Pfeiffer, D. U., and Stevens, K. B. (2015). Spatial and temporal epidemiological analysis in the big data era. *Prev. Vet. Med.* 122, 213–220. doi: 10.1016/j.prevetmed.2015.05.012

Qiu, J., Nie, W., Ding, H., Dai, J., Wei, Y., Li, D., et al. (2024). PB-LKS: a python package for predicting phage-bacteria interaction through local K-mer strategy. *Brief. Bioinform.* 25:bbae010. doi: 10.1093/bib/bbae010

Ramachandran, P. S., Ramesh, A., Creswell, F. V., Wapniarski, A., Narendra, R., Quinn, C. M., et al. (2022). Integrating central nervous system metagenomics and host response for diagnosis of tuberculosis meningitis and its mimics. *Nat. Commun.* 13:1675. doi: 10.1038/s41467-022-29353-x

Rapún-Araiz, B., Sorzabal-Bellido, I., Asensio-López, J., Lázaro-Díez, M., Ariz, M., Sobejano de la Merced, C., et al. (2023). *In vitro* modeling of polyclonal infection dynamics within the human airways by *Haemophilus influenzae* differential fluorescent labeling. *Microbiol. Spectr.* 11:e0099323. doi: 10.1128/spectrum.00993-23

Rawal, K., Sinha, R., Abbasi, B. A., Chaudhary, A., Nath, S. K., Kumari, P., et al. (2021). Identification of vaccine targets in pathogens and design of a vaccine using computational approaches. *Sci. Rep.* 11:17626. doi: 10.1038/s41598-021-96863-x

Rees, C., and Müller, B. (2022). All that glitters is not gold: trustworthy and ethical AI principles. *AI Ethics* 3, 1241–1254. doi: 10.1007/s43681-022-00232-x

Rhodes, N. J., Rohani, R., Yarnold, P. R., Pawlowski, A. E., Malczynski, M., Qi, C., et al. (2023). Machine learning to stratify methicillin-resistant *Staphylococcus aureus* risk among hospitalized patients with community-acquired pneumonia. *Antimicrob. Agents Chemother.* 67:e0102322. doi: 10.1128/aac.01023-22

Rodrigues Lopes, I., Alcantara, L. M., Silva, R. J., Josse, J., Vega, E. P., Cabrerizo, A. M., et al. (2022). Microscopy-based phenotypic profiling of infection by *Staphylococcus aureus* clinical isolates reveals intracellular lifestyle as a prevalent feature. *Nat. Commun.* 13:7174. doi: 10.1038/s41467-022-34790-9

Santa Maria, J. P., Park, Y., Yang, L., Murgolo, N., Altman, M. D., Zuck, P., et al. (2017). Linking high-throughput screens to identify MoAs and novel inhibitors of *Mycobacterium tuberculosis* Dihydrofolate reductase. *ACS Chem. Biol.* 12, 2448–2456. doi: 10.1021/acschembio.7b00468

Schwartz, I. S., Link, K. E., Daneshjou, R., and Cortés-Penfield, N. (2024). Black box warning: large language models and the future of infectious diseases consultation. *Clin. Infect. Dis.* 78, 860–866. doi: 10.1093/cid/ciad633

Senescau, A., Kempowsky, T., Bernard, E., Messier, S., Besse, P., Fabre, R., et al. (2018). Innovative DendrisChips® Technology for a Syndromic Approach of *in vitro* diagnosis: application to the respiratory infectious diseases. *Diagnostics (Basel)* 8:77. doi: 10.3390/diagnostics8040077

Shen, Y., Yuan, K., Chen, D., Colloc, J., Yang, M., Li, Y., et al. (2018). An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. *Artif. Intell. Med.* 86, 20–32. doi: 10.1016/j.artmed.2018.01.003

Sherry, N. L., Horan, K. A., Ballard, S. A., Gonçalves da Silva, A., Gorrie, C. L., Schultz, M. B., et al. (2023). An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. *Nat. Commun.* 14:60. doi: 10.1038/s41467-022-35713-4

Stracy, M., Snitser, O., Yelin, I., Amer, Y., Parizade, M., Katz, R., et al. (2022). Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. *Science (New York, N.Y.)* 375, 889–894. doi: 10.1126/science.abg9868

Tagliaferri, T. L., Jansen, M., and Horz, H.-P. (2019). Fighting pathogenic Bacteria on two fronts: phages and antibiotics as combined strategy. *Front. Cell. Infect. Microbiol.* 9:22. doi: 10.3389/fcimb.2019.00022

Tilton, C. S., and Johnson, S. W. (2019). Development of a risk prediction model for hospital-onset *Clostridium difficile* infection in patients receiving systemic antibiotics. *Am. J. Infect. Control* 47, 280–284. doi: 10.1016/j.ajic.2018.08.021

Ting Sim, J. Z., Fong, Q. W., Huang, W., and Tan, C. H. (2023). Machine learning in medicine: what clinicians should know. *Singapore Med. J.* 64, 91–97. doi: 10.11622/smedj.2021054

Váradi, L., Luo, J. L., Hibbs, D. E., Perry, J. D., Anderson, R. J., Orenga, S., et al. (2017). Methods for the detection and identification of pathogenic bacteria: past, present, and future. *Chem. Soc. Rev.* 46, 4818–4832. doi: 10.1039/c6cs00693k

Viertel, T. M., Ritter, K., and Horz, H.-P. (2014). Viruses versus bacteria-novel approaches to phage therapy as a tool against multidrug-resistant pathogens. *J. Antimicrob. Chemother.* 69, 2326–2336. doi: 10.1093/jac/dku173

Villarroel, J., Kleinheinz, K. A., Jurtz, V. I., Zschach, H., Lund, O., Nielsen, M., et al. (2016). HostPhinder: a phage host prediction tool. *Viruses* 8:116. doi: 10.3390/v8050116

Volynets, G. P., Usenko, M. O., Gudzera, O. I., Starosyla, S. A., Balanda, A. O., Syniugin, A. R., et al. (2022). Identification of dual-targeted *Mycobacterium tuberculosis* aminoacyl-tRNA synthetase inhibitors using machine learning. *Future. Med. Chem.* 14, 1223–1237. doi: 10.4155/fmc-2022-0085

Waddington, C., Carey, M. E., Boinett, C. J., Higginson, E., Veeraraghavan, B., and Baker, S. (2022). Exploiting genomics to mitigate the public health impact of antimicrobial resistance. *Genome Med.* 14:15. doi: 10.1186/s13073-022-01020-2

Walsh, T. R., Gales, A. C., Laxminarayan, R., and Dodd, P. C. (2023). Antimicrobial resistance: addressing a global threat to humanity. *PLoS Med.* 20:e1004264. doi: 10.1371/journal.pmed.1004264

Wang, H., Ceylan Koydemir, H., Qiu, Y., Bai, B., Zhang, Y., Jin, Y., et al. (2020). Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light Sci. Appl.* 9:118. doi: 10.1038/s41377-020-00358-9

Wang, H.-Y., Chung, C.-R., Wang, Z., Li, S., Chu, B.-Y., Horng, J.-T., et al. (2021). A large-scale investigation and identification of methicillin-resistant *Staphylococcus aureus* based on peaks binning of matrix-assisted laser desorption ionization-time of flight MS spectra. *Brief. Bioinform.* 22:bbaa138. doi: 10.1093/bib/bbaa138

Wang, M., Wei, Z., Jia, M., Chen, L., and Ji, H. (2022). Deep learning model for multi-classification of infectious diseases from unstructured electronic medical records. *BMC Med. Inform. Decis. Mak.* 22:41. doi: 10.1186/s12911-022-01776-y

Wells, J., Grant, R., Chang, J., and Kayyali, R. (2021). Evaluating the usability and acceptability of a geographical information system (GIS) prototype to visualise socio-economic and public health data. *BMC Public Health* 21:2151. doi: 10.1186/s12889-021-12072-1

Wieser, A., Schneider, L., Jung, J., and Schubert, S. (2012). MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). *Appl. Microbiol. Biotechnol.* 93, 965–974. doi: 10.1007/s00253-011-3783-4

Wilson, M. L. (2015). Diagnostic microbiology: the accelerating transition from culture-based to molecular-based methods. *Am. J. Clin. Pathol.* 143, 766–767. doi: 10.1309/AJCPIC9GPLHCV1NT

Wong, F., De La Fuente-Nunez, C., and Collins, J. J. (2023). Leveraging artificial intelligence in the fight against infectious diseases. *Science* 381, 164–170. doi: 10.1126/science.adh1114

Yan, Y., Chen, C., Liu, Y., Zhang, Z., Xu, L., and Pu, K. (2021). Application of machine learning for the prediction of etiological types of classic fever of unknown origin. *Front. Public Health* 9:800549. doi: 10.3389/fpubh.2021.800549

Yi, Q., Cai, D., Xiao, M., Nie, M., Cui, Q., Cheng, J., et al. (2019). Direct antimicrobial susceptibility testing of bloodstream infection on SlipChip. *Biosens. Bioelectron.* 135, 200–207. doi: 10.1016/j.bios.2019.04.003

Ying, C., Li, X., Lv, S., Du, P., Chen, Y., Fu, H., et al. (2022). T-SPOT with CT image analysis based on deep learning for early differential diagnosis of nontuberculous mycobacteria pulmonary disease and pulmonary tuberculosis. *Int. J. Infect. Dis.* 125, 42–50. doi: 10.1016/j.ijid.2022.09.031

Youn, J., Rai, N., and Tagkopoulos, I. (2022). Knowledge integration and decision support for accelerated discovery of antibiotic resistance genes. *Nat. Commun.* 13:2360. doi: 10.1038/s41467-022-29993-z

Younes, S., Al-Sulaiti, A., Nasser, E. A. A., Najjar, H., and Kamareddine, L. (2020). Drosophila as a model organism in host-pathogen interaction studies. *Front. Cell. Infect. Microbiol.* 10:214. doi: 10.3389/fcimb.2020.00214

Yu, J., Lin, Y.-T., Chen, W.-C., Tseng, K.-H., Lin, H.-H., Tien, N., et al. (2023b). Direct prediction of carbapenem-resistant, carbapenemase-producing, and colistin-resistant *Klebsiella pneumoniae* isolates from routine MALDI-TOF mass spectra using machine learning and outcome evaluation. *Int. J. Antimicrob. Agents* 61:106799. doi: 10.1016/j.ijantimicag.2023.106799

Yu, J., Lin, H.-H., Tseng, K.-H., Lin, Y.-T., Chen, W.-C., Tien, N., et al. (2023a). Prediction of methicillin-resistant Staphylococcus aureus and carbapenem-resistant *Klebsiella pneumoniae* from flagged blood cultures by combining rapid Sepsityper MALDI-TOF mass spectrometry with machine learning. *Int. J. Antimicrob. Agents* 62:106994. doi: 10.1016/j.ijantimicag.2023.106994

Zhang, Y.-M., Tsao, M.-F., Chang, C.-Y., Lin, K.-T., Keller, J. J., and Lin, H.-C. (2023). Rapid identification of carbapenem-resistant *Klebsiella pneumoniae* based on matrix-assisted laser desorption ionization time-of-flight mass spectrometry and an artificial neural network model. *J. Biomed. Sci.* 30:25. doi: 10.1186/s12929-023-00918-2

frontiers | Frontiers in Microbiology

TYPE Technology and Code
PUBLISHED 19 August 2024
DOI 10.3389/fmicb.2024.1453870

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Heng Zhang,
University of Science and Technology of China, China
Baohua Qiang,
Guilin University of Electronic Technology, China
Yuanyuan Jia,
Chongqing Medical University, China

*CORRESPONDENCE
Xiaojuan Liu
✉ liuxiaojuan0127@cqut.edu.cn

RECEIVED 24 June 2024
ACCEPTED 05 August 2024
PUBLISHED 19 August 2024

CITATION
Liu X, Xiang C, Lan L, Li C, Xiao H and Liu Z (2024) Lesion region inpainting: an approach for pseudo-healthy image synthesis in intracranial infection imaging.
Front. Microbiol. 15:1453870.
doi: 10.3389/fmicb.2024.1453870

Lesion region inpainting: an approach for pseudo-healthy image synthesis in intracranial infection imaging

Xiaojuan Liu[1,2]*, Cong Xiang[1], Libin Lan[3], Chuan Li[2], Hanguang Xiao[1] and Zhi Liu[1]

[1]College of Artificial Intelligence, Chongqing University of Technology, Chongqing, China, [2]College of Big Data and Intelligent Engineering, Chongqing College of International Business and Economics, Chongqing, China, [3]College of Computer Science and Engineering, Chongqing University of Technology, Chongqing, China

The synthesis of pseudo-healthy images, involving the generation of healthy counterparts for pathological images, is crucial for data augmentation, clinical disease diagnosis, and understanding pathology-induced changes. Recently, Generative Adversarial Networks (GANs) have shown substantial promise in this domain. However, the heterogeneity of intracranial infection symptoms caused by various infections complicates the model's ability to accurately differentiate between pathological and healthy regions, leading to the loss of critical information in healthy areas and impairing the precise preservation of the subject's identity. Moreover, for images with extensive lesion areas, the pseudo-healthy images generated by these methods often lack distinct organ and tissue structures. To address these challenges, we propose a three-stage method (localization, inpainting, synthesis) that achieves nearly perfect preservation of the subject's identity through precise pseudo-healthy synthesis of the lesion region and its surroundings. The process begins with a Segmentor, which identifies the lesion areas and differentiates them from healthy regions. Subsequently, a Vague-Filler fills the lesion areas to construct a healthy outline, thereby preventing structural loss in cases of extensive lesions. Finally, leveraging this healthy outline, a Generative Adversarial Network integrated with a contextual residual attention module generates a more realistic and clearer image. Our method was validated through extensive experiments across different modalities within the BraTS2021 dataset, achieving a healthiness score of 0.957. The visual quality of the generated images markedly exceeded those produced by competing methods, with enhanced capabilities in repairing large lesion areas. Further testing on the COVID-19-20 dataset showed that our model could effectively partially reconstruct images of other organs.

KEYWORDS

pseudo-healthy image synthesis, generative adversarial networks, intracranial infection, data augmentation, contextual residual attention module lesion inpainting for pseudo-healthy synthesis

1 Introduction

Intracranial infections, involving the brain and its adjacent structures, pose significant clinical challenges due to their potential to cause severe outcomes. Characterized by inflammation and infection within the cranial cavity, these conditions affect the brain parenchyma, meninges, and other intracranial structures. A wide range of pathogens, including bacteria, viruses, fungi, and parasites, can instigate various infections such as meningitis, encephalitis, brain abscesses, and subdural empyemas (Foerster et al., 2007). Magnetic Resonance Imaging (MRI) is crucial in detecting, assessing, and monitoring these central nervous system infections and inflammations (Zimmerman and Girard, 2004; Mitchell and Dehkharghani, 2014). By providing comprehensive imaging of the brain and its meningeal coverings, MRI helps identify distinct patterns and features indicative of different types of intracranial infections. For example, MRI excels in distinguishing between pyogenic and fungal abscesses; pyogenic abscesses typically present with a well-defined rim and surrounding edema. In cases of ventriculitis (Luque-Paz et al., 2021), MRI can display ventricular enlargement, ependymal enhancement, and intraventricular debris. Enhancing MRI images in instances of intracranial infection is thus essential, as it provides clinicians with critical diagnostic information, improving both diagnostic accuracy and efficiency. Figure 1 illustrates an instance of intracranial infection (Deng and Gaillard, 2014).

In recent years, the development of pseudo-healthy image synthesis technology has become a pivotal tool in data augmentation and medical image anomaly detection. In the realm of data augmentation, generating pseudo-healthy images significantly enriches datasets by creating numerous representations from the same subjects' pathological images (Xia et al., 2020). This technique not only bolsters the model's generalization capabilities but also mitigates challenges associated with data imbalance and limited sample availability. In anomaly detection within medical imaging, synthesizing pseudo-healthy images allows models to simulate representations of healthy tissues (Tsunoda et al., 2014). By contrasting these images with their pathological counterparts, clinicians can more accurately pinpoint lesions. Thus, the production of high-quality pseudo-healthy images is crucial for enhancing the detection and diagnosis of conditions like intracranial infections. Furthermore, comparing pathological images with pseudo-healthy ones deepens the understanding of pathology-induced alterations, thereby advancing insights into disease progression and pathology development processes.

The process of synthesizing pseudo-healthy images involves generating apparently normal, lesion-free images from pathological data using sophisticated computer imaging and machine learning techniques. Ideally, a pseudo-healthy image should possess two essential attributes (Zhang et al., 2022): First, the image must maintain a healthy appearance, closely mimicking a genuine healthy image. This is the primary goal of pseudo-healthy image synthesis. Second, the synthesized image must originate from the same individual as the pathological image. This requirement is equally important, as producing healthy images from different individuals does not aid in medical diagnosis (Bowles et al., 2016). Typically, it is not feasible for the tissues or organs of a single patient to exhibit both pathological and healthy states simultaneously. Therefore, identifying an exact corresponding pseudo-healthy image for a specific pathological image is inherently complex and fraught with uncertainties. In the context of pseudo-healthy synthesis for intracranial infection, the varied manifestations of the disease in MRI images present significant challenges. For instance, severe cerebral edema in lesion areas can cause a mass effect, compressing and deforming adjacent brain ventricles. Consequently, pseudo-healthy synthesis for intracranial infection should focus on restoring the anatomical integrity in affected regions and accommodating the disease's diverse presentations.

The synthesis of pseudo-healthy images entails creating seemingly normal, lesion-free images from pathological data through the use of advanced computer image processing and machine learning techniques. Determining whether an image is truly pseudo-healthy hinges on the absence of pathological features, while maintaining the subject's identity depends on the intact preservation of non-pathological regions. Consequently, in pseudo-healthy synthesis, accurately localizing pathological regions and reconstructing their healthy analogs is paramount. Several Generative Adversarial Network (GAN)-based approaches for pseudo-healthy image synthesis have been previously proposed (Baumgartner et al., 2018; Chen and Konukoglu, 2018; Baur et al., 2019, 2020). These methods typically employ a generator, structured as an encoder-decoder network, to convert pathological images into their healthy-looking equivalents. Simultaneously, a discriminator, competing against the generator, utilizes a classifier to differentiate between the synthesized healthy images and actual healthy images. Through this adversarial training process, the generator and classifier refine their capabilities in a dynamic interplay. However, a significant limitation of these methods is their inability to effectively learn and incorporate pathological information, which complicates the task of maintaining the subject's identity in the synthesized pseudo-healthy images. To overcome these challenges, Xia et al. (2020) and Zhang et al. (2022) introduced the use of a segmentor alongside pixel-level annotations. This strategy involves the collaborative training of both the generator and the segmentor. The segmentor's training loss is fed back to the generator, encouraging it to differentiate pathological information from the subject's identity while preserving any healthy attributes present in the pathological image. Despite these advancements, the methods still face several drawbacks.

(1) Integrating the segmentor into the model results in an overdependence on the segmentor's efficacy for lesion localization. The varied etiologies underlying intracranial infections lead to significantly diverse symptoms. Thus, a singular segmentation strategy is evidently inadequate to meet these demands.

(2) The generator creates segments devoid of lesion regions, whereas the classifier's visual focus is primarily on healthy areas. This causes the generator to employ images from different subjects to deceive the discriminator, unintentionally erasing the unique identity of the subject.

(3) These models demonstrate a deficiency in learning anatomical structures from healthy images, thus hindering their capacity to accurately reconstruct anatomical features within lesion areas, especially in cases involving extensive lesions.

To address the challenges and accommodate the anatomical alterations caused by intracranial infections, we introduce a novel three-stage pseudo-healthy image synthesis model called the Lesion Region Inpainting Generative Adversarial Network (LRI-GAN). This model is specifically tailored to manage the varying characteristics of infection areas in brain imaging. It ensures the preservation of the subject's identity by accurately synthesizing

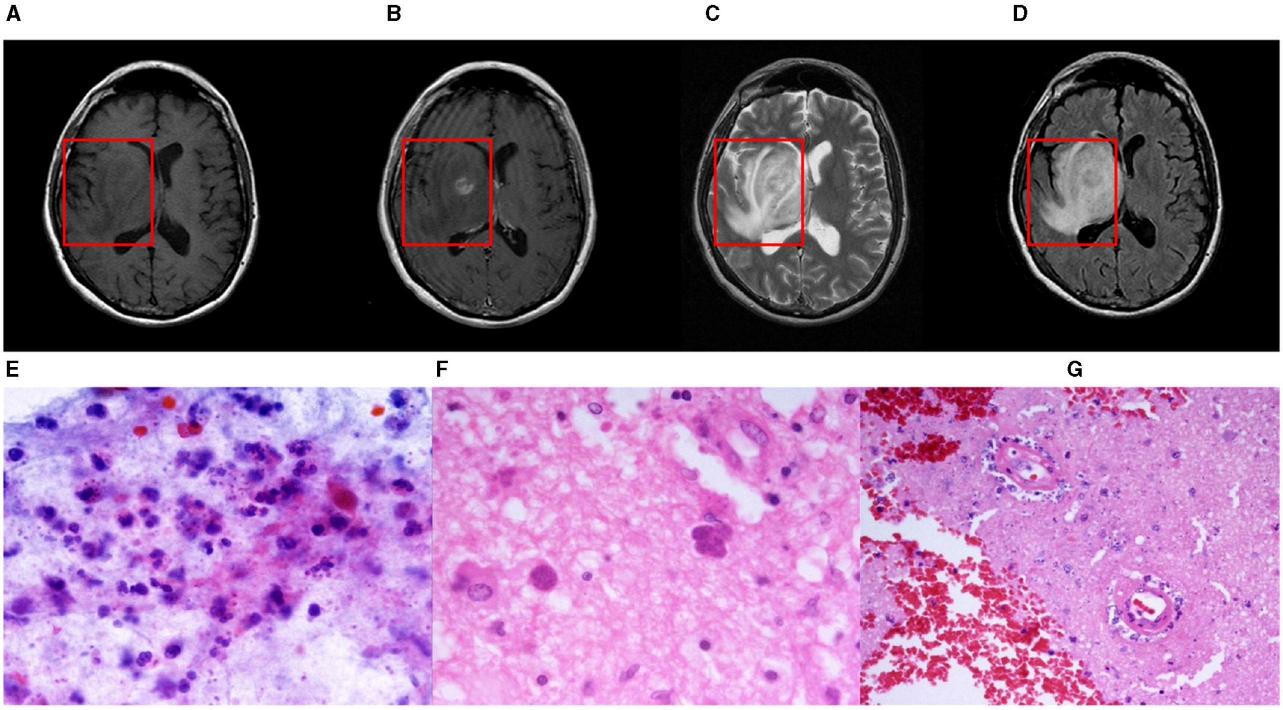


FIGURE 1
A nodular mass is centered in the right basal nuclei with an irregularly thick contrast-enhancing rim. The central portion of the mass is mildly hypointense to gray matter on T1-weighted imaging and hyperintense on T2-weighted imaging with poor suppression on FLAIR. Histological examination of paraffin sections 1 and 2 confirmed the presence of numerous Toxoplasma gondii tachyzoites and sporadic bradyzoite cysts. These parasites are embedded within a context of extensive cerebral parenchymal necrosis, where a distinct boundary demarcates the necrotic areas from the adjacent viable tissue. Imaging of necrotic regions reveals evidence of necrotic blood vessels, with the accumulation of neutrophils and chronic inflammatory cells, as well as nuclear debris in the surrounding and perivascular regions. Within the surviving parenchyma, reactive proliferation of small blood vessels is observed. Paraffin section 3 displays both cortical and white matter structures. In the deep white matter, there is a significant presence of Toxoplasma gondii tachyzoites along with sporadic bradyzoite cysts situated in areas of congestion and focal necrosis. Sporadic cysts can also be identified in the more superficial white matter and cortical regions. These findings are consistent with a diagnosis of cerebral toxoplasmosis. **(A)** T1-weight, **(B)** T2-weight **(C)** T1ce-weight **(D)** Flair-weight **(E)** Paraffin section 1 **(F)** Paraffin section 2 and **(G)** Paraffin section 3.

and replacing lesioned areas with pseudo-healthy regions, thus maintaining image integrity. The three-stage architecture enhances the model's effectiveness in constructing accurate healthy contours. Initially, various segmentation models are pre-trained, based on specific pathological requirements, or employing pixel-level annotations from clinical experts to precisely pinpoint lesion areas. Subsequently, in the second stage, a "Vague-filler" network fills the identified lesion regions, including an adjacent 5 mm area, capturing the essential characteristics of healthy tissues. The final stage employs a Generator network, enhanced with a contextual residual attention module, which adeptly learns from real healthy images and extracts relevant features from non-lesioned parts of the pathological image. This innovative approach results in pseudo-healthy images that not only reflect a clearer visual quality but also enhance diagnostic accuracy, as demonstrated in Figure 2. The LRI-GAN thus represents a significant advancement in medical imaging, particularly in the synthesis of images for diagnostic and treatment planning in cases of intracranial infection.

To assess the efficacy of our proposed method, we utilized image slices from the BraTS2021 dataset, featuring various conditions like edema, hemorrhage, and deformation. Our extensive testing shows that this method surpasses contemporary leading techniques in performance. Further validation was conducted using the COVID-19-20 dataset to evaluate the model's versatility across different organs, confirming consistent high performance.

Key contributions of our study include:

(1) Development of an advanced pseudo-healthy image synthesis approach tailored for intracranial infections, which preserves the identity of the pathological region with meticulous lesion area restoration.

(2) Introduction of a novel generator network architecture, incorporating a flipped symmetrical structure and a contextual residual attention mechanism, designed specifically to accurately mend lesioned areas.

(3) Establishment of a new evaluation metric called "Structure Healthiness" (SH), designed to gauge the capability of models to restore the anatomical integrity of lesion areas.

2 Related works

In the field of medical image analysis, the synthesis of pseudo-healthy images has attracted significant interest due to its potential benefits for various downstream applications. Research in this area can be categorized into two main groups based on the

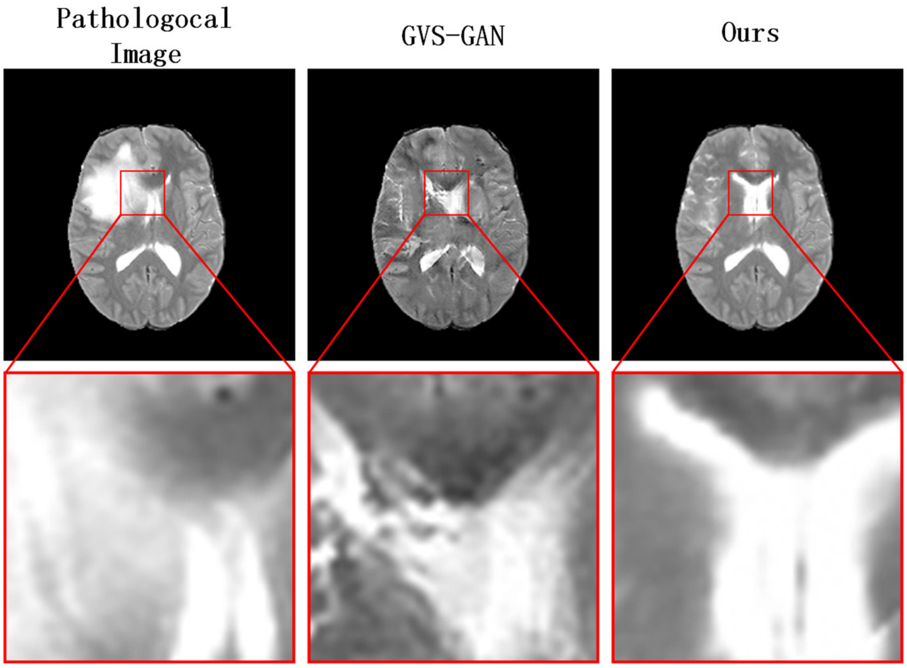


FIGURE 2
Our model generates clearer pseudo-healthy images in the presence of large lesions compared to the current state-of-the-art model GVS-GAN.

nature of the training data utilized (Zhang et al., 2022). The first category is Pathology-deficiency based methods. These methods exclusively rely on healthy images during the training process and are consequently devoid of pathological information. They do not require pathological data for training and are often closely associated with unsupervised medical image segmentation techniques (Bowles et al., 2017; Baumgartner et al., 2018; Tao et al., 2023; Rahman Siddiquee et al., 2024). The second category comprises Pathology-sufficiency based methods, which utilize a comprehensive dataset containing both pathological and healthy images during training. These approaches address the challenge of pseudo-healthy image synthesis from an image translation perspective. They incorporate pathological images along with their corresponding image-level or pixel-level pathological annotations to ensure that the synthesized pseudo-healthy images closely resemble the characteristics of healthy tissues (Sun et al., 2020; Xia et al., 2020; Zhang et al., 2022). This methodology facilitates more accurate and clinically relevant outputs by incorporating essential pathological details into the training process.

2.1 Pathology-deficiency based methods

Pathology-deficiency based methods begin by learning the normative distribution, leveraging techniques focused on compressing and recovering structures of healthy anatomical features during training. Subsequently, during the testing phase, these methods compress pathological images into a latent space. The underlying hypothesis is that the resultant latent representations closely approximate those of pseudo-healthy images, leading to the reconstruction of pseudo-healthy images from these representations. Chen and Konukoglu (2018) utilized an autoencoder-based approach to capture the distribution of brain MRIs from healthy subjects. Their objective was to map images to regions proximate to corresponding healthy images in latent space, employing specific constraints to guide this process. In a similar vein, Baur et al. (2019) modeled the distribution of healthy brain MRIs to identify pathological alterations through erroneous reconstructions. They implemented a Laplacian pyramid technique to compress and reconstruct healthy brain MRIs, which resulted in higher reconstruction fidelity at greater resolutions. Nevertheless, such methods are founded on idealized assumptions that often do not hold in practical scenarios. Specifically, the challenge lies in identifying an optimal latent representation that aligns with pseudo-healthy images when pathological images are compressed into the latent space. This difficulty frequently leads to a failure to preserve the identity of the pseudo-healthy images. Therefore, while the theoretical foundation of these methods is strong, their practical application is hindered by limitations in capturing and maintaining the true characteristics of the subject's healthy state in the synthesized images.

2.2 Pathology-sufficiency based methods

To synthesize higher-quality pseudo-healthy images, VA-GAN (Baumgartner et al., 2018) introduces a GAN-based framework that incorporates pathological information. This framework comprises a generator, tasked with synthesizing images that appear healthy while preserving the subject's identity, and a discriminator, which distinguishes between these synthesized images and real, unpaired healthy images. However, this method relies primarily on image-level annotations, which limits its ability to accurately differentiate between lesioned and non-lesioned areas, consequently impacting the preservation of the subject's identity in the synthesized images. To mitigate these limitations, PHS-GAN (Xia et al., 2020) and ANT-GAN (Sun et al., 2020), both variants of Cycle-GAN, incorporate pixel-level annotations. PHS-GAN addresses the one-to-many issue characteristic of medical images with variable pathology by employing a segmenter alongside pixel-level pathological annotations. This configuration allows precise localization of lesions, facilitating the separation of pathological information from healthy tissue, thus enhancing the precision of pseudo-healthy image synthesis. This method effectively manages pathological data to improve the accuracy and realism of the generated images. ANT-GAN, on the other hand, utilizes the L2 loss calculated between non-lesioned areas of the pathological and pseudo-healthy images. By reintegrating this feedback into the entire cyclic network, ANT-GAN ensures that the identity of the subject is maintained in the resultant images. To improve the localization of lesions, GVS-GAN (Zhang et al., 2022) attempts to resolve discrepancies between how healthy and pseudo-healthy images are perceived by the segmenter, aiming for a harmonized outcome. Nonetheless, these strategies, by trying to make the segmenter less sensitive to lesions, may not truly achieve the creation of "pseudo-healthy" images in the strictest sense. A persistent challenge with these methods is their struggle to fully grasp the anatomical features of a healthy brain, especially when faced with images featuring extensive lesions. This often leads to the generated images obscuring rather than restoring the anatomical structure of the affected areas. Consequently, while these approaches advance the field of pseudo-healthy image synthesis by better managing pathological information and improving image realism, they still face significant hurdles in accurately rendering and restoring the detailed anatomy in areas affected by pathology.

2.3 Our method

To facilitate the synthesis of pseudo-healthy images for intracranial infections, we have integrated the aforementioned methods and introduced a segmentation-first, then-repair strategy for pseudo-healthy synthesis. This approach differs from previous methodologies, which incorporated the segmentor within the generative network during the training phase, thus performing segmentation and generation simultaneously. Instead, our method employs the segmentor specifically to localize lesion areas, a strategy that prevents the segmentor from excessively influencing the generative network during training and ensures that the generator does not focus disproportionately on concealing lesions. Additionally, this segmentation-first approach allows for the flexible replacement of the segmentor, enhancing the model's adaptability to the varied manifestations of intracranial infections evident in MRI images. Our method executes the synthesis of pseudo-healthy images in a structured three-stage process. Initially, in the first stage, lesion areas are precisely identified using either a pre-trained segmentor model or manual annotations. Following this, the second stage employs a Vague-Filler network designed to infill these localized lesion areas, effectively mimicking the appearance of healthy tissue. In the final stage, a generator equipped with an inverted symmetrical structure and a contextual residual attention module (Yi et al., 2020) is utilized. This sophisticated arrangement enables the generator to learn effectively from both flipped images and features outside the lesion areas, thereby enhancing its capability to synthesize more accurate pseudo-healthy images tailored to the specific requirements of intracranial infection cases.

3 Methods

The architecture of LRI-GAN comprises three distinct components aligned with the workflow: a Segmentor (responsible for localization), a Vague-Filler (responsible for coarse filling), and a Generator (responsible for fine reconstruction). Both the Vague-Filler and the Generator are trainable elements, whereas the Segmentor is a pre-trained deep learning model or manually annotated pixel-level pathology. The structure of this paper is as follows: Section 3.1 provides an overview of the problem; Section 3.2 introduces the Segmentor; Section 3.3 describes the Vague-Filler; Section 3.4 elaborates on the Generator; Section 3.5 discusses the loss function of LRI-GAN; Section 3.6 outlines the training process of LRI-GAN; and Section 3.7 details the inference procedure of LRI-GAN.

3.1 Problem overview

As illustrated in Figure 3, we consider a set of images $\{x_i\}_{i=1}^{N} \in X$, with each i representing a slice, alongside their binary annotations $\{y_i\}_{i=1}^{N} \in Y$. These images are classified into two subsets based on their labels: pathological images $\{p_i\}_{i=1}^{M}$ and healthy images $\{h_i\}_{i=1}^{N-M}$. The data distributions of the pathological and healthy samples are denoted as $p_i \sim f_p$ and $h_i \sim f_h$, respectively.

In the inference pipeline, for a given pathological image p_i that contains lesion regions, our objective is to derive the corresponding y_i (where 0 indicates normal regions and 1 indicates pathological regions) via the Segmentor S. Subsequently, y_i is combined with p_i and fed into the Vague-filler V to produce a vague pseudo-healthy image $vague\hat{h}_i$. This image is then refined by the Generator G to yield a clearer pseudo-healthy image $\hat{h}_i$, ensuring that $\hat{h}_i$ adheres to the distribution of healthy images (i.e., $\hat{h}_i \sim f_h$). Moreover, we aim to maintain the normal anatomical structure of p_i within $\hat{h}_i$.

In the training pipeline, we emphasize healthy images to comprehensively learn their latent features. For a given healthy image h_i, we randomly mask 30%–60% of the regions to emulate the process of a pre-trained Segmentor detecting lesion regions, resulting in the corresponding mask y_i. This mask is then combined with h_i and input into the Vague-filler V to generate a vague pseudo-healthy image $vague\hat{h}_i$. The Generator G is then utilized to refine $vague\hat{h}_i$, producing a clearer pseudo-healthy image $\hat{h}_i$, which ensures that the masked regions in $\hat{h}_i$ closely resemble the original healthy image h_i.

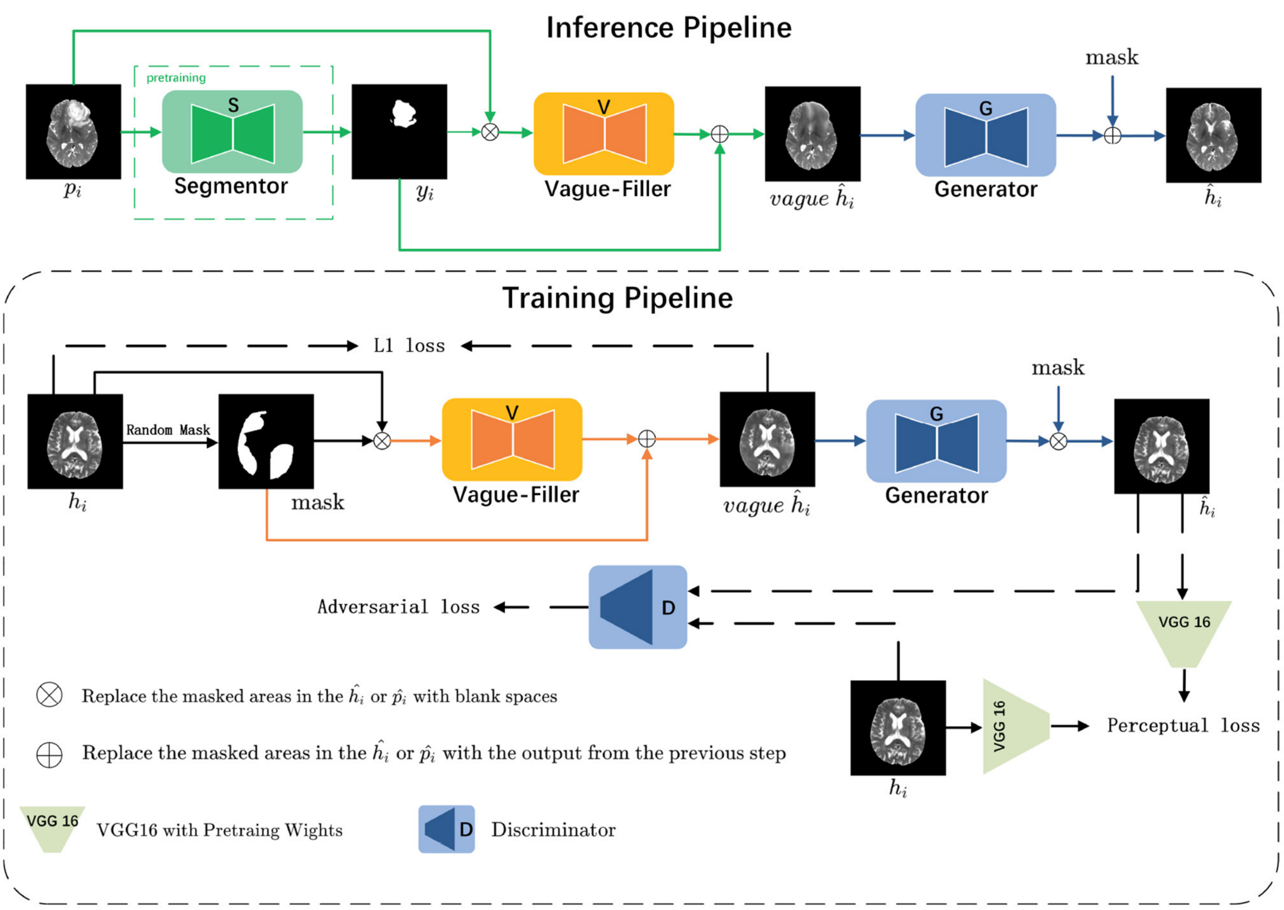


FIGURE 3
The LRI-GAN mainly consists of three components: the Segmentor, the Vague-Filler, and the Generator. During the training stage, the model is trained on healthy images. In the inference stage, it is applied to pathological images.

3.2 Segmentor

Before commencing the synthesis of pseudo-healthy images, accurately identifying lesion locations within pathological images is crucial. The primary aim during the Segmentor phase is to obtain pixel-level annotations y_i that precisely delineate lesion areas in the pathological image p_i. However, acquiring such detailed pathological annotations is often expensive and time-consuming. Therefore, for pathological images lacking specific annotations, we utilize a pre-trained segmentor, S, to automatically generate these annotations. In this study, we employ the U-Net architecture, renowned for its effectiveness in medical image segmentation, as the pre-trained segmentor (Ronneberger et al., 2015).

3.3 Vague-Filler

The Vague-Filler processes the pathological image p_i, where lesion regions are replaced by blanks, to produce a preliminary pseudo-healthy image, $vague\hat{h}_i$. Detailed insights into the Vague-Filler's methodology are provided in the "Vague Filler" section illustrated in Figure 4. This component accepts an image alongside a binary mask of lesion regions as inputs and outputs a filled-in image. It incorporates gated convolution as its sole learnable mechanism. The Vague-Filler operates on a "straight-line" residual network architecture devoid of skip connections, preserving the input and output dimensions at H × W pixels. To broaden the receptive field and minimize computational demands, the input image is initially down-sampled to $\frac{H}{2} \times \frac{W}{2}$ pixels prior to convolution. Subsequent convolutions further reduce the resolution to $\frac{H}{4} \times \frac{W}{4}$ pixels using two gated convolutions. The image then undergoes additional processing at the $\frac{H}{4} \times \frac{W}{4}$ scale via a sequence of gated convolutions, which vary in stride and padding, yet maintain a consistent size throughout the input and output stages.

3.4 Generator

The Generator's fundamental role is to enhance a vaguely defined pseudo-healthy image, denoted as $vague\hat{h}_i$, into a distinctly clearer image $\hat{h}_i$. This enhancement recognizes the symmetric nature of brain medical imagery, incorporating a flip-symmetric architecture detailed in the Generator section of Figure 3. Initially, $vague\hat{h}_i$ undergoes a flipping

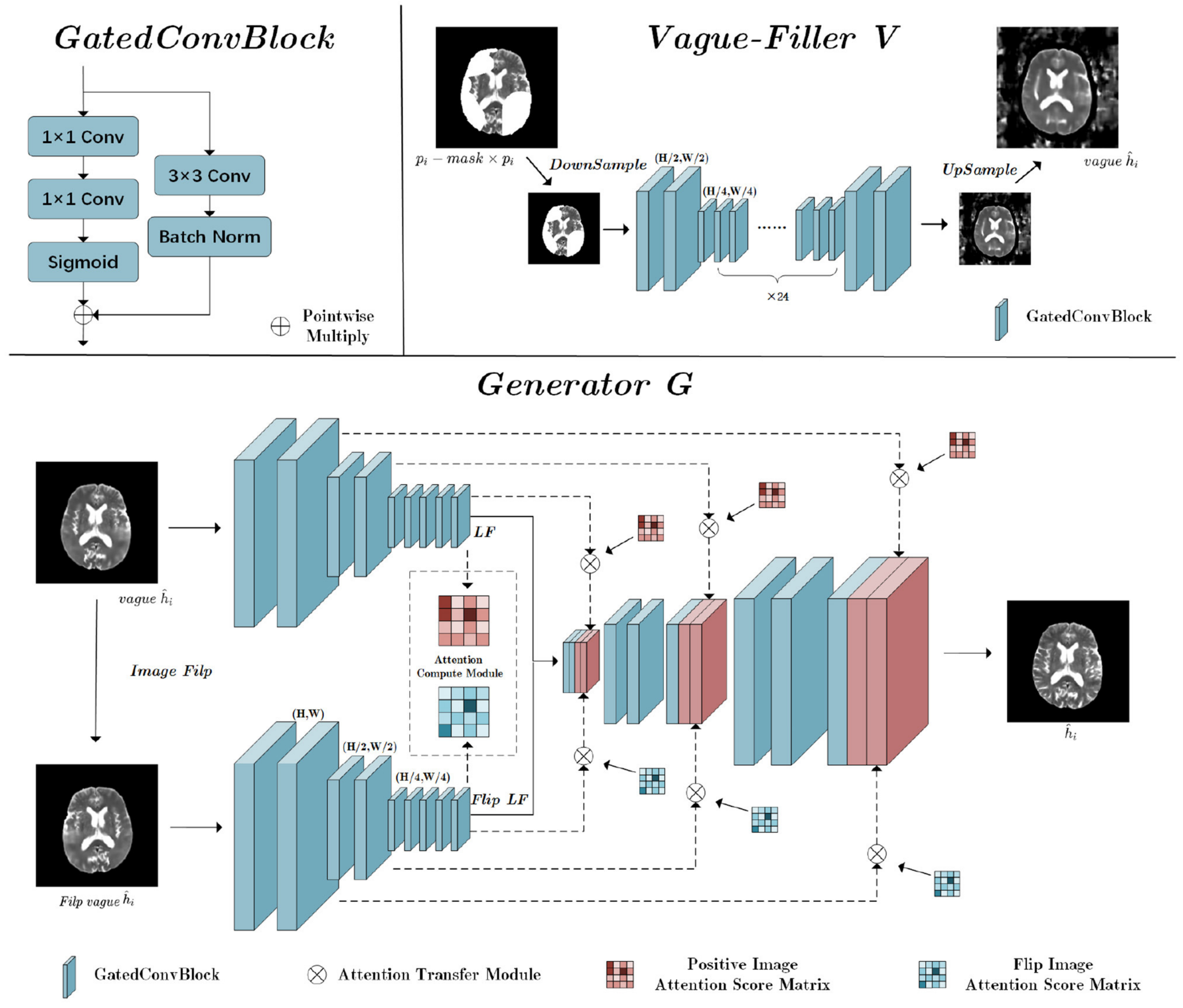


FIGURE 4
The structure of Vague-Filler V and Generator G; H is the height of the input image, and W is the width of the input image.

operation to prepare for convolutional processing. Both the original and flipped versions of $vague\hat{h}_i$ undergo parallel convolutional operations.

$$flip\ vague\hat{h}_i = flip(vague\ \hat{h}_i) \tag{1}$$

The convolution phase features a fully symmetric dual-path structure that optimizes feature extraction:

$$LF = down\_conv(vague\ \hat{h}_i) \tag{2}$$

$$LF\_flip = down\_conv(flip\ vague\hat{h}_i) \tag{3}$$

LF represents the latent features derived post-convolution, and $down\_conv$ refers to the down-sampling convolution process. The Attention Calculation Module (ACM) calculates attention score matrices for both the forward and flipped images:

$$PIAS\ Matrix = ACM(LF) \tag{4}$$

$$FIAS\ Matrix = ACM(LF\_flip) \tag{5}$$

The PIAS Matrix denotes the Positive Image Attention Score Matrix, detailing the interactions of the forward image with the mask area, while the FIAS Matrix is the Flip Image Attention Score Matrix, detailing interactions of the flipped image components with the mask. After calculating these matrices, both pathways integrate the residuals within the masked areas using their respective Attention Transfer Modules (ATM), based on the attention scores and contextual residuals:

$$\hat{h}_i = conv(LF, LF\_flip, ATM(LF, PIAS\ Matrix), ATM(LF\_flip, FIAS\ Matrix)) \tag{6}$$

The ACM uses cosine similarity measures for establishing image attention scores across high-level feature maps:

$$c_{i,j} = \left\langle \frac{b_i}{\|b_i\|}, \frac{b_j}{\|b_j\|} \right\rangle \tag{7}$$

b_i and b_j represent the patches outside and inside the mask area, respectively. The resultant similarity scores are squared and normalized to derive attention scores for each patch:

$$s_{i,j} = \frac{c_{i,j}^2}{\sum_{i=1}^{N} c_{i,j}^2} \tag{8}$$

N represents the number of patches outside the mask area. Despite the heterogeneity in lesion areas, a 256–256 matrix uniformly stores potential affinity scores between any pair of patches.

Finally, the ATM utilizes these attention scores to fill gaps in the low-level feature map with contextually weighted patches:

$$b_j = \sum_{i=1}^{N} s_{i,j} b_i \tag{9}$$

b_i is extracted from outside the masked area and b_j fills within the mask. Each patch measures 16 × 16, allowing for the extraction of 256 patches in total.

Through residual aggregation, the model reconstructs detailed aspects of the lesion area:

$$R_j = \sum_{i=1}^{N} s_{i,j} R_i \tag{10}$$

R denotes the residual image, with R_i and R_j representing the patches involved in filling the masked area. These patches cover all pixels seamlessly, ensuring a coherent integration of filled residuals with the surrounding tissue. The resultant aggregated residual image is then merged with the up-sampled blurry image from the generator to enhance clarity.

3.5 Loss function

3.5.1 Vague-Filler loss

L1 Loss: To ensure uniformity throughout the training process of the Vague-Filler, we utilize the L1 loss function. The formula for this is given by:

$$L_v = \frac{1}{N} \sum_{i=1}^{N} |h_i - \text{vague}\hat{h}_i|$$

In training the Vague-Filler, our objective is to enhance the model's focus on the contour structures of healthy images, while allowing a greater tolerance toward their textural features. The L1 loss function is chosen because it minimally penalizes large discrepancies and accommodates outliers effectively, making it an appropriate choice for this application.

3.5.2 Generator loss

To enhance the stability of the generator's training, we use the hinge loss method for adversarial training. Additionally, to enrich texture details in the generated images, we incorporate perceptual loss.

Adversarial loss: For the adversarial training of the generator, we employ the hinge loss method. The primary goal is to maximize the separation between positive and negative samples, thus enhancing categorical distinctions. This approach is based on the methodology used in the Geometric GAN (Lim and Ye, 2017), which has demonstrated improvements in the effectiveness of adversarial training. The adversarial losses for the discriminator and the generator are defined as follows:

$$L_D = \mathbb{E}\left[\max(0, 1 - D(h_i))\right] + \mathbb{E}\left[\max(0, 1 + D(G(p_i)))\right]$$

$$L_G = -\mathbb{E}[D(G(h_i))]$$

Here, G represents the generator, and D represents the discriminator. For D, only positive samples where $D(X) < 1$ and negative samples where $D(G(z)) > -1$ impact the outcome, implying that a small fraction of samples exceeding these thresholds will not influence the gradients. This results in more stable training dynamics.

Perceptual loss: To ensure that the images generated by the generator network during high-definition reconstruction closely align with the visual characteristics of healthy tissues, we incorporate perceptual loss. Perceptual loss emphasizes the perceptual quality of the restored images rather than solely focusing on pixel-level differences. This loss is widely used in medical imaging to enhance the restoration of textural details (Yang et al., 2018; Li et al., 2021). The perceptual loss is defined as follows:

$$L_{perc}(G) = \lambda \mathbb{E}\left[\|\phi(h_i) - \phi(G(h_i))\|\right]$$

Here, ϕ denotes the feature extraction function from the VGG16 network, and λ, the weight of the texture loss, is set to 64. The final loss function of the generator is:

$$\text{loss} = L_G + L_{perc}(G)$$

3.6 Training pipeline

The training process is elucidated within the "Training Pipeline" section, as depicted in Figure 3. During this phase, the grayscale values of all images are linearly adjusted to a range of $[-1, 1]$. Masks are designated by a value of 1 for missing regions and 0 for background areas. In this context, h_i represents the input healthy image, $\tilde{h}_i$ denotes the generated healthy image, and m indicates the mask for missing regions. The operation $\odot$ signifies element-wise multiplication. The Vague-Filler V interprets the concatenated masked image and mask as inputs to forecast the vague image $\text{vague}_y = V(h_i, m)$, maintaining the same dimensions as the input image. Following this, the Generator G uses the combined vague image and mask to predict $y = G(\text{vague}_y, m)$,

```
Input: Training data $H$, Vague-filler model $V$, Generator model $G$, Discriminator model $D$
Output: Trained Vague-filler model $V$, Generator model $G$, Discriminator model $D$
Initialization;
while V or G has not converged do
    // In order to amplify the adversarial impact, the discriminator undergoes two updates for each single update of the generator.
    for 1, 2 do
        Sampling batch images $h_i$ from training data $H$;
        Generating random masks $m$ for $h_i$;
        Getting inpainted vague $y \leftarrow V(h_i, m)$;
        Pasting back vague $\tilde{h}_i \leftarrow y \odot m + h_i \odot (1 - m)$;
        Getting Pseudo-Healthy inpainted vague $y \leftarrow G(\tilde{h}_i, m)$;
        Pasting back $\tilde{h}_i \leftarrow y \odot m + h_i \odot (1 - m)$;
        Updating the Discriminator $D$ with loss $L_D$;
    end for
    Sampling batch images $h_i$ from training data $H$;
    Generating random masks $m$ for $h_i$;
    Getting inpainted vague $y \leftarrow V(h_i, m)$;
    Updating the Vague-filler $V$ with loss $L_V$;
    Pasting back vague $\tilde{h}_i \leftarrow y \odot m + h_i \odot (1 - m)$;
    Getting Pseudo-Healthy inpainted vague $y \leftarrow G(\tilde{h}_i, m)$;
    Pasting back $\tilde{h}_i \leftarrow y \odot m + h_i \odot (1 - m)$;
    Updating the Generator $G$ with loss $L_G$;
end while
Return Trained Vague-filler model $V$, Generator model $G$, Discriminator model $D$;
```

Algorithm 1. Training of our approach.

producing a pseudo-healthy image with dimensions identical to those of the input image. Detailed descriptions of this training process are provided in Algorithm 1.

3.7 Inference pipeline

The inference process is detailed within the "Inference Pipeline" section, as outlined in Figure 3. During inference, the grayscale values of all images are linearly adjusted to range from $[-1, 1]$. Masks are used to indicate pathological regions with a value of 1 and background areas with a value of 0. In this context, p_i represents the input pathological image, while $\tilde{h}_i$ signifies the generated healthy image post vague filling. The operation $\odot$ stands for element-wise multiplication. The Vague-Filler V processes the concatenated masked image and mask as input, forecasting a vague filled image vague $y = V(h_i, m)$ that retains the dimensions of the input image. Subsequently, the Generator G utilizes the combined vague image and mask to generate $y = G(\text{vague}(\tilde{h}_i), m)$, resulting in the pseudo-healthy image $\tilde{h}_i$. This process is comprehensively described in Algorithm 2.

```
Input: Test data $T$, Vague-filler model $V$, Generator model $G$, Segmentor model $S$, Vague-filler model weights $V_{weights}$, Generator model weights $G_{weights}$, Segmentor model weights $S_{weights}$
Output: Pseudo-healthy images $\tilde{h}_i$
Initialization;
Load weights for Vague-filler model $V$ from $V_{weights}$;
Load weights for Generator model $G$ from $G_{weights}$;
Load weights for Segmentor model $S$ from $S_{weights}$;
Sampling batch images $p_i$ from test data $T$;
Getting masks $m \leftarrow S(p_i)$;
Getting inpainted vague $y \leftarrow V(p_i, m)$;
Pasting back vague $\tilde{x}_{h_i} \leftarrow y \odot m + p_i \odot (1 - m)$;
Getting Pseudo-Healthy inpainted vague $y \leftarrow G(\tilde{x}_{h_i}, m)$;
Pasting back $\tilde{h}_i \leftarrow y \odot m + p_i \odot (1 - m)$;
Return $\tilde{h}_i$;
```

Algorithm 2. Inferencing of our approach.

4 Experiments

4.1 Datasets

The proposed model was rigorously evaluated using the T1 and T2 modalities of the BraTS2021 dataset, demonstrating effectiveness across the T1ce and FLAIR modalities as well. The model's versatility was further assessed by examining its adaptability to viral lesions in other organs with the COVID-19-20 dataset.

BraTS2021 Dataset (Menze et al., 2015; Bakas et al., 2017; Baid et al., 2021): The BraTS2021 Dataset (Brain Tumor Segmentation Challenge 2021 Dataset) is designed for the task of medical image segmentation, specifically aimed at evaluating and advancing algorithms for brain tumor segmentation. It comprises MRI scans of the brain collected from multiple medical centers. Each case in the dataset includes four different MRI modalities: T1-weighted, T2-weighted, T1-weighted with contrast enhancement (T1ce), and Fluid-Attenuated Inversion Recovery (FLAIR), along with corresponding ground truth tumor segmentation. Comprising 1,251 cases in the training set, 219 in the validation set, and 530 in the test set, the BraTS2021 dataset ensures comprehensive evaluation. All cases are skull-stripped, resampled to an isotropic resolution of 1 mm^s, and co-registered. Each volume presents four modalities: T1, T2, T1ce, and FLAIR, measured at dimensions of $240 \times 240 \times 155$ (L×W×H).

COVID-19-20 Dataset (Roth et al., 2022): The COVID-19-20 challenge facilitates the evaluation of innovative techniques for segmenting and quantifying lung lesions induced by SARS-CoV-2 through CT images. Drawn from multiple institutions across various countries, these images depict a diverse cohort in terms of age, gender, and disease severity. The dataset includes 199 training images and 50 validation images, each with a resolution of 512–512 pixels. Notably, these images detail lung lesions caused by SARS-CoV-2 and include ground truth annotations derived from non-contrast chest CT scans with confirmed positive RT-PCR results.

4.2 Implementation details and baseline comparisons

Environment: Windows 11, CUDA 11.7.

Framework: The methodology is implemented using the PyTorch framework.

Optimizer: Model training is facilitated using the Adam optimizer.

Learning rate: The initial learning rate is set at 0.001 and reduces by 50% every 5 epochs.

Batch size: Given the slice dimensions of the BraTS2021 dataset at 240 × 240 and those of the COVID-19-20 dataset at 512 × 512, batch sizes are accordingly adjusted. A batch size of 16 is employed for the BraTS2021 dataset, while a smaller batch size of 4 is utilized for the COVID-19-20 dataset.

Training hardware: The model is trained on an NVIDIA GeForce 4080 Super 16GB GPU.

Compared methods: The effectiveness of the proposed method is assessed against three pathologically-informed pseudo-healthy synthesis approaches [GVS-GAN (Zhang et al., 2022), PHS-GAN (Xia et al., 2020), and VA-GAN (Baumgartner et al., 2018)] and two widely-used generative adversarial models [AAE (Makhzani et al., 2016) and Cycle GAN (Zhu et al., 2020)].

Code sources: For the implementation, official codebases are used for GVS-GAN, VA-GAN, and PHS-GAN, while the most popular GitHub repositories are sourced for AAE and Cycle GAN.

Data processing: For the BraTS2021 dataset, we extracted one slice every five slices, resulting in a total of 13,759 slices. For the COVID-19-20 dataset, we filtered the slices to include only those with clearly visible lungs, extracting one slice every two slices, which yielded a total of 2,965 slices.

4.3 Structure healthiness

In certain instances, significant deformations are often observed in pathological images, particularly when large lesion areas are present. Figure 5 illustrates pseudo-healthy images and their corresponding Canny edge maps synthesized under such conditions. Notably, it is common for models to still generate pseudo-healthy images with deformations. To address this issue, Xia et al. (2020) suggested the use of a classifier to categorize Canny edge maps of both healthy and lesioned images to evaluate the presence of deformations. Despite this approach, our statistical analysis of 13,759 pathological slices revealed that only 1,059 slices presented large lesions, where the lesion area exceeded 20% of the total brain area. This indicates that large lesions are relatively rare among pathological slices. Therefore, solely classifying Canny edge maps of healthy and lesioned images does not provide a reliable assessment of a model's deformation correction capability in cases with extensive lesions.

Building on the methodology, we introduce the concept of "Structural Health" (SH) to more accurately explore models' abilities to correct deformations in images with substantial lesion areas. We specifically employed the BraTS21 dataset for this purpose, analyzing Canny edge maps of medical images both with and without extensive lesions. A binary classifier, trained on the VGG network, was utilized. This classifier demonstrated a high level of performance, achieving an average accuracy of 91.2% during its pre-training phase, which underscores its efficacy in detecting deformations in images. The classifier's output, a continuous value ranging from 0 to 1, indicates the likelihood of an image being free from deformations. During the evaluation phase, we focused exclusively on pseudo-healthy images generated from samples with extensive lesions. Here, SH is quantified as the average probability that these images maintain structural integrity and are free from deformations.

$$SH = E_{x_p \sim P}\left[C_p\left(GeN(x_p)\right)\right] \tag{11}$$

In this formula, x_p denotes the pathological image, C_p represents the pre-trained edge map classification model, and GeN indicates the pseudo-healthy synthesis network.

4.4 Other metrics

4.4.1 Healthiness

To evaluate the "healthiness" of pseudo-healthy images, Xia et al. (2020) developed a metric named "healthiness." This metric utilizes a pre-trained segmentation model, which is further refined on a validation set. The fundamental role of this segmenter is to identify pathological regions within both generated pseudo-healthy images and their original pathological counterparts. The healthiness metric is quantified by the proportion of matching pathological pixels found in these images, where a higher percentage indicates a more extensive presence of pathological regions, thus denoting a lower healthiness. The healthiness index (HEALTHINESS, H) is calculated using the following formula:

$$H = 1 - \frac{\mathbb{E}_{x_p \sim P}[N(f_p(\mathrm{GeN}(x_p)))]}{\mathbb{E}_{m_p \sim P_m}[N(f_p(x_p))]}$$

Here, x_p represents the pathological image, f_p is the pre-trained segmentation model, $N(\cdot)$ denotes the number of pixels identified as pathological by f_p, and GeN refers to the pseudo-healthy generation network. The denominator incorporates the segmentation output $f_p(x_p)$ of the pathological image rather than the actual mask m_p, to counter potential biases from the pre-trained model. Subtracting this term from 1 indicates that a reduction in the pathological mask correlates with an increase in H, signifying enhanced healthiness.

4.4.2 Identity preservation

The metric for Identity Preservation quantifies the degree to which the generated pseudo-healthy images maintain the subject's identity (Zhang et al., 2022), specifically assessing the likelihood that both the synthesized pseudo-healthy image and the input pathological image are derived from the same subject. This metric evaluates the structural similarity and peak signal-to-noise ratio of non-pathological regions between the pseudo-healthy image and its corresponding pathological counterpart. The calculations are based on the following formulas:

$$MP = PSNR\Big[(1 - y_t) \odot G\big(x_p\big), (1 - y_t) \odot x_p\Big] \tag{12}$$

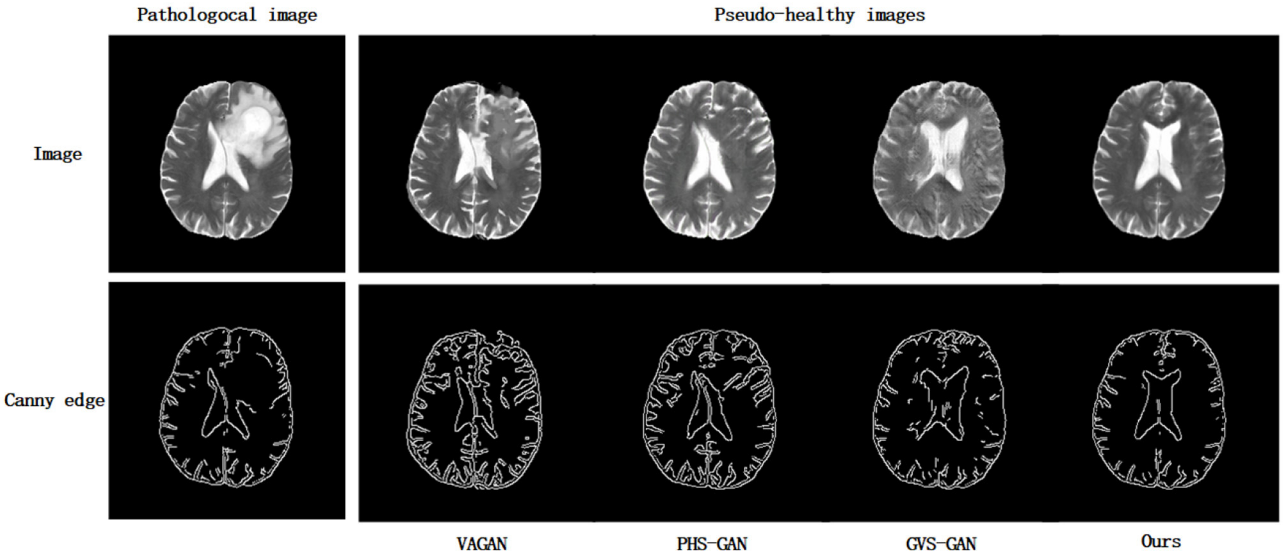


FIGURE 5
The pseudo-healthy synthetic images generated by VAGAN, PHS-GAN, GVS-GAN, and the method proposed in this paper, alongside their corresponding edge maps that display the anatomical structures.

$$MS = SSIM\left[(1 - y_t) \odot G(x_p), (1 - y_t) \odot x_p\right] \tag{13}$$

Where x_p represents the pathological image, y_t denotes the corresponding pathological mask, $\odot$ signifies element-wise multiplication, and PSNR and SSIM are abbreviations for Peak Signal-to-Noise Ratio and Multi-Scale Structural Similarity Index, respectively.

4.5 Evaluation of healthiness and identity preservation

We conducted a thorough evaluation of our proposed method alongside five other models, examining them across four essential dimensions: Healthiness (H), Mask Peak Signal-to-Noise Ratio (MPSNR), Mask Structural Similarity Index Measure (MSSIM), and Structural Healthiness (SH). The outcomes for the T1 modality are detailed in Table 1, and those for the T2 modality appear in Table 2. Under the T1 modality, the AAE model achieved the highest Healthiness score and maintained strong performance in the T2 modality. This superior performance is primarily due to the blurriness of the images it generated, which impacts the segmentor's ability to accurately locate lesion regions, thus resulting in higher health metrics. On the other hand, the PHS-GAN and GVS-GAN models, tailored specifically for brain medical imaging, significantly outshine the other models in both health and subject identity metrics. However, their heavy reliance on the segmentor for identifying lesion regions slightly compromises subject identity preservation. The AAE, VAGAN, and CycleGAN models exhibit a noticeable deficiency in preserving subject identity compared to other models, as they do not incorporate pixel-level pathological annotations, leading to less precise lesion region localization. Our method, which accurately replaces the pathological region and its adjacent 5mm area, nearly flawlessly preserves subject identity. Additionally, extensive training with healthy brain medical images allows the pseudo-healthy brain images synthesized by our method to be more coherent, ensuring a superior Healthiness score.

4.6 Evaluation of visual quality

We conducted experiments comparing our model against five baseline models, assessing their performance in synthesizing pseudo-healthy images under the T1 and T2 modalities, as depicted in Figure 6. Each method's efficacy was evaluated based on subject identity and healthiness. Healthiness is assessed by how well pathological and normal regions integrate in the synthesized images. Images where pathological regions blend seamlessly with normal areas are considered "healthy," while those where pathological areas are distinctly separate are regarded as "unhealthy." Our findings indicate that images generated by the AAE model often do not maintain the subject identity of the input images and appear notably blurred. The VAGAN-produced images can reconstruct lesion regions to a degree, but the quality of reconstruction is poor, and the inaccurate localization of lesions leads to a loss of subject identity. PHS-GAN, similar to CycleGAN, and CycleGAN itself both face challenges in preserving subject identity while repairing extensive lesion regions, resulting in some images losing subject identity and having less coherent repaired organ structures. The performance of GVS-GAN relies heavily on the segmentor's accuracy during the generation process, with errors leading to the creation of lesion-free but blurred tissue structures in the synthesized images. In contrast, our method effectively preserves subject identity in pseudo-healthy images by specifically replacing

TABLE 1 Quantitative comparison of health and identity preservation metrics for AAE, CycleGAN, VAGAN, PHS-GAN, GVS-GAN, and the proposed method under the T1 modality.

Models	Healthiness	MPSNR	MSSIM	Structure healthiness
AAE (Makhzani et al., 2016)	**0.968**	20.64	0.795	0.702
CycleGAN (Zhu et al., 2020)	0.701	31.63	0.968	0.294
VAGAN (Baumgartner et al., 2018)	0.721	21.50	0.899	0.422
PHS-GAN (Xia et al., 2020)	0.831	32.18	0.987	0.580
GVS-GAN (Zhang et al., 2022)	0.909	33.32	0.993	0.749
Ours	0.929	**34.92**	**0.995**	**0.843**

Bold indicates the highest values.

TABLE 2 Quantitative comparison of health and identity preservation metrics for AAE, CycleGAN, VAGAN, PHS-GAN, GVS-GAN, and the proposed method under the T2 modality.

Models	Healthiness	MPSNR	MSSIM	Structure healthiness
AAE (Makhzani et al., 2016)	0.849	21.93	0.775	0.733
CycleGAN (Zhu et al., 2020)	0.744	32.98	0.964	0.496
VAGAN (Baumgartner et al., 2018)	0.783	22.47	0.898	0.499
PHS-GAN (Xia et al., 2020)	0.887	32.55	0.977	0.621
GVS-GAN (Zhang et al., 2022)	0.945	33.11	0.984	0.589
Ours	**0.957**	**33.65**	**0.992**	**0.749**

Bold indicates the highest values.

pathological regions. Enhanced by a context residual mechanism, the synthesized images exhibit a balanced tissue structure distribution, clear visual quality, and consistent preservation of subject identity.

4.7 Other modalities

We conducted comparative experiments focusing on the T1ce and Flair modalities, alongside the PHS-GAN and GVS-GAN models, which previously showed promising results in T1 and T2 modalities. As depicted in Figure 7, the qualitative analysis reveals that images from all three models exhibit a degree of blurring in the T1ce modality. Our model, however, demonstrates superior performance in lesion repair and restoration of brain structures compared to PHS-GAN and GVS-GAN. In the Flair modality, both PHS-GAN and our model show areas of high signal intensity, with PHS-GAN's high signal areas extending throughout the brain. Meanwhile, the images generated by GVS-GAN display no significant high signal areas but fall short in restoring brain structures effectively.

The quantitative results, as presented in Table 3, show that our approach significantly surpasses the other methods in the T1ce modality. In the Flair modality, while GVS-GAN excels in terms of healthiness, our method outperforms in other significant metrics. Overall, the qualitative and quantitative outcomes underscore our method's comparative advantage in both T1ce and Flair modalities over competing approaches, affirming its efficacy in producing more accurate and clinically relevant pseudo-healthy images.

4.8 COVID-19-20 dataset

Our method was applied to the COVID-19-20 dataset to generate pseudo-healthy images, specifically targeting viral lesions. Despite this, the challenges inherent in COVID-19 segmentation and the complex nature of pneumonia cases mean that pixel-level annotations are not sufficiently precise. Consequently, there is a noticeable disparity between the synthesized pseudo-healthy images and actual healthy images. As shown in Figure 8, our approach achieves some success in cases with small-scale lesions and relatively straightforward backgrounds. However, in scenarios involving extensive lung lesions, the synthesized images significantly diverge from true healthy lung images, highlighting the limitations in current segmentation and synthesis techniques in handling complex clinical scenarios.

4.9 Ablation study

To evaluate the effectiveness of the proposed flip symmetry, we conducted both qualitative and quantitative analyses on three variations of GAN networks within the T1 modality: the standard GAN, GAN with Contextual Residual Attention (GAN+CRA), and GAN with Contextual Residual Attention plus Flip Symmetry Network (GAN+CRA+FLIP). The qualitative results are illustrated in Figure 9, and the quantitative outcomes are detailed in Table 4. The findings demonstrate that networks equipped with Contextual Residual Attention significantly surpass the basic GAN in both quantitative and qualitative evaluations. Furthermore, from a qualitative standpoint,

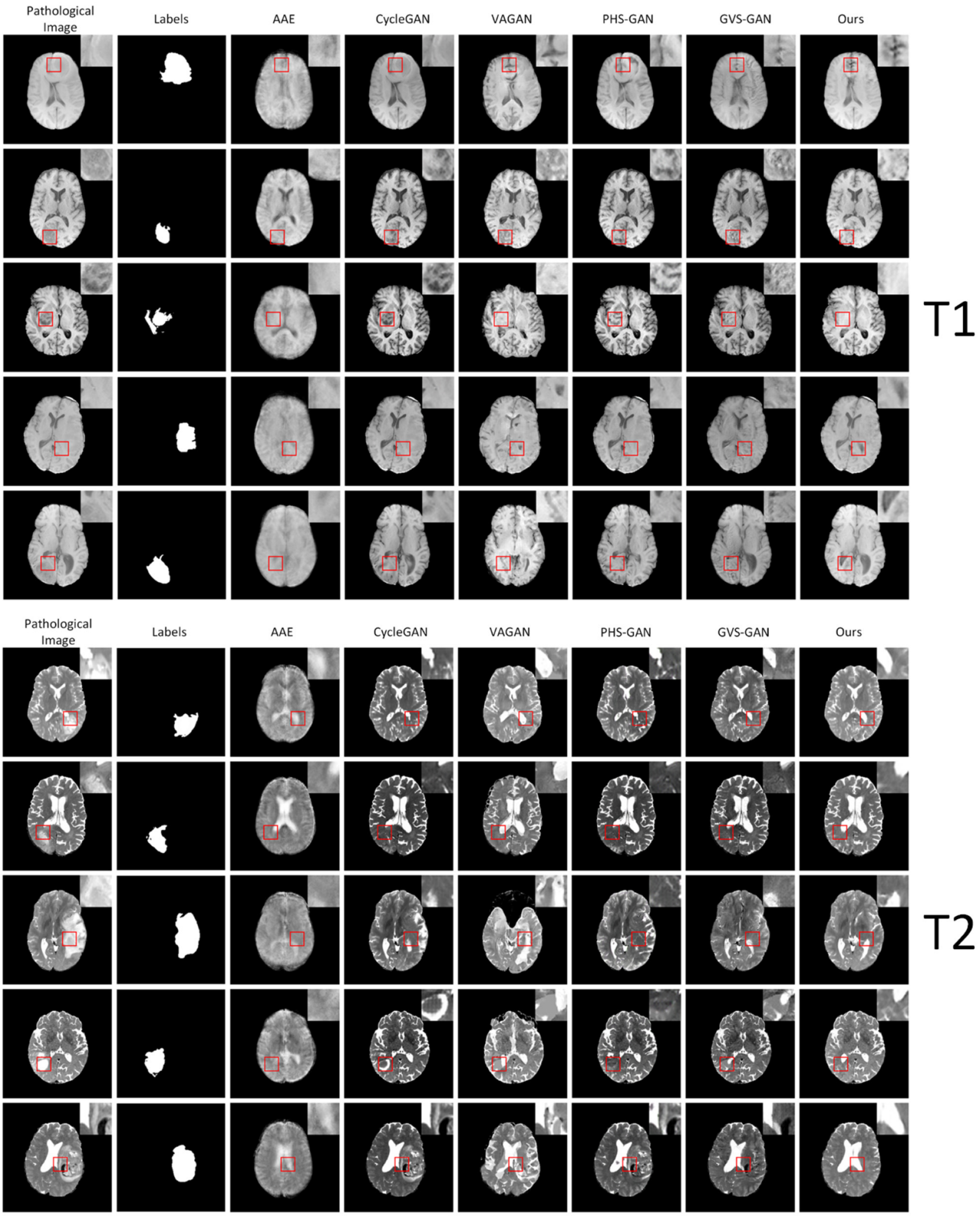


FIGURE 6
Shows experimental results on five samples (one per row) from the BraTS2021 dataset under the T1 and T2 modalities. The columns, from left to right, display the original pathological images, followed by the synthesized pseudo-healthy images generated by AAE, CycleGAN, VAGAN, PHS-GAN, GVS-GAN, and the proposed method.

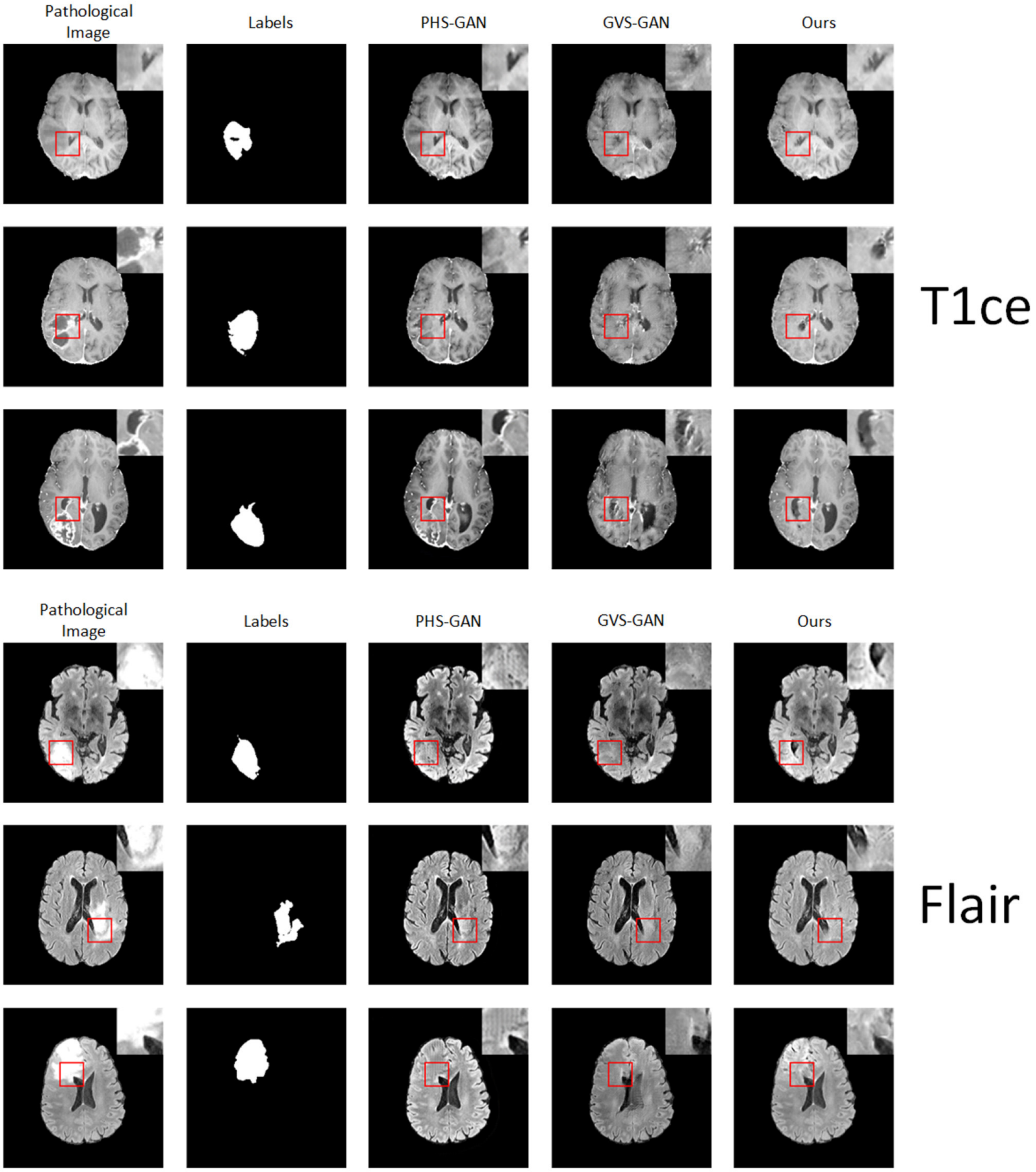


FIGURE 7
Experimental results of five samples each for T1 and T2 modalities on the BraTS dataset: original pathological images and pseudo-healthy images synthesized by PHS-GAN, GVS-GAN, and our method (from left to right).

networks incorporating the flip structure produce pseudo-healthy images that exhibit greater symmetry compared to those without the flip structure. This difference in STRUCTURAL HEALTHINESS confirms that the images generated by networks with the flip structure align more closely with established health standards.

5 Conclusion

We have introduced a novel pseudo-healthy synthesis method that utilizes an inpainting approach to generate images for intracranial infections. Unlike previous methods, our approach prioritizes the visual quality of the synthesized images. It consists

TABLE 3 **Quantitative comparison of health and identity preservation metrics for PHS-GAN, GVS-GAN, and the proposed method under the T1ce modality and flair modality.**

Modal	Models	Healthiness	MPSNR	MSSIM	Structure healthiness
T1ce	**PHS-GAN** (Xia et al., 2020)	0.831	31.78	0.979	0.571
	GVS-GAN (Zhang et al., 2022)	0.909	33.13	0.989	0.529
	Ours	**0.935**	**37.53**	**0.994**	**0.679**
FLAIR	**PHS-GAN** (Xia et al., 2020)	0.731	31.27	0.966	0.552
	GVS-GAN (Zhang et al., 2022)	**0.912**	32.16	0.979	0.602
	Ours	0.891	**35.58**	**0.992**	**0.605**

Bold indicates the highest values.

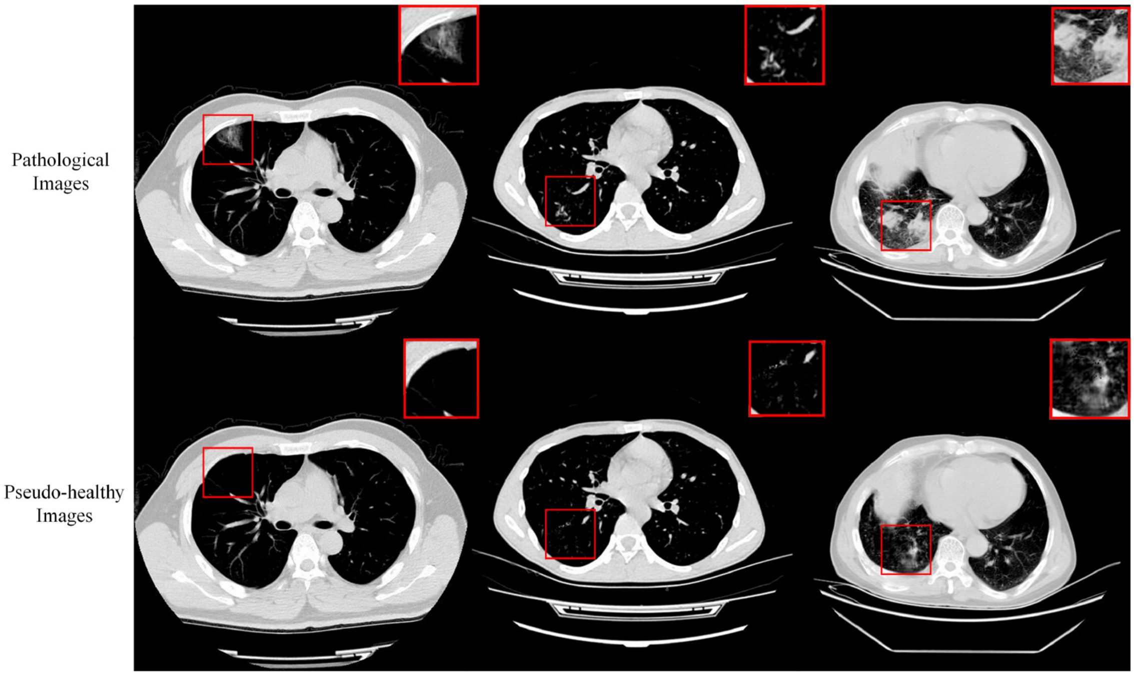


FIGURE 8
Displays the pseudo-healthy images generated by the proposed method on the COVID-19-20 dataset, with three samples shown, one per column.

of three components: a Segmentor, a Vague-Filler, and a Generator. The Segmentor identifies and localizes pathological regions, the Vague-Filler constructs inpainted pseudo-healthy images, and the Generator refines the reconstructions of the pathological input images. We have also established numerical evaluation metrics to assess the anatomical structure quality of the synthesized images. Demonstrated on the BraTS2021 dataset, our method exceeds current state-of-the-art benchmarks in qualitative, quantitative, and subjective evaluations.

Looking ahead, several promising research directions emerge from our work and the broader field. Our method effectively patches lesion regions, enhancing the preservation of subject identity. Post-patching, the Generator leverages global information, allowing the synthesized pseudo-healthy regions to integrate more seamlessly with adjacent areas. Although our results are impressive, our approach is limited by the need for dense, accurate segmentation annotations, which are challenging to amass in clinical settings. Future research should aim to reduce the reliance on precise pixel-level annotations, possibly through more sophisticated segmentation models or unsupervised learning techniques (Ma et al., 2024). Additionally, we have proposed a method to repair regions surrounding lesions to counteract pathologies beyond the lesion areas, though further refinement is needed for more accurate synthesis. Our method also shows limitations in synthesizing pseudo-healthy images of other organs (Liu et al., 2022), prompting future efforts to integrate more advanced localization techniques for a broader application of pseudo-healthy synthesis.

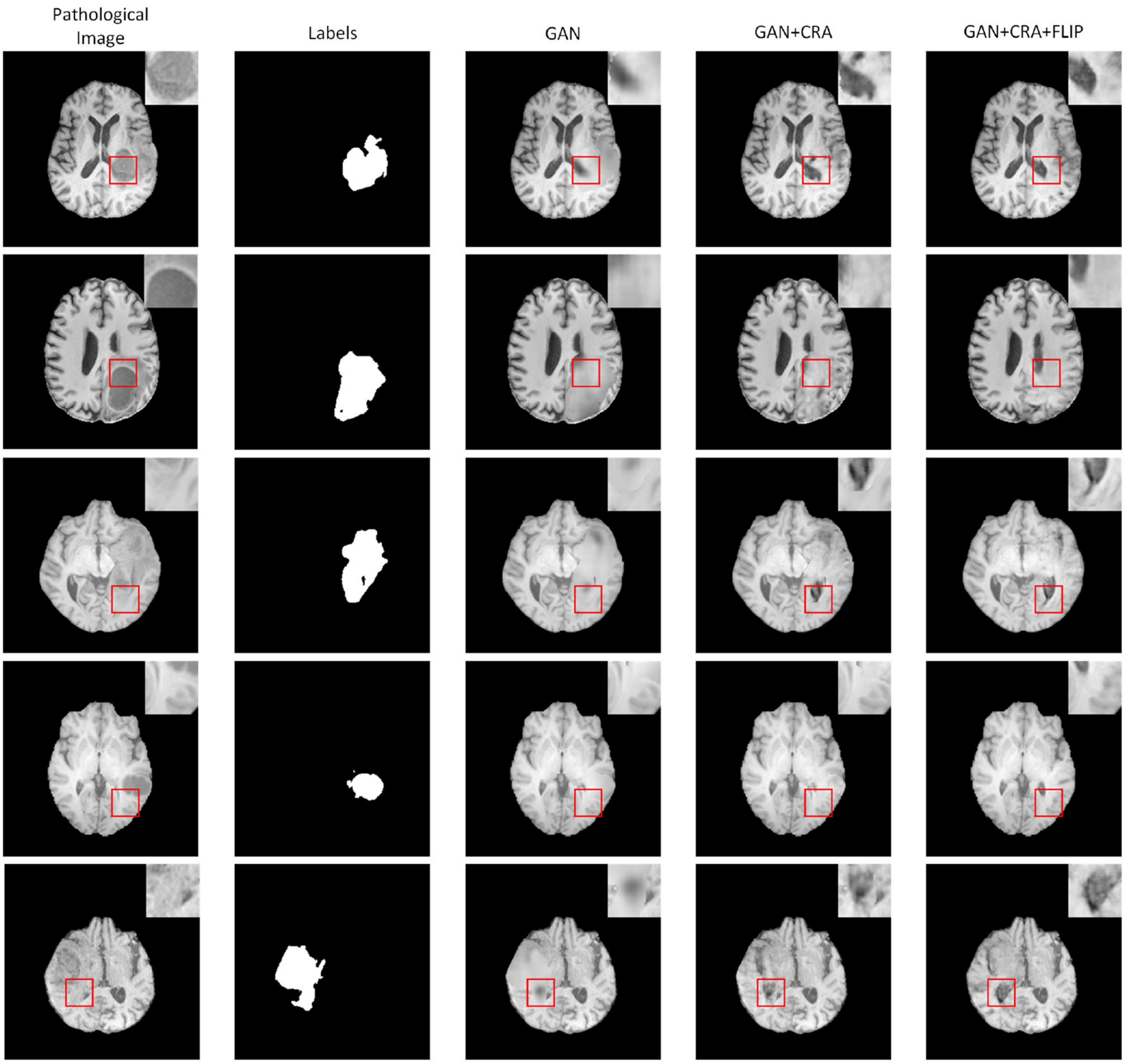


FIGURE 9
Comparison of pseudo-healthy images generated on the BraTS dataset: basic GAN, GAN with contextual residual attention, and GAN with contextual residual attention plus flipped symmetrical network across five samples (one sample per column).

TABLE 4 **Quantitative results for basic GAN networks, GAN networks with contextual residual attention mechanism, and GAN networks featuring both contextual residual attention mechanism and mirrored symmetry network.**

Models	Healthiness	MPSNR	MSSIM	Structure healthiness
GAN	0.591	32.72	0.962	0.245
GAN + CRA	0.882	34.03	0.989	0.660
GAN + CRA + FLIP	**0.929**	**34.92**	**0.995**	**0.843**

Bold indicates the highest values.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

XL: Funding acquisition, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. CX: Methodology, Software, Visualization, Writing – original draft. LL: Funding acquisition, Writing – review & editing. CL: Writing – review & editing. HX: Writing – review & editing. ZL: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study has received funding from the Youth Project of Science and Technology Research Program of Chongqing Education Commission of China (Grant Nos. KJQN202301145 and KJQN202301162), the Technology Campus Teaching Reform Foundation of Chongqing University of Technology (Grant No. 0121249254), and the Scientific Research Foundation of Chongqing University of Technology (Grant Nos. 0121230235 and 0103210650).

Acknowledgments

We would like to thank Shanxiong Chen for constructive criticism of this 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

References

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., et al. (2021). The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314.*

Bakas, S., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., et al. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Sci. Data* 4, 1–13. doi: 10.1038/sdata.2017.117

Baumgartner, C. F., Koch, L. M., Tezcan, K. C., Ang, J. X., and Konukoglu, E. (2018). "Visual feature attribution using wasserstein GANs," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT, USA: IEEE), 8309–8319. doi: 10.1109/CVPR.2018.00867

Baur, C., Wiestler, B., Albarqouni, S., and Navab, N. (2019). "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4* (Springer International Publishing), 161–169. doi: 10.1007/978-3-030-11723-8_16

Baur, C., Wiestler, B., Albarqouni, S., and Navab, N. (2020). Scale-space autoencoders for unsupervised anomaly segmentation in brain MRI. *arXiv:2006.12852 [cs, eess, q-bio].*

Bowles, C., Qin, C., Guerrero, R., Gunn, R., Hammers, A., Dickie, D. A., et al. (2017). Brain lesion segmentation through image synthesis and outlier detection. *NeuroImage* 16, 643–658. doi: 10.1016/j.nicl.2017.09.003

Bowles, C., Qin, C., Ledig, C., Guerrero, R., Gunn, R., Hammers, A., et al. (2016). "Pseudo-healthy image synthesis for white matter lesion segmentation," in *Simulation and Synthesis in Medical Imaging*, eds. S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince (Cham: Springer International Publishing), 87–96. doi: 10.1007/978-3-319-46630-9_9

Chen, X., and Konukoglu, E. (2018). Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. *arXiv:1806.04972 [cs].*

Deng, F., and Gaillard, F. (2014). "Cerebral toxoplasmosis," in *Radiopaedia.org.* doi: 10.53347/rID-30082

Foerster, B., Thurnher, M., Malani, P., Petrou, M., Carets-Zumelzu, F., and Sundgren, P. (2007). Intracranial infections: clinical and imaging characteristics. *Acta Radiol.* 48, 875–893. doi: 10.1080/02841850701477728

Li, Y., Sixou, B., and Peyrin, F. (2021). A review of the deep learning methods for medical images super resolution problems. *IRBM* 42, 120–133. doi: 10.1016/j.irbm.2020.08.004

Lim, J. H., and Ye, J. C. (2017). Geometric GAN. *arXiv:1705.02894 [cond-mat, stat].*

Liu, W., Li, C., Xu, N., Jiang, T., Rahaman, M. M., Sun, H., et al. (2022). CVM-Cervix: A hybrid cervical Pap-smear image classification framework using CNN, visual transformer and multilayer perceptron. *Pattern Recognit.* 130:108829. doi: 10.1016/j.patcog.2022.108829

Luque-Paz, D., Revest, M., Eugène, F., Boukthir, S., Dejoies, L., Tattevin, P., et al. (2021). "Ventriculitis: a severe complication of central nervous system infections," in *Open Forum Infectious Diseases* (Oxford University Press US), ofab216. doi: 10.1093/ofid/ofab216

Ma, Y., Chen, S., Xiong, H., Yao, R., Zhang, W., Yuan, J., et al. (2024). LVONet: automatic classification model for large vessel occlusion based on the difference information between left and right hemispheres. *Phys. Med. Biol.* 69:035012. doi: 10.1088/1361-6560/ad1d6a

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2016). Adversarial autoencoders. *arXiv:1511.05644 [cs].*

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2015). The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imag.* 34, 1993–2024. doi: 10.1109/TMI.2014.2377694

Mitchell, B. C., and Dehkharghani, S. (2014). Imaging of intracranial infectious diseases in adults. *Appl. Radiol.* 43:6. doi: 10.37549/AR2052

Rahman Siddiquee, M. M., Shah, J., Wu, T., Chong, C., Schwedt, T. J., Dumkrieger, G., et al. (2024). "Brainomaly: unsupervised neurologic disease detection utilizing unannotated T1-weighted Brain MR images," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (Waikoloa, HI, USA: IEEE), 7558–7567. doi: 10.1109/WACV57701.2024.00740

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. *arXiv:1505.04597 [cs].*

Roth, H. R., Xu, Z., Tor-Díez, C., Jacob, R. S., Zember, J., Molto, J., et al. (2022). Rapid artificial intelligence solutions in a pandemic–the covid-19-20 lung ct lesion segmentation challenge. *Med. Image Analy.* 82:102605. doi: 10.1016/j.media.2022.102605

Sun, L., Wang, J., Huang, Y., Ding, X., Greenspan, H., and Paisley, J. (2020). An adversarial learning approach to medical image synthesis for lesion detection. *IEEE J. Biomed. Health Inform.* 24, 2303–2314. doi: 10.1109/JBHI.2020.2964016

Tao, Y., Ma, X., Zhang, Y., Huang, K., Ji, Z., Fan, W., et al. (2023). LAGAN: lesion-aware generative adversarial networks for edema area segmentation in SD-OCT images. *IEEE J. Biomed. Health Inform*. 27, 2432–2443. doi: 10.1109/JBHI.2023.3252665

Tsunoda, Y., Moribe, M., Orii, H., Kawano, H., and Maeda, H. (2014). "Pseudo-normal image synthesis from chest radiograph database for lung nodule detection," in *Advanced Intelligent Systems* (Springer), 147–155. doi: 10.1007/978-3-319-05500-8_14

Xia, T., Chartsias, A., and Tsaftaris, S. A. (2020). Pseudo-healthy synthesis with pathology disentanglement and adversarial learning. *Med. Image Analy*. 64:101719. doi: 10.1016/j.media.2020.101719

Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., et al. (2018). Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Trans. Med. Imag*. 37, 1348–1357. doi: 10.1109/TMI.2018.2827462

Yi, Z., Tang, Q., Azizi, S., Jang, D., and Xu, Z. (2020). Contextual residual aggregation for ultra high-resolution image inpainting. *arXiv:2005.09704 [cs]*.

Zhang, Y., Lin, X., Zhuang, Y., Sun, L., Huang, Y., Ding, X., et al. (2022). Harmonizing pathological and normal pixels for pseudo-healthy synthesis. *IEEE Trans. Med. Imag*. 41, 2457–2468. doi: 10.1109/TMI.2022.3164095

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2020). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv:1703.10593 [cs]*.

Zimmerman, R. A., and Girard, N. (2004). "Imaging of intracranial infections," in *Infections of the Central Nervous System. 3rd edition*. Philadelphia: Lippincott-Raven, 31–56.

frontiers | Frontiers in Microbiology

TYPE Technology and Code
PUBLISHED 01 October 2024
DOI 10.3389/fmicb.2024.1483084

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Dongmei Ai,
University of Science and Technology Beijing, China
Yinan Guo,
China University of Mining and Technology, Beijing, China
Ying Wei,
Northeastern University, China

*CORRESPONDENCE
Jie Wang
✉ wangjie@cnu.edu.cn

RECEIVED 19 August 2024
ACCEPTED 16 September 2024
PUBLISHED 01 October 2024

CITATION
Peng Y, Liu Y, Liu Y and Wang J (2024) Comprehensive data optimization and risk prediction framework: machine learning methods for inflammatory bowel disease prediction based on the human gut microbiome data.
Front. Microbiol. 15:1483084.
doi: 10.3389/fmicb.2024.1483084

Comprehensive data optimization and risk prediction framework: machine learning methods for inflammatory bowel disease prediction based on the human gut microbiome data

Yan Peng[1], Yue Liu[1], Yifei Liu[2] and Jie Wang[1]*

[1]School of Management, Capital Normal University, Beijing, China, [2]School of Mathematical Sciences, Capital Normal University, Beijing, China

Over the past decade, the prevalence of inflammatory bowel disease (IBD) has significantly increased, making early detection crucial for improving patient survival rates. Medical research suggests that changes in the human gut microbiome are closely linked to IBD onset, playing a critical role in its prediction. However, the current gut microbiome data often exhibit missing values and high dimensionality, posing challenges to the accuracy of predictive algorithms. To address these issues, we proposed the comprehensive data optimization and risk prediction framework (CDORPF), an ensemble learning framework designed to predict IBD risk based on the human gut microbiome, aiding early diagnosis. The framework comprised two main components: data optimization and risk prediction. The data optimization module first employed triple optimization imputation (TOI) to impute missing data while preserving the biological characteristics of the microbiome. It then utilized importance-weighted variational autoencoder (IWVAE) to reduce redundant information from the high-dimensional microbiome data. This process resulted in a complete, low-dimensional representation of the data, laying the foundation for improved algorithm efficiency and accuracy. In the risk prediction module, the optimized data was classified using a random forest (RF) model, and hyperparameters were globally optimized using improved aquila optimizer (IAO), which incorporated multiple strategies. Experimental results on IBD-related gut microbiome datasets showed that the proposed framework achieved classification accuracy, recall, and F1 scores exceeding 0.9, outperforming comparison models and serving as a valuable tool for predicting IBD onset risk.

KEYWORDS

gut microbiome, inflammatory bowel disease, novel risk warning framework, machine learning, data imputation, parameter optimization

1 Introduction

Inflammatory bowel disease (IBD), which includes ulcerative colitis (UC) and Crohn's disease (CD), is a group of chronic inflammatory disorders of the gastrointestinal tract (Flynn and Eisenstein, 2019). IBD is associated with an increased risk of intestinal malignancies (Faye et al., 2022), and it can also lead to complications involving the joints, skin, eyes, and central nervous system (Rogler et al., 2021). Additionally, patients with IBD frequently experience comorbid depression and anxiety (Bisgaard et al., 2022). Although no specific pathogen has been definitively implicated in the etiology of IBD, a growing body of evidence suggests a significant association between the human gut microbiome and the development of IBD (Kostic et al., 2014).

The development of high-throughput sequencing technologies has enabled researchers to capture a comprehensive snapshot of the microbial community of interest (Almeida et al., 2019). Among these, 16S rRNA gene sequencing stands out as an efficient and cost-effective method for identifying and classifying bacteria and archaea within microbial populations (Johnson et al., 2019).

Although new technologies have significantly enhanced our ability to characterize the human gut microbiome and its potential in predicting IBD, several key challenges remain in effectively utilizing these data to construct predictive models. Firstly, due to the high-dimensional sparsity of microbiome data and the limitations of sequencing technologies, data missingness is a prevalent issue. Most current studies employ simple mean imputation, zero-filling methods or K-nearest neighbors (KNN) imputation (Liñares-Blanco et al., 2022), which fail to adequately capture the intrinsic structure and complex relationships within the data, potentially leading to decreased model performance. While multiple imputation by chained equations (MICE) is widely utilized for data imputation, it possesses several constraints. Its performance could be considerably influenced if the data fails to meet the assumption of missing completely at random (MCAR) (Azur et al., 2011). Furthermore, MICE is susceptible to the selection of model parameters (Doove et al., 2014), particularly with non-linear relationships or interactions, which might result in less reliable imputation outcomes. The approach also brings in uncertainty, as the results may vary across different datasets or subsets of the same dataset. Most crucially, MICE is inclined to overfitting when dealing with high-dimensional data (Tang and Ishwaran, 2017). Hence, considering the high-dimensional characteristic of gut microbiome data, MICE might not be the optimal choice for imputation in this context. Secondly, the human gut microbiome involves thousands of genes or microbial features, many of which are irrelevant or noisy, obscuring the relationship between key features and health, leading to overfitting. High-dimensional risk factors increase computation time (Wang et al., 2023a), and complex interrelationships reduce prediction accuracy. In situations with a small sample size, a large number of features can lead to the curse of dimensionality, rendering the data sparse within the feature space.

A review (Armstrong et al., 2022) evaluates dimensionality reduction techniques for microbiome data, including principal component analysis (PCA), non-metric multidimensional scaling (nMDS), t-SNE and UMAP. PCA and nMDS are not suitable for handling sparse data, whereas t-SNE and UMAP, although effective in capturing non-linear patterns, are highly sensitive to parameter settings, making their results less reliable and harder to reproduce. Additionally, some studies have also explored nonlinear techniques such as Variational Autoencoders (VAE) (Rezende et al., 2014), which introduce probabilistic generative models and nonlinear transformations to achieve more representative low-dimensional representations (Li et al., 2020). However, these models encounter challenges such as training instability, slow convergence, and the issue of vanishing gradients when dealing with ultra-high-dimensional datasets. Thirdly, individual machine learning models exhibit variable performance across different datasets, resulting in inconsistent predictions and limited generalization (Ansarullah and Kumar, 2019). While deep learning models can enhance prediction accuracy, they require substantial data, computational resources and face challenges in multi-dimensional data processing (Yekkala et al., 2017). Ensemble learning has yielded promising results across various prediction tasks (Peng et al., 2023; Kalaiselvi and Geetha, 2024). Similarly, it has demonstrated efficacy in predicting the risk of IBD (Alfonso Perez and Castillo, 2023). Studies have shown that random forest (RF), as an ensemble learning method, performs exceptionally well in predicting IBD. One study demonstrated that RF model based on laboratory markers exhibit high accuracy in classifying IBD, particularly achieving AUC values of 97% for Crohn's disease and 91% for ulcerative colitis (Kraszewski et al., 2021). A study developed a RF model using baseline clinical and serological parameters, achieving an AUC of 0.90 to successfully predict CD patients' response to IFX treatment, outperforming a logistic regression model (Li et al., 2021). Moreover, research by Alfonso Perez and Castillo (2023) further confirmed that RF models excel in handling complex medical data, making them an excellent choice for IBD prediction, outperforming many other commonly used machine learning algorithms. In addition, hyperparameter settings significantly impact ensemble learning model accuracy. Traditional hyperparameter optimization methods like random search (RS) and grid search (GS) are computationally intensive (Hutter et al., 2019), while Bayesian optimization (BO) (Chen et al., 2023) and particle swarm optimization (PSO) (Wang et al., 2023b), and Gray Wolf optimization (GWO) (Mafarja and Mirjalili, 2017) can get trapped in local optima. Direct use of aquila optimizer (AO) (Abualigah et al., 2021) also risks local optima issues.

To address these issues, we proposed the CDORPF, a comprehensive data optimization and risk prediction framework. This framework was divided into two main modules: data optimization and risk prediction. In the data optimization module, we first employed triple optimization imputation (TOI) to impute missing data while preserving the biological characteristics of the gut microbiome data. Next, we introduced the importance-weighted variational autoencoder with integrated evaluation (IWVAE) method, which incorporated feature importance ranking and a comprehensive scoring approach based on VAE, to enhance the dimensionality reduction process by retaining critical features. This resulted in a more complete dataset and a low-dimensional representation, laying a solid foundation for improving algorithm efficiency and accuracy. In the risk prediction module, the optimized data was classified using the RF model, while the improved aquila optimizer (IAO), enhanced with multiple strategies, was employed for global hyperparameter optimization of the RF model. The effectiveness of the CDORPF framework had been validated through multiple comparative experiments.

2 Materials and methods

2.1 CDORPF

The overall framework structure of CDORPF is illustrated in Figure 1.

Where, the input is a N^*M matrix, and the output is a N'^*M' matrix. N represents the number of original data items, M refers to the feature dimension of the data items, and N' and M' denote the sample size and feature dimension, respectively, after data optimization.

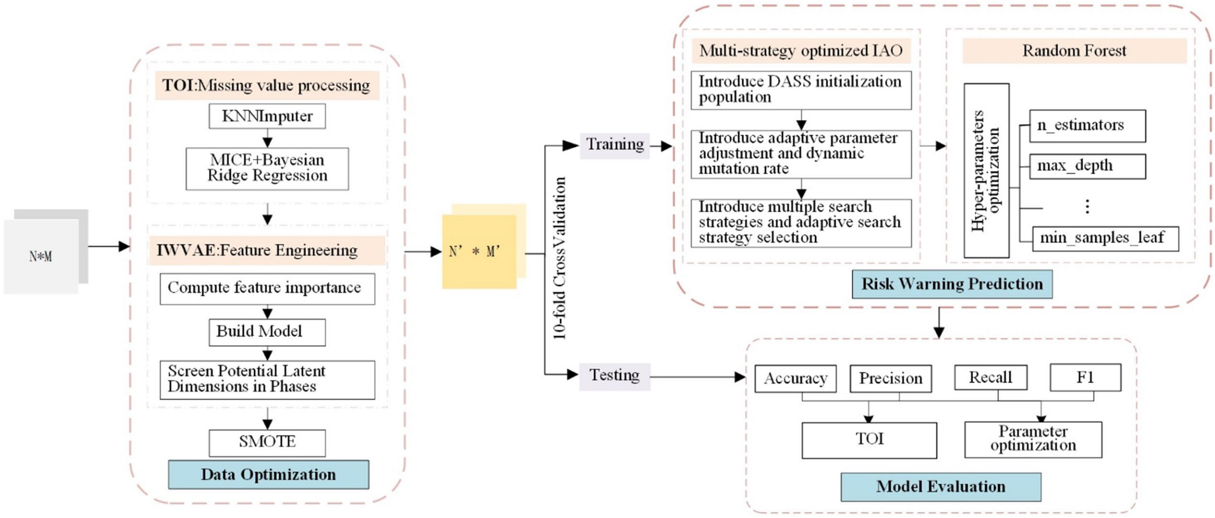


FIGURE 1
Flow work of CDORPF.

The workflow of the model is as follows:

1. Data optimization: TOI is applied to impute missing data in the original dataset. IWVAE is then utilized for dimensionality reduction, and SMOTE (Feng et al., 2023) is employed to address sample imbalance, ensuring consistency and reliability of the human gut microbiome data while maintaining biological accuracy.
2. Predictive model construction: Initial parameter ranges for the RF model are set, including n_estimators, max_depth, min_samples_split and min_samples_leaf. IAO optimizes these parameters through iterative searches. Enhancements such as dynamic adjustment Sobol sequence (DASS), adaptive parameter adjustment, and dynamic mutation rates are introduced. The optimized hyperparameters are used to train the RF model, which is then evaluated on a validation set.
3. Model evaluation: Performance metrics for the predictive model are proposed and compared with classical and widely-used models to assess CDORPF.

2.2 Data imputation based on TOI

Gut microbiome data frequently exhibit high-dimensional sparsity (Xie and Lederer, 2021), and multicollinearity may be present among the features (Die et al., 2022). For sparse data, relying solely on KNN often encounters challenges in identifying sufficiently similar neighboring samples, while using MICE alone may fail to capture strong feature correlations. KNN excels at managing locally similar samples and effectively captures local structural characteristics between them, whereas MICE employs regression models that leverage global feature correlations for imputation. By combining KNN and MICE, both local and global information can be leveraged to more comprehensively fill in missing values. Additionally, introducing ridge regression during MICE imputation can effectively reduce model instability caused by multicollinearity, enhancing the robustness of the model and making the imputation results more stable and accurate. Ridge regression achieves this through regularization, which prevents overfitting to noise inherent in sparse data while preserving reasonable correlations among features.

Based on the above three methods, we proposed TOI. TOI not only preserves the dataset's integrity but also retains the intrinsic structure and relationships within the microbiome data, supporting reliable subsequent analysis and model development.

The process is as follows:

Step 1. Initial imputation: For a dataset X containing missing values, the KNN is employed for preliminary imputation.

1. Distance calculation: The distance between record i and other records in the dataset is computed using the appropriate Equation 1.

$$d(i,n)=\sqrt{\sum_{m=1}^{p}\left(x_{i,m}-x_{n,m}\right)^2} \tag{1}$$

2. Select nearest neighbors: Based on the calculated distances, select the k records that are closest.
3. Calculate mean: Compute the mean of feature j across these k nearest neighbors using Equation 2 and use this value to impute the missing data.

$$\hat{x}_{i,j}=\frac{1}{k}\sum_{n\in N_k(i)} x_{n,j} \tag{2}$$

where, $\hat{x}_{i,j}$ represents the imputed value of feature j for record i, $N_k(i)$ denotes the set of indices corresponding to the k nearest neighbors of record i, $x_{n,j}$ refers to the values of feature j among these nearest neighbors.

Step 2. Iterative optimization of imputation: Building on the initial imputation, multiple imputation is performed with iterative refinement. In each iteration, a Bayesian ridge regression model is used to predict the imputed values.

(1) Target feature selection: In each iteration, select feature *j* as the target variable, with the remaining features serving as predictors.
(2) Model construction: Using the other feature values $x_{i,\neg j}^{(t)}$ of record *i*, construct the Bayesian Ridge Regression model f_j and compute the regression coefficients β, as shown in Equation 3.

$$\hat{\beta} = \left(X^{\mathrm{T}} X + \lambda I\right)^{-1} X^{\mathrm{T}} y \tag{3}$$

(3) Based on the Equation 4, predictions are made to obtain updated imputed values.

$$\hat{x}_{i,j}^{(t+1)} = f_j\left(x_{i,\neg j}^{(t)}\right) + \varepsilon_{i,j}^{(t)} \tag{4}$$

where, $\hat{x}_{i,j}^{(t+1)}$ represents the imputed value of feature *j* for record *i* in the $t+1$ iteration, f_j denotes the regression model for feature *j*, $x_{i,\neg j}^{(t)}$ refers to all feature values of record *i* excluding feature *j* in the *t* iteration, $\varepsilon_{i,j}^{(t)}$ is the residual value.

Step 3. Iterative refinement: Repeat the iterative process, selecting each feature for imputation and continuously optimizing the imputed values until convergence is achieved or the maximum number of iterations is reached.

Step 4. Final imputation results: Obtain the final optimized imputation results, filling in all missing values.

In summary, TOI combines the simplicity of KNN, the iterative refinement capability of MICE, and the regularization strength of Bayesian ridge regression. TOI effectively captures both linear and nonlinear relationships in the data, ensuring data integrity and enhancing model prediction performance for more accurate and stable missing data handling.

2.3 Data dimensionality reduction based on IWVAE

Building on VAE, we propose IWVAE, which integrates feature importance ranking and a comprehensive scoring mechanism to effectively reduce data dimensionality while maintaining high classification performance.

Step 1. Calculate feature importance: Feature importance scores are computed using RF, and features are ranked accordingly. This approach prioritizes the retention of the most critical features for classification tasks, enhancing the efficiency of the dimensionality reduction process.

Step 2. Definition of VAE: The encoder maps high-dimensional data into a low-dimensional latent space, while the decoder reconstructs the high-dimensional data from this latent representation. The encoder outputs include the latent mean μ and the latent log-variance $\log\left(\sigma^2\right)$, with the KL divergence loss term constraining the distribution in the latent space as shown in Equation 5.

$$\mathrm{KL}\left(q(z|x)\|p(z)\right) = -\frac{1}{2}\sum_{j=1}^{J}\left(1 + \log\left(\sigma^2\right) - \mu_j^2 - \sigma_j^2\right) \tag{5}$$

The decoder reconstructs the data using the latent variable *z*, and the reconstruction error is computed based on the following Equation 6:

$$\text{Reconstruction error} = \frac{1}{N}\sum_{i=1}^{N} x_i - \hat{x}_i^{\,2} \tag{6}$$

where, x_i represents the original data, $\hat{x}_i$ represents the reconstructed data, *N* denotes the sample size.

Step 3. Preliminary screening stage: Features are selected at intervals of 1/10 of the total dimensionality. The trained VAE model is used to calculate reconstruction errors, and RF is trained on the dimensionally reduced data. Classification accuracy is evaluated through cross-validation. By balancing reconstruction error and classification accuracy, the introduction of a comprehensive score avoids bias and overfitting, enabling a more thorough model evaluation and ensuring an optimal balance between preserving data features and predictive capability. Both the reconstruction error and classification accuracy are standardized using specific Equations 7, 8, and a comprehensive score is computed using Equation 9. The optimal latent dimensions are then recorded.

$$\mathrm{SRE} = \frac{\text{Reconstruction error} - \min\left(\text{reconstruction error}\right)}{\max\left(\text{Reconstruction error}\right) - \min\left(\text{Reconstruction error}\right)} \tag{7}$$

$$\mathrm{SMO} = \frac{\text{Model accuracy} - \min\left(\text{Model accuracy}\right)}{\max\left(\text{Model accuracy}\right) - \min\left(\text{Model accuracy}\right)} \tag{8}$$

$$\text{Combined score} = \mathrm{SRE} + \left(1 - \mathrm{SMO}\right) \tag{9}$$

Step 4. Refined screening phase: Conduct a more detailed screening around the optimal latent dimensions identified in the preliminary screening phase. All steps from the preliminary screening are repeated within this refined range to ensure precision.

2.4 IAO-RF risk prediction model construction

2.4.1 IAO

The traditional AO initializes the population randomly, which may result in insufficient exploration during the early stages and an increased risk of becoming trapped in local optima. Furthermore, its fixed parameter selection and singular search strategy can cause an imbalance between global exploration and local exploitation. To overcome these limitations, IAO incorporates multiple strategies to enhance AO, achieving a balance between local and global optimization while improving search efficiency. The following optimizations have been implemented:

(1) DASS is employed to initialize the population, enhancing the diversity of the initial population.

The Sobol sequence is a quasi-random sequence used to generate low-discrepancy samples. The DASS refines this by incorporating feature importance information to adjust the search space dynamically, overcoming the limitation of the traditional Sobol sequence, which cannot adapt to problem-specific characteristics.

a Calculate feature importance: The importance of each feature is calculated using a baseline model, as described in Equation 10.

$$\text{importance}_i = \frac{1}{n_{\text{trees}}} \sum_{t=1}^{n_{\text{trees}}} \Delta \text{Gini}_t(i) \tag{10}$$

where, $\Delta \text{Gini}_t(i)$ represents the contribution of feature i to the Gini index in the t-th tree.

b Dynamic adjustment of search space: The search space is adjusted based on feature importance. If a feature's importance exceeds a certain threshold, its search range is expanded as specified in Equation 11.

$$\text{bounds}_j = \begin{cases} \left[x_{\min,j}, x_{\max,high,j}\right] & \text{if importance}_j > \theta \\ \left[x_{\min,j}, x_{\max,j}\right] & \text{if importance}_j \le \theta \end{cases} \tag{11}$$

where, θ is a predefined threshold.

c Sobol sequence generation: The Sobol sequence is used to generate uniformly distributed points, as defined in Equation 12.

$$u = \text{Sobol}(d, N) \tag{12}$$

where, u is a $d \times N$ matrix, d represents the dimensionality of the hyperparameters, N refers the population size.

d Mapping to the dynamically adjusted search space: Using Equation 13, the values from the Sobol sequence in the range of [0, 1] are mapped to the dynamic adjustment range of each parameter.

$$x_{ij} = x_{\min,j} + u_{ij} \cdot \left(x_{\max,j} - x_{\min,j}\right) \tag{13}$$

where, x_{ij} represents the j-th parameter of the i-th individual, $x_{\min,j}$ and $x_{\max,j}$ denote the minimum and maximum ranges of the j-th parameter.

Figure 2 shows the distribution of a 2D initial population of size 200 generated using the DASS, the traditional Sobol sequence, and random generation methods. It is evident that the population generated by the dynamically adjusted Sobol sequence is more uniformly distributed, providing broader coverage of the solution space. Notably, within the parameter ranges of higher feature importance, this method maintains better population diversity, which can enhance the optimization speed and convergence accuracy of the algorithm.

(2) Adaptive parameter adjustment and dynamic mutation rates are introduced to adaptively modify the search range and mutation rates at different stages of the optimization process. This effectively balances global exploration and local exploitation, thereby enhancing overall optimization performance.

a Adaptive parameter adjustment: As shown in the Equation 14, parameters are dynamically adjusted based on the number of iterations. This approach strengthens global exploration in the early iterations and enhances local exploitation in the later stages, preventing premature convergence.

$$\alpha = \frac{\text{max\_iterations} - \text{iteration}}{\text{max\_iterations}} \tag{14}$$

b Dynamic mutation rate: As shown in the Equation 15, the mutation rate is dynamically adjusted based on the number of iterations, enhancing population diversity and preventing premature convergence.

$$\text{mutation\_rate} = 0.1 \times \left(1 - \frac{\text{iteration}}{\text{max\_iterations}}\right) \tag{15}$$

(3) Incorporate a position update strategy, as different strategies can facilitate exploration and exploitation during the optimization process.

a Exploration strategy: The position of individuals is updated using the current best individual (x_{best}) as a reference point. A wide range of movement is achieved through a random factor (rand) and an adaptive parameter (α), expanding the search space and enhancing global exploration capabilities. As shown in the Equation 16:

$$x_i^{\text{new}} = x_i + \alpha \cdot rand \cdot \left(x_{\text{best}} - x_i\right) \tag{16}$$

where, x_i^{new} represents the new position of individual i, x_i is the current position of individual i, α is an adaptive parameter that gradually decreases with the number of iterations, rand is a random number between 0 and 1.

b Exploitation strategy: The position of individuals is updated using the current worst individual (x_{worst}) as a reference point. Local exploitation is achieved through a random factor (rand) and an adaptive parameter (α), enhancing local search capabilities. This allows for finer exploration within the current search region, preventing premature convergence to local optima. As shown in the Equation 17:

$$x_i^{\text{new}} = x_i + \alpha \cdot rand \cdot \left(x_{\text{worst}} - x_i\right) \tag{17}$$

c Levy flight: Individual x_i approaches the current best individual (x_{best}) using a random step length $\text{Levy}(\beta)$ based on the Levy distribution and an adaptive parameter (α), as shown in Equations 18, 19. This strategy helps to overcome the limitations of local optima and further enhances global exploration capabilities. The formula is as follows:

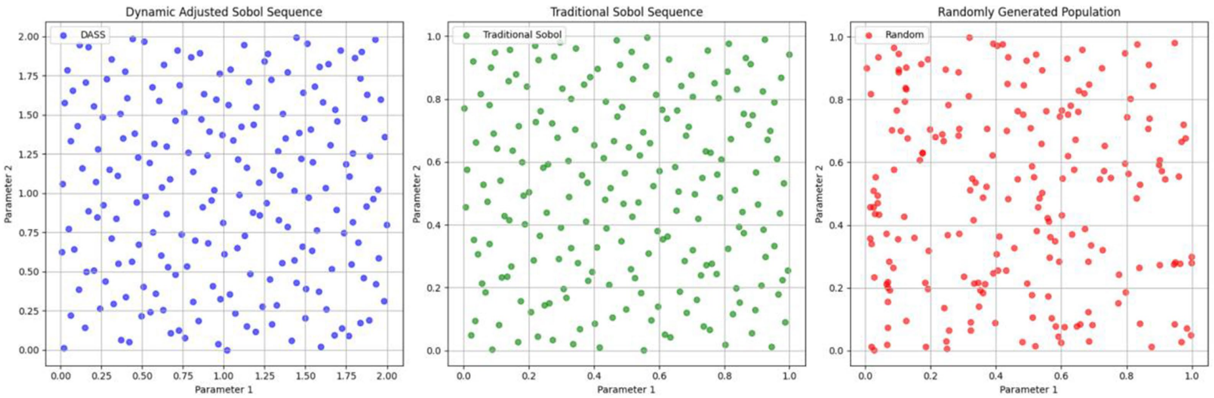


FIGURE 2
Comparison of population initialization.

$$x_i^{\text{new}} = x_i + \alpha \times \text{Levy}(\beta) \times (x_i - x_{\text{best}}) \quad (18)$$

$$\text{Levy}(\beta) = \frac{u}{|v|^{1/\beta}} \quad (19)$$

where, u and v are random variables that follow a normal distribution, and x_{mean} is a parameter of the Levy distribution.

d Gradual convergence strategy: In the middle to late stages of the optimization process, individuals gradually converge toward the population mean, as shown in Equation 20. This approach balances global exploration and local exploitation, leading to a gradual convergence. The formula is as follows:

$$x_i^{\text{new}} = x_i + \alpha \cdot rand \cdot (x_i - x_{\text{mean}}) \quad (20)$$

(4) Introduce diversity measurement and adaptive strategy selection to dynamically adjust optimization strategies, enabling the algorithm to better balance exploring new solution spaces and optimizing the current solution space.

IAO dynamically selects different search strategies at various optimization stages based on population diversity. When diversity is high, the algorithm favors the exploration strategy and Levy flight to expand the search space. Conversely, when diversity is low, it leans toward the exploitation strategy and gradual convergence strategy to optimize the current solution. The standard deviation across each dimension of the population is calculated using Equation 21 to measure population diversity, effectively reflecting the distribution of the population within the search space.

$$D = \frac{1}{n} \sum_{j=1}^{n} \sigma_j \quad (21)$$

where, D represents the population diversity, n is the number of dimensions, and σ_j is the standard deviation of the population in the j-th dimension.

2.4.2 The algorithmic process of IAO-optimized RF

The key steps in optimizing RF with IAO are as follows:

Step 1: Initialize parameters: population size, parameter dimensions (dim), parameter ranges (x_min and x_max) and maximum iterations (max_iterations).

Step 2: Define the fitness function.

Step 3: Initialize the population using DASS.

Step 4: Iterative optimization process:

- Calculate the population mean (x_{mean}) and the worst individual (x_{worst}).
- Calculate adaptive adjustment parameters based on the Equation 14.
- Compute the dynamically adjusted mutation rate using the Equation 15.
- Retain the current best individual (x_{best}) using an elite strategy.
- Execute the search strategy for each individual:

    Select the search strategy based on population diversity.
    Implement exploration, exploitation, Levy and gradual convergence strategies.

- Calculate the fitness of new individuals and update the population.

Step 5: Output the optimal parameters to construct the IAO-RF model for prediction tasks.

2.5 Model performance measures

Using stratified random sampling to maintain class distribution consistency, the dataset was divided into two subsets: 80% for the training dataset and 20% for the test dataset. The training dataset was used to train the machine learning models, and the test dataset was used to evaluate model performance. Ten-fold cross-validation was performed. Based on the test dataset and the model predicted target variables, five statistical measures were used to evaluate the model performance: accuracy, precision, recall, F1-score and AUC.

2.6 Baseline methods

To demonstrate the effectiveness of the proposed model, we compared it with several widely used models across different categories. For evaluating the effectiveness of various imputation methods, we employed classifiers such as logistic regression (LR), SVM, MLP, XGBoost, LightGBM and RF. To assess the performance of dimensionality reduction methods, we compared PCA, VAE and IVAE. For evaluating parameter optimization methods, we compared RS, GS and BO.

3 Results

3.1 Data selection

The experimental data used in this study is from the Inflammatory Bowel Disease Multi'omics Database (IBDMDB) within the Integrated Human Microbiome Project (iHMP) (Lloyd-Price et al., 2019). Microbial community structure and diversity were analyzed using 16S rRNA gene sequencing, specifically targeting the V4 region. This gut microbiome dataset comprises 178 participant records, in which 137 with IBD and 41 without. Each record consists of 983 fields. Notably, one field serves as an indicator for the presence or absence of IBD, while the remaining 982 fields represent an array of microbial features. However, many feature values in most samples are either close to zero or exactly zero, leading to a sparse data distribution in high-dimensional space. Despite this overall sparsity, some samples exhibit high abundance in specific feature dimensions, creating locally dense regions. Additionally, approximately 491 features have missing values, with missing rates ranging from 0.56% to 9.55%, and an average missing rate of 5.27%.

In summary, this dataset is characterized by high-dimensional sparsity, the presence of missing values, differences in class distribution, and local density regions.

3.2 Data optimization

3.2.1 Data imputation based on TOI and effectiveness analysis

3.2.1.1 Validation of the rationale for using TOI

Data missingness can be classified as missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR). To validate the rationale for using the TOI for data imputation, the first step is to identify the type of missing data within the dataset.

3.2.1.1.1 Analysis of missing data types in the dataset

A correlation matrix was used to analyze the types of missing data by displaying the correlations between missing values across different features. The matrix colors range from blue (negative correlation) to red (positive correlation). Blue indicates that when one feature is missing, another is likely to be present, while red suggests that missing values in two features tend to occur together.

As shown in Figure 3, the correlation matrix of missing values in the IBD dataset is predominantly blue, indicating no significant correlation between the missing values of different variables. This suggests that the missing values in this dataset are likely to be MAR, meaning that the missingness of certain variables may be related to other observed variables, but not to the missing data itself. The red areas are mainly along the diagonal, showing that each variable is perfectly correlated with its own missing values, which is expected. Therefore, this figure suggests that most variables have independent missing values, indicating the missingness mechanism in this dataset is likely MAR.

3.2.1.1.2 Analysis of grouped statistical analysis and hypothesis testing results

Grouped statistical analysis provides a statistical comparison between missing and non-missing groups by analyzing their means and standard deviations. Hypothesis testing, through *t*-tests on key variables, evaluates whether significant differences exist between missing values and other variables. Selected experimental results are shown in Table 1.

Most variables show no significant differences between missing and non-missing samples (p-value >0.05), indicating a weak association with missingness. However, some features exhibit significant differences (p-value <0.05), suggesting a potential relationship, while others show near-significant differences (p-value close to 0.05), indicating a possible but not conclusive association. These findings suggest that the majority of missing data in the dataset are not significantly related to other variables, implying a likely MAR.

The combined results of both experiments confirm that the missing data mechanism in the IBD dataset is MAR, supporting the rationale for using TOI for imputation.

3.2.1.2 TOI-based data imputation

3.2.1.2.1 Imputation results of the dataset

The original IBD dataset contains 491 features with missing values, with missing rates ranging from 0.56 to 9.55%. TOI successfully imputed all missing values. A comparison of the dataset before and after imputation is presented in Figure 4.

3.2.1.2.2 Comparison of distributions before and after imputation

By comparing the data distributions before and after TOI imputation, the impact of TOI on the data was visually assessed, validating the effectiveness and fidelity of the imputed data. If the post-imputation distribution aligns with the original data, it indicates that the imputation method is appropriate. Selected experimental results are shown in Figure 5.

Figure 5 illustrates the comparative distributions of variables with substantial missing rates, both prior to and following imputation. Each subplot displays the distribution of an individual variable, where the original data are depicted in green and the imputed data in purple. The histograms illustrate the frequency distribution of the variables, and by contrasting the green and purple histograms, one can distinctly perceive the alterations in data frequency across diverse value ranges before and after imputation. The kernel density estimation (KDE) curves offer a smooth estimation of the probability density, further highlighting the distribution tendencies of the data.

The outcomes imply that the distribution of the variable__ Erysipelotrichaceae_UCG_003 remains highly consistent before and

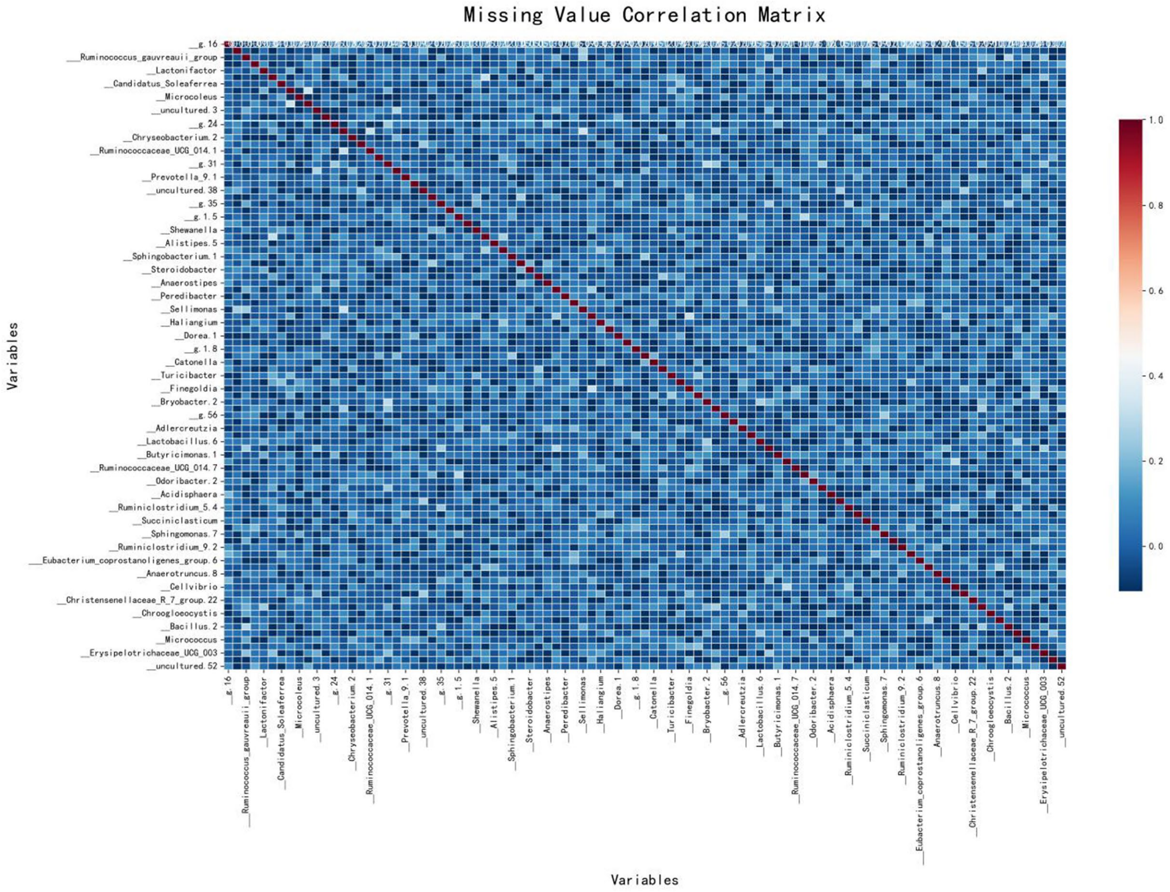


FIGURE 3
Missing value correlation matrix.

TABLE 1 Results of grouped statistical analysis and hypothesis testing.

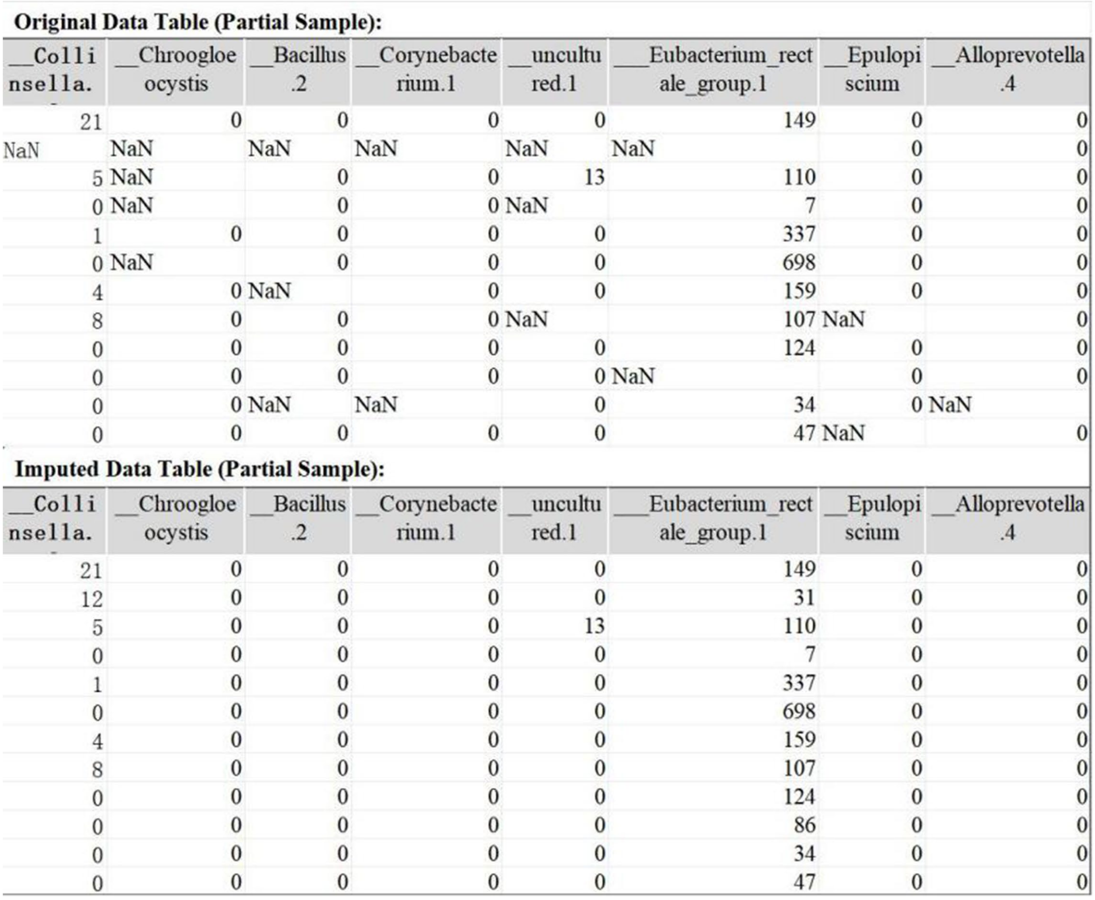
Feature	Variable	Mean (missing)	Mean (non-missing)	*t*-statistic	*p*-value
_Tepidimonas	_Bacteroides	3348.25	2366.95679	−1.136126196	0.257484265
_Tepidimonas	_Bacteroides.6	2147.166667	1618.566265	−0.617270031	0.537854396
_Prevotella	_Bacteroides	1857.416667	2477.388889	0.716181482	0.474850527
_Prevotella	_Faecalibacterium.2	1785.666667	2335.963855	0.737912748	0.461550206
_Prevotella	_Escherichia_Shigella	447.6666667	626.9036145	0.300885428	0.763857015
_Belnapia	_Bacteroides.6	434	1661.096045	0.426897256	0.669975915
_Belnapia	_Dialister.2	318	147.3619632	−0.662256296	0.508748001
_Belnapia	_Lachnoclostridium.1	272	86.81920904	−0.781271529	0.435692463

after imputation, indicating that the imputed data effectively preserves the traits of the original data. For the variable__Defluviitaleaceae_UCG_011.2, the imputed data fills the sparse regions of the original distribution, leading to a more smoother distribution, which attests to the efficacy of the imputation technique. On the whole, the imputed distributions (purple) closely resemble the original distributions (green) for the majority of variables, as evidenced by the similarity in the histograms and KDE curves in each subplot. These findings indicate that our imputation methodology effectively preserves the key distributional properties of the original dataset. For variables exhibiting long-tailed distributions, the imputed data preserve this characteristic, thereby underscoring the efficacy of the imputation technique in addressing sparsity.

3.2.1.3 Evaluation of imputation results validity

To further validate the effectiveness of TOI, comparative experiments were conducted using datasets imputed with KNN, MICE and TOI. Classifier models such as LR, SVM, MLP, XGBoost, LightGBM and RF were selected for analysis. The experimental results are presented in Figure 6.

Figure 6 compares the performance of different imputation methods across various classifier models: red for TOI, green for MICE



Original Data Table (Partial Sample):

__Collinsella.	__Chroogloeocystis	__Bacillus.2	__Corynebacterium.1	__uncultured.1	___Eubacterium_rectale_group.1	__Epulopiscium	__Alloprevotella.4
21	0	0	0	0	149	0	0
NaN	NaN	NaN	NaN	NaN	NaN	0	0
5	NaN	0	0	13	110	0	0
0	NaN	0	0	NaN	7	0	0
1	0	0	0	0	337	0	0
0	NaN	0	0	0	698	0	0
4	0	NaN	0	0	159	0	0
8	0	0	0	NaN	107	NaN	0
0	0	0	0	0	124	0	0
0	0	0	0	0	NaN	0	0
0	0	NaN	NaN	0	34	0	NaN
0	0	0	0	0	47	NaN	0

Imputed Data Table (Partial Sample):

__Collinsella.	__Chroogloeocystis	__Bacillus.2	__Corynebacterium.1	__uncultured.1	___Eubacterium_rectale_group.1	__Epulopiscium	__Alloprevotella.4
21	0	0	0	0	149	0	0
12	0	0	0	0	31	0	0
5	0	0	0	13	110	0	0
0	0	0	0	0	7	0	0
1	0	0	0	0	337	0	0
0	0	0	0	0	698	0	0
4	0	0	0	0	159	0	0
8	0	0	0	0	107	0	0
0	0	0	0	0	124	0	0
0	0	0	0	0	86	0	0
0	0	0	0	0	34	0	0
0	0	0	0	0	47	0	0

FIGURE 4
Example of data tables before and after imputation (partial sample).

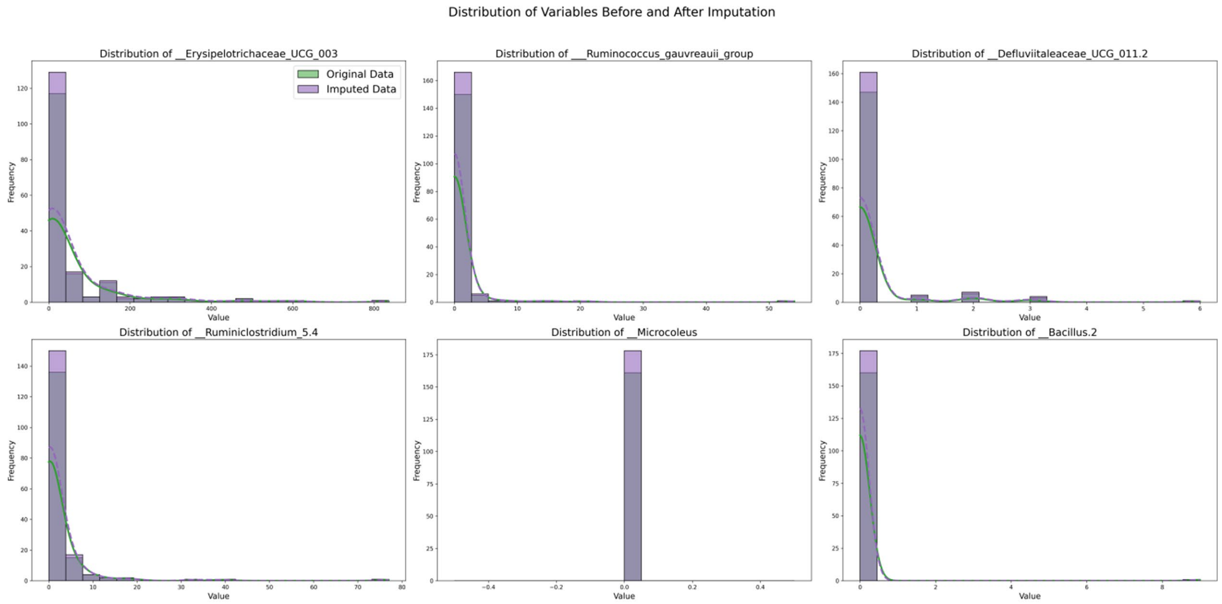


FIGURE 5
Distribution of variables before and after imputation.

and blue for KNN. The results show that the TOI imputation method achieves the highest AUC values in the evaluations of LR, SVM, MLP, LightGBM and RF models, indicating superior classification performance. Although MICE slightly outperforms TOI in the XGBoost, the AUC difference is only 0.0042. Overall, the TOI imputation method provides better classification performance, demonstrating excellent generalization capability and robustness.

Furthermore, when comparing the performance of various imputation methods, the RF model stands out as particularly exceptional across the entire dataset. With the TOI, the RF achieves an AUC of 0.7896, which is higher than that of most other models using the same method. For example, the AUC for the SVM, LR and XGBoost models under the TOI are 0.7561, 0.7326 and 0.7346, respectively. Additionally, even with the MICE and KNN, the RF's AUC values remain highly competitive, outperforming those of most other models. These results indicate that the RF model not only maintains a high level of classification performance when handling imputed data but also demonstrates greater stability across different imputation methods, making it the optimal choice for this dataset.

3.2.2 Data dimensionality reduction based on IWVAE and effectiveness analysis

3.2.2.1 Selection of data dimensions

Figure 7 shows the impact of latent dimensions on reconstruction error (blue curve), model performance (red curve) and the combined score (green curve) during the refinement phase. The blue curve, representing reconstruction error, exhibits fluctuations but generally trends downward as the number of dimensions increases, indicating that data after dimensionality reduction effectively reconstructs the original dataset. The red curve illustrates variations in model performance, with accuracy oscillating between 0.765 and 0.795, suggesting that predictive capability remains relatively stable across different dimensionalities. The green curve, reflecting the combined score, demonstrates significant variability—particularly at higher dimensions—indicating that the dimensionality reduction method is more effective at specific levels. From this figure, it is evident that while both reconstruction error and model performance remain relatively stable as latent dimensions vary, the fluctuations in combined scores imply substantial changes in overall model efficacy across different dimensions. Based on this analysis, 147 dimensions were identified as optimal for achieving an ideal balance between minimizing reconstruction error and enhancing model performance.

3.2.2.2 Analysis of the effectiveness of dimensionality reduction

To further validate the effectiveness of IWVAE, comparative experiments were conducted using the imputed dataset as the experimental dataset. The RF model was applied to the full feature set, as well as to the feature sets reduced by PCA, VAE and IWVAE. The experimental results are presented in Table 2.

As shown in Table 2, IWVAE outperforms all other methods across all metrics, with accuracy, precision, recall and F1 scores all exceeding 0.8. This demonstrates that IWVAE has superior feature extraction capabilities when handling this type of data, significantly enhancing the overall performance of the model.

Following data optimization, the dataset comprises 274 samples with 147 features and no missing values. Of these, 137 samples are from IBD patients, while the remaining 137 are from healthy controls.

3.3 IBD risk prediction based on the IAO-RF

3.3.1 Sensitivity analyses

The primary role of sensitivity analysis is to evaluate the model's response to variations in input parameters and to help identify key

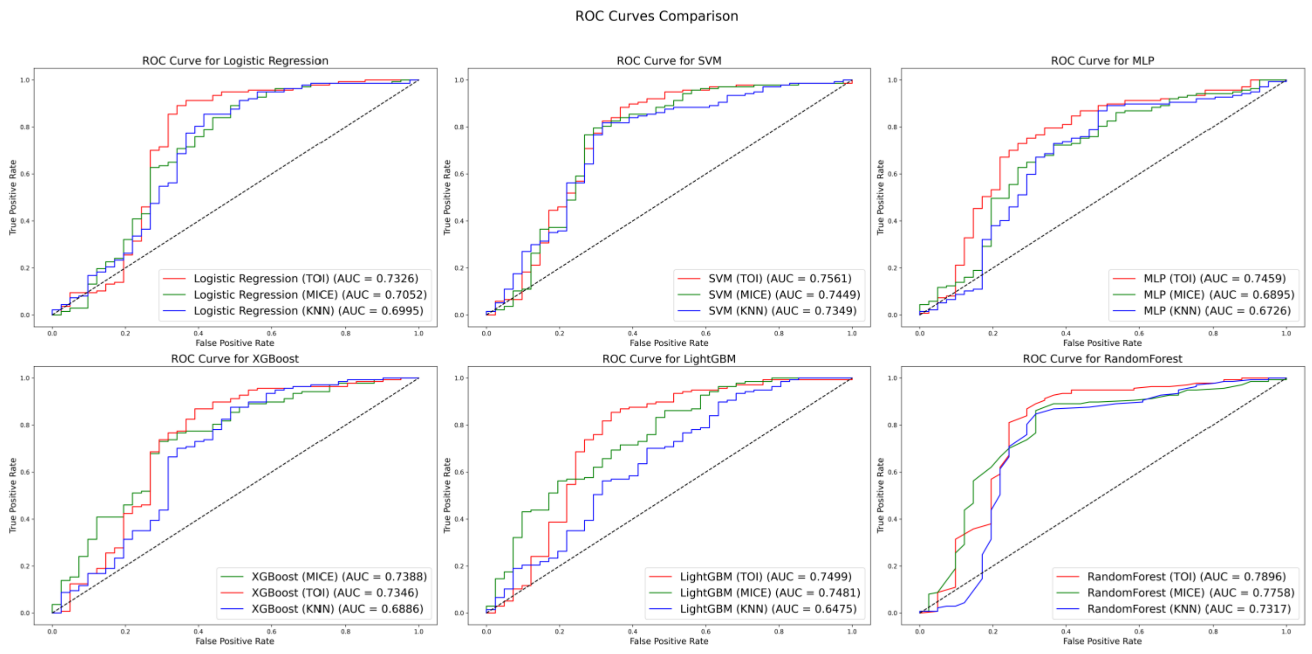


FIGURE 6
ROC curves comparison.

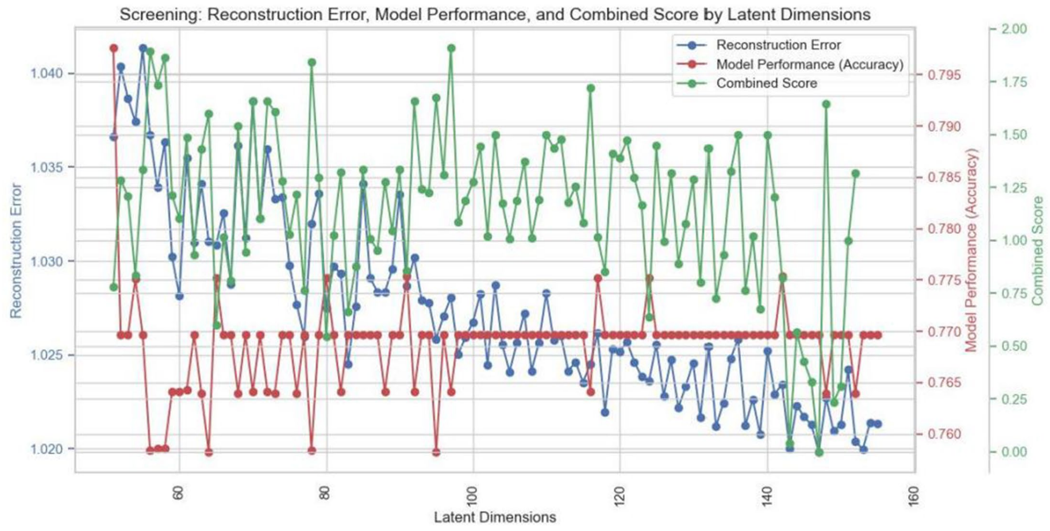


FIGURE 7
Optimal latent dimension selection.

TABLE 2 Results of IBD risk prediction using different dimensionality reduction strategies.

Method	Accuracy	Precision	Recall	F1-score
Full features	0.7863	0.6888	0.7863	0.7098
PCA	0.7980	0.7817	0.7980	0.7711
VAE	0.7696	0.5926	0.7696	0.6695
IWVAE	0.8369	0.8459	0.8369	0.8146

parameters that have the most significant impact on the model's output. Sensitivity analysis allows us to understand how the model performs under different parameter settings, determining its stability and robustness, thereby preventing overfitting or underfitting.

Figure 8 shows the sensitivity analysis of four hyperparameters on the accuracy of the RF model. The top-left plot indicates that increasing the n_estimators from 100 to 120 significantly improves accuracy, which then stabilizes beyond 120 trees, suggesting limited benefits from adding more trees after this point. The top-right plot demonstrates that increasing the max_depth improves accuracy until it plateaus at a depth of 10, implying that deeper trees capture data complexity better, but further increases do not boost accuracy. The bottom-left plot reveals that setting min_samples_split around 5 achieves optimal accuracy, with further increases leading to a decline, indicating that too high a threshold for node splitting can cause underfitting. Lastly, the bottom-right plot indicates that setting min_samples_leaf to 2 maximizes accuracy; further increases in this value negatively impact accuracy, implying that a higher number of samples in leaf nodes may reduce model complexity and predictive power.

In summary, hyperparameter selection is crucial for RF performance. Proper tuning of parameters like n_estimators and max_depth can significantly enhance accuracy, while excessively high values for min_samples_split and min_samples_leaf can hinder performance. Consequently, careful adjustment based on the dataset is essential for optimal results.

3.3.2 Parameter settings

The RF model includes multiple hyperparameters within a large parameter space. In this experiment, IAO was employed to optimize four key hyperparameters: n_estimators, max_depth, min_samples_split and min_samples_leaf. The optimization results are presented in Table 3.

Comparative experiments and model evaluation

To further validate the effectiveness of IAO, a comparative experiment was conducted using the processed dataset. The RF model was optimized using no optimization, RS, GS, BO and IAO. The experimental results are presented in Table 4.

As shown in Table 4, IAO outperformed all other methods on all metrics, achieving accuracy, precision, recall and F1 scores above 0.9, significantly surpassing other approaches. In contrast, BO and GS failed to notably improve model performance, and although RS provided some enhancement, it remained inferior to IAO. Overall, IAO exhibited a distinct advantage in parameter optimization, leading to enhanced model accuracy, precision, recall and F1 scores.

3.4 Effectiveness of CDORPF

To further validate the effectiveness of CDORPF, we conducted a comparative experiment. This experiment compared commonly used machine learning models with CDORPF and the results are presented in Table 5.

Table 5 clearly demonstrates that the CDORPF framework significantly outperforms other commonly used machine learning models across all evaluation metrics, confirming its superior accuracy and reliability. Moreover, CDORPF exhibits a notable level of consistency in accuracy, precision, recall and F1-score, which is critically important for practical applications.

4 Discussion and interpretation

We developed a framework named CDORPF to address the issue of missing values in microbiome data, transforming high-dimensional

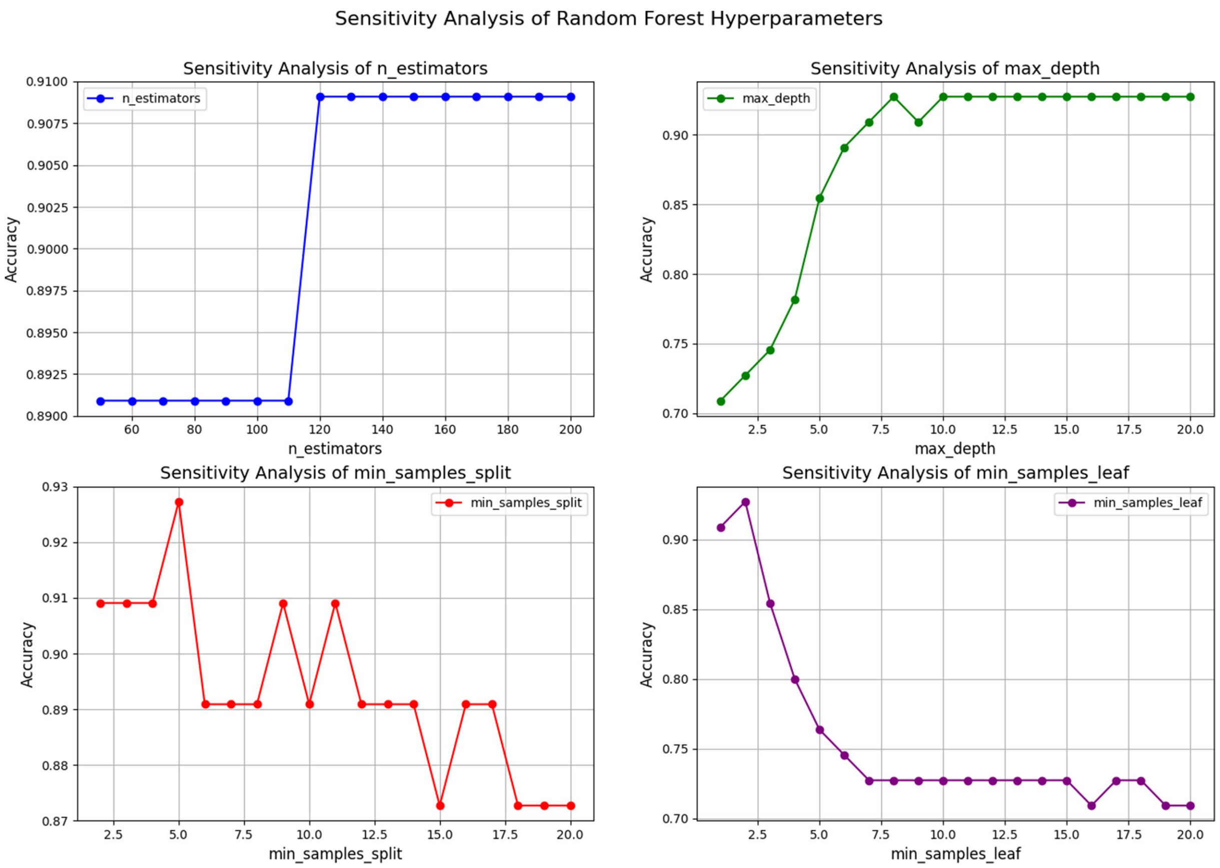


FIGURE 8
Sensitivity analysis of RF hyperparameters.

TABLE 3 The hyper-parameters tuning results of IAO-optimized RF.

Hyper-parameter	Description	Parameter range	Tuning result
n_estimators	The number of trees in the forest	[10, 200]	200
max_depth	The maximum depth of the trees	[1, 50]	9
min_samples_split	The minimum number of samples required to split an internal node	[2, 10]	2
min_samples_leaf	The minimum number of samples required at a leaf node	[1, 5]	1

TABLE 4 Results of different parameter optimization methods.

Method	Accuracy	Precision	Recall	F1-score
No optimization	0.8727	0.8731	0.8727	0.8726
BO	0.8545	0.8626	0.8545	0.8540
GS	0.8727	0.8775	0.8727	0.8725
RS	0.8909	0.8996	0.8909	0.8905
IAO	0.9043	0.9084	0.9043	0.9040

microbiome profiles into low-dimensional representations and constructing classification models based on these representations.

In the initial phases of this research, our primary objective was data imputation, as most machine learning algorithms are not designed to effectively handle missing values, which can lead to bias. We explored traditional KNN and MICE methods for imputation; however, the results were suboptimal. To preserve the inherent structure and relationships within microbiome data, we proposed the TOI method that integrates KNN, MICE and Bayesian ridge regression. By analyzing the types of missing data present in the IBD dataset, we validated the rationale behind the TOI method and successfully imputed all missing values while maintaining internal structural integrity. As illustrated in Figure 5, experimental results indicate that the distribution of imputed data closely aligns with that of original data—demonstrating that TOI enhances completeness while preserving critical features—thereby laying a solid foundation for subsequent analyses and model development.

Building upon this groundwork, we conducted dimensionality reduction experiments since high dimensionality in 16S rRNA data introduces noise detrimental to downstream predictions. Our proposed IWVAE method outperformed PCA and VAE by effectively reducing dimensions while retaining essential features. As presented in Table 2, IWVAE achieved superior performance across metrics such as accuracy, precision, recall and F1 score—significantly enhancing overall model efficacy and showcasing its exceptional capability in feature extraction.

TABLE 5 Effectiveness of CDORPF.

Model	Accuracy	Precision	Recall	F1-score
XGBoost	0.7500	0.6481	0.6607	0.6535
LightGBM	0.7222	0.3889	0.5000	0.4375
CatBoost	0.7778	0.6563	0.5893	0.6000
SVM	0.7778	0.3889	0.5000	0.4375
Mice + RF	0.8333	0.8627	0.8333	0.7914
Mice + SVM	0.7778	0.6049	0.7778	0.6806
CDORPF	0.9043	0.9084	0.9043	0.9040

In our model optimization experiments involving global hyperparameter tuning using IAO-RF demonstrated notable advantages over RS, GS and BO (Table 4). This further substantiates the effectiveness of IAO optimization strategies in improving model performance particularly when dealing with complex datasets characterized by enhanced accuracy and stability.

In summary, our CDORPF framework exhibits significant strengths in addressing issues related to incompleteness, high dimensionality and sparsity within microbiome datasets. As evidenced by Table 5, CDORPF surpasses traditional machine learning models across all evaluated metrics offering improved accuracy alongside consistency thus affirming its potential applicability within real-world scenarios.

By adeptly integrating components such as data imputation, dimensionality reduction and risk prediction, the CDORPF framework effectively confronts challenges associated with microbiome information. Future investigations could further validate this framework's robustness on larger more intricate datasets whilst exploring prospective applications across diverse fields.

5 Conclusion

In recent years, IBD has become a global health challenge with a substantial treatment burden. Research has consistently shown a strong association between the human gut microbiome and IBD pathogenesis, making it crucial for risk prediction. To address the challenges of high-dimensional, sparse, and incomplete microbiome data, this paper introduces a novel integrated data optimization and risk prediction framework, CDORPF. Compared to traditional methods, this approach excels in handling complex microbiome data by preserving the inherent structure of the data, minimizing biases from missing data, and significantly enhancing data integrity and analytical reliability. Additionally, it effectively retains the core information during dimensionality reduction, while markedly improving model predictive performance. This approach offers a comprehensive solution to the challenges of missing values and high dimensionality commonly found in microbiome data.

In clinical workflows, CDORPF can serve as a complementary tool to existing diagnostic methods by providing additional risk assessment information through microbiome analysis. This not only enhances diagnostic accuracy but also optimizes the treatment process, making patient management more refined and personalized. For example, in the initial screening phase, CDORPF can leverage gut microbiome data to help identify high-risk patients, prioritizing further diagnostic or intervention measures and reducing unnecessary delays. Future research can further explore the performance of CDORPF in large-scale, multi-center clinical trials to validate its applicability and robustness across different populations and disease subtypes. Moreover, the successful application of the CDORPF framework offers new research directions for the early diagnosis of other complex diseases, such as cardiovascular disease or cancer, through microbiome analysis, thereby advancing broader applications in personalized medicine.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://hmpdacc.org/ihmp/.

Ethics statement

The studies involving humans were approved by Institutional Review Board at Washington University in St. Louis. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

YP: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Project administration, Writing – review & editing. YuL: Conceptualization, Data curation, Software, Visualization, Writing – original draft, Writing – review & editing. YiL: Data curation, Validation, Visualization, Writing – review & editing, Software. JW: Investigation, Methodology, Project administration, Validation, Writing – review & editing, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work is supported by the National Natural Science Foundation of China (Grant no. 62172287).

Acknowledgments

The authors express our deep gratitude to everyone who has provided assistance with this 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

References

Abualigah, L., Yousri, D., Abd Elaziz, M., Ewees, A. A., Al-qaness, M. A. A., and Gandomi, A. H. (2021). Aquila optimizer: a novel meta-heuristic optimization algorithm. *Comput. Ind. Eng.* 157:107250. doi: 10.1016/j.cie.2021.107250

Alfonso Perez, G., and Castillo, R. (2023). Gene identification in inflammatory bowel disease via a machine learning approach. *Medicina* 59:1218. doi: 10.3390/medicina59071218

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., et al. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568, 499–504. doi: 10.1038/s41586-019-0965-1

Ansarullah, S. I., and Kumar, P. (2019). A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method. *Int. J. Recent Technol. Eng.* 7, 1009–1015.

Armstrong, G., Rahman, G., and Martino, C. (2022). Applications and comparison of dimensionality reduction methods for microbiome data. *Front. bioinform.* 2:82186. doi: 10.3389/fbinf.2022.821861

Azur, M. J., Stuart, E. A., Frangakis, C., and Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *Int. J. Methods Psychiatr. Res.* 20, 40–49. doi: 10.1002/mpr.329

Bisgaard, T. H., Allin, K. H., Keefer, L., Ananthakrishnan, A. N., and Jess, T. (2022). Depression and anxiety in inflammatory bowel disease: epidemiology, mechanisms and treatment. *Nat. Rev. Gastroenterol. Hepatol.* 19, 717–726. doi: 10.1038/s41575-022-00634-6

Chen, X., He, R., Chen, X., Jiang, L., and Wang, F. (2023). Optimizing dose-schedule regimens with Bayesian adaptive designs: opportunities and challenges. *Front. Pharmacol.* 14:1261312. doi: 10.3389/fphar.2023.1261312

Die, D., Zhu, J., Sun, C., Li, M., Liu, J., Wu, S., et al. (2022). GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res.* 50, D777–D784. doi: 10.1093/nar/gkab1019

Doove, L. L., Van Buuren, S., and Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Comput. Stat. Data Anal.* 72, 92–104. doi: 10.1016/j.csda.2013.10.025

Faye, A. S., Holmer, A. K., and Axelrad, J. E. (2022). Cancer in inflammatory bowel disease. *Gastroenterol. Clin. N. Am.* 51, 649–666. doi: 10.1016/j.gtc.2022.05.003

Feng, T., Peng, Y., and Wang, J. (2023). ISGS: a combinatorial model for hysteresis effects. *Acta Electron. Sin.* 51, 2504–2509. doi: 10.12263/DZXB.20220238

Flynn, S., and Eisenstein, S. (2019). Inflammatory bowel disease presentation and diagnosis. *Surg. Clin. North Am.* 99, 1051–1062. doi: 10.1016/j.suc.2019.08.001

Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). Automated machine learning: methods, systems, challenges. Cham: Springer.

Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10:5029. doi: 10.1038/s41467-019-13036-1

Kalaiselvi, B., and Geetha, S. (2024). "Ensemble voting classifier-based machine learning model for predictive modeling of campus student placements" in Science and technology: recent updates and future prospects, 93–107.

Kostic, A. D., Xavier, R. J., and Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146, 1489–1499. doi: 10.1053/j.gastro.2014.02.009

Kraszewski, S., Szczurek, W., Szymczak, J., Reguła, M., and Neubauer, K. (2021). Machine learning prediction model for inflammatory bowel disease based on laboratory markers. *J. Clin. Med.* 10:4745. doi: 10.3390/jcm10204745

Li, I., Fabbri, A., Hingmire, S., and Radev, D. (2020). R-VGAE: relational-variational graph autoencoder for unsupervised prerequisite chain learning. Proceedings of the 28th International Conference on Computational Linguistics

Li, Y., Pan, J., Zhou, N., Fu, D., Lian, G., Yi, J., et al. (2021). A random forest model predicts responses to infliximab in Crohn's disease based on clinical and serological parameters. *Scand. J. Gastroenterol.* 56, 1030–1039. doi: 10.1080/00365521.2021.1939411

Liñares-Blanco, J., Fernandez-Lozano, C., Seoane, J. A., and López-Campos, G. (2022). Machine learning based microbiome signature to predict inflammatory bowel disease subtypes. *Front. Microbiol.* 13:872671. doi: 10.3389/fmicb.2022.872671

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. doi: 10.1038/s41586-019-1237-9

Mafarja, M. M., and Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* 260, 302–312. doi: 10.1016/j.neucom.2017.04.053

Peng, Y., Liu, Y., Wu, H., Liu, L., and Wang, J. (2023). High-dimensional data mining algorithm based on SPCA-GWR. *J. Nonlinear Convex Anal.* 24, 1303–1314.

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. Proceedings of the International Conference on Machine Learning. 1278–1286.

Rogler, G., Singh, A., Kavanaugh, A., and Rubin, D. T. (2021). Extraintestinal manifestations of inflammatory bowel disease: current concepts, treatment, and implications for disease management. *Gastroenterology* 161, 1118–1132. doi: 10.1053/j.gastro.2021.07.042

Tang, F., and Ishwaran, H. (2017). Random forest missing data algorithms. *Stat. Anal. Data Min.* 10, 363–377. doi: 10.1002/sam.11348

Wang, J., Li, J., Wang, Z., Zhou, S., and Peng, Y. (2023a). An interpretable prediction model for heart disease risk based on improved whale optimized LightGBM. *J. Beijing Univ. Posts Telecommun.* 46:39. doi: 10.13190/j.jbupt.2023-015

Wang, J., Wang, Z., Li, J., and Peng, Y. (2023b). An interpretable depression prediction model for the elderly based on ISSA optimized LightGBM. *J. Beijing Inst. Technol.* 32, 168–180. doi: 10.15918/j.jbit1004-0579.2023.010

Xie, F., and Lederer, J. (2021). Aggregating knockoffs for false discovery rate control with an application to gut microbiome data. *Entropy* 23:230. doi: 10.3390/e23020230

Yekkala, I., Dixit, S., and Jabbar, M. A. (2017). Prediction of heart disease using ensemble learning and particle swarm optimization. Proceedings of the 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon). IEEE. 691–698.

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 02 October 2024
DOI 10.3389/fmicb.2024.1474078

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Mengyuan Ji,
University of Padua, Italy
Hanyan Li,
University of Oklahoma, United States

*CORRESPONDENCE
Min Liu
✉ liuminxiaoqi@163.com

RECEIVED 01 August 2024
ACCEPTED 23 September 2024
PUBLISHED 02 October 2024

CITATION
Gao Y and Liu M (2024) Application of machine learning based genome sequence analysis in pathogen identification. *Front. Microbiol.* 15:1474078.
doi: 10.3389/fmicb.2024.1474078

Application of machine learning based genome sequence analysis in pathogen identification

Yunqiu Gao[1,2] and Min Liu[1,3]*

[1]Department of Dermatology, The First Hospital of China Medical University, Shenyang, China, [2]Key Laboratory of Immunodermatology, Ministry of Education and NHC, National Joint Engineering Research Center for Theranostics of Immunological Skin Diseases, Shenyang, China, [3]Institute of Respiratory Disease, China Medical University, Shenyang, China

Infectious diseases caused by pathogenic microorganisms pose a serious threat to human health. Despite advances in molecular biology, genetics, computation, and medicinal chemistry, infectious diseases remain a significant public health concern. Addressing the challenges posed by pathogen outbreaks, pandemics, and antimicrobial resistance requires concerted interdisciplinary efforts. With the development of computer technology and the continuous exploration of artificial intelligence(AI)applications in the biomedical field, the automatic morphological recognition and image processing of microbial images under microscopes have advanced rapidly. The research team of Institute of Microbiology, Chinese Academy of Sciences has developed a single cell microbial identification technology combining Raman spectroscopy and artificial intelligence. Through laser Raman acquisition system and convolutional neural network analysis, the average accuracy rate of 95.64% has been achieved, and the identification can be completed in only 5 min. These technologies have shown substantial advantages in the visible morphological detection of pathogenic microorganisms, expanding anti-infective drug discovery, enhancing our understanding of infection biology, and accelerating the development of diagnostics. In this review, we discuss the application of AI-based machine learning in image analysis, genome sequencing data analysis, and natural language processing (NLP) for pathogen identification, highlighting the significant role of artificial intelligence in pathogen diagnosis. AI can improve the accuracy and efficiency of diagnosis, promote early detection and personalized treatment, and enhance public health safety.

KEYWORDS

artificial intelligence (AI), antibiotic resistance, pathogenic microorganisms, machine learning (ML), diagnosis

Introduction

Pathogenic microorganisms include viruses, bacteria, parasites, and fungi that can cause infections in humans and animals. They spread rapidly through aerosols, body fluids, food, and direct contact, leading to various infectious diseases and even death (Zhang et al., 2018). Early detection, diagnosis, and treatment are crucial for preventing infectious diseases. Since the discovery of penicillin in 1928, antibiotics have become vital public health tools, saving countless lives globally (Fleming, 2001; Davies and Davies, 2010). Today, a wide range of antibacterial, antifungal, and antiviral drugs are used in clinical practice. However, the misuse of these antimicrobial drugs has led to increased drug resistance in microorganisms, reducing the effectiveness of these treatments, a phenomenon known as antimicrobial resistance

(Prestinaci et al., 2015). According to the World Health Organization (WHO) in 2020, antimicrobial resistance (AMR) is among the top ten global public health threats facing humanity. In 2022, The Lancet published a systematic analysis of the global burden of bacterial AMR, including data from over 200 countries. The study revealed that AMR poses a significant threat to global health. In 2019, AMR infections directly caused approximately 1.27 million deaths and indirectly resulted in about 4.95 million deaths worldwide (Antimicrobial Resistance Collaborators, 2022). By 2022, around 1.3 million deaths were related to antibiotic resistance (Ranjbar and Alam, 2023). If left unaddressed, it is projected that by 2050, antibiotic-resistant infections could cause 10 million deaths annually, with direct economic losses exceeding $10 trillion (de Kraker et al., 2016; Ventola, 2015). Developing new antimicrobial drugs is becoming increasingly difficult, often taking 10–15 years and costing over 6 billion (Wouters et al., 2020; DiMasi et al., 2016). The emergence of more severe multidrug-resistant bacteria will pose significant treatment challenges. These data highlight the substantial burden that infectious diseases and antimicrobial resistance place on human health and the global economy.

The technologies in pathogen detection include nucleic acid and immunological methods (Whiley and Taylor, 2016) (Figure 1). These technologies help identify pathogenic bacteria or potential health risks, making accurate and rapid detection crucial for diagnosing and preventing diseases in public health, environmental pollution monitoring (Zhang et al., 2023), and clinical diagnosis (Smith and Kirby, 2020). However, current detection techniques often fall short of clinical needs due to long processing times, cumbersome procedures, and reliance on large instruments, limiting fast and efficient identification. The traditional methods for identifying pathogenic microorganisms, including smear microscopy, isolation and cultivation, biochemical assays etc., are not without limitations. These methods are often characterized by prolonged timeframes, intricate procedures, and suboptimal sensitivity. A case in point is the identification of mycobacterial strains, which can extend to a lengthy period of 30 to 40 days. Furthermore, certain fastidious bacteria and viruses demand cultivation conditions that are so stringent they may prove unattainable, or the organisms may be refractory to culture altogether. Molecular diagnostic techniques, anchored in PCR, have made strides in addressing some of the aforementioned challenges in pathogen detection. However, they encounter significant hurdles when it comes to the identification of unknown microorganisms. The absence of known nucleic acid sequences renders the design of specific primers an insurmountable obstacle for these technologies. While immunological and PCR methods boast high sensitivity and specificity, enabling the detection of a broad spectrum of pathogens, they are constrained by their targeted nature. This means that a single experiment is typically capable of detecting only one pathogen, which can lead to diminished diagnostic efficiency. The indistinguishable symptoms and signs of many infectious diseases further complicate matters, as identical clinical presentations may be induced by a variety of pathogens or result from co-infections. The laborious and time-consuming process of detecting pathogens one at a time can potentially lead to diagnostic delays.

A key breakthrough in overcoming these limitations is the deployment of AI driven genome sequencing tools, which analyze complex genomic data to quickly and accurately identify pathogenic microorganisms with high throughput and speed. For example, DeepVariant is a mutation caller based on deep learning that can improve the accuracy score of single nucleotide mutation and Indel detection (Poplin et al., 2018). Integrating image processing and big data analysis into detection methods is therefore highly significant (Kothari et al., 2014; Jain et al., 2016). Recent advancements in AI, particularly in computer vision and image processing, have shown promising potential in the morphological detection of pathogenic microorganisms.

The development of AI has progressed through several key stages. It began in 1945 with Alan Turing's idea of using computers to simulate the human brain. During the 1950s to the 1970s, AI started to become practical with the creation of the first generation of AI systems. The 2010s saw an explosion in AI capabilities, driven by advances in deep learning and big data technologies like chatGPT (LeCun et al., 2015; Esteva et al., 2017). Today, AI excels in numerous fields, including disease diagnosis, risk management, facial recognition (Figure 2).

AI has significant applications in microbial diagnosis. It uses machine learning algorithms to analyze microbial genome data, identify antibiotic resistance genes, speed up pathogen identification, and improve diagnostic accuracy. AI can also process vast amounts of complex data, provide real-time diagnostic support, aid in the early detection and control of infectious diseases, and enhance public health prevention and control efforts (Esteva et al., 2017).

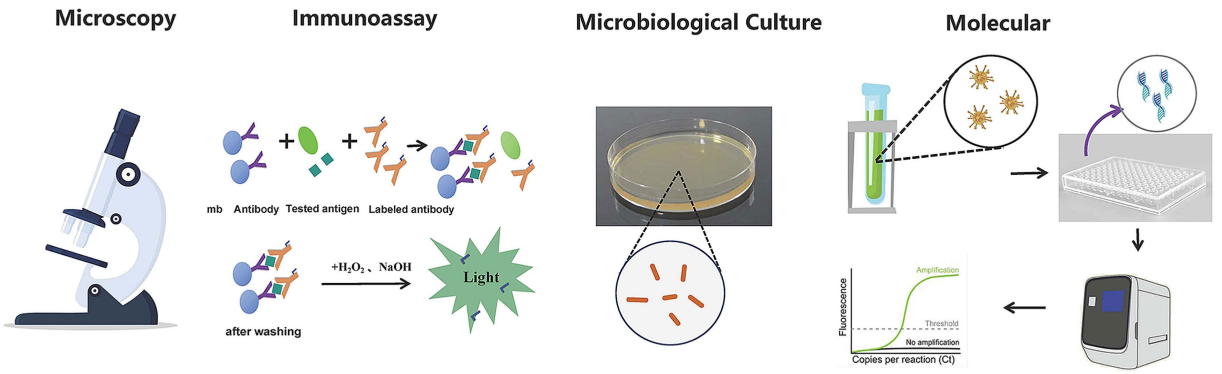


FIGURE 1
Pathogenic microorganism detection technology.

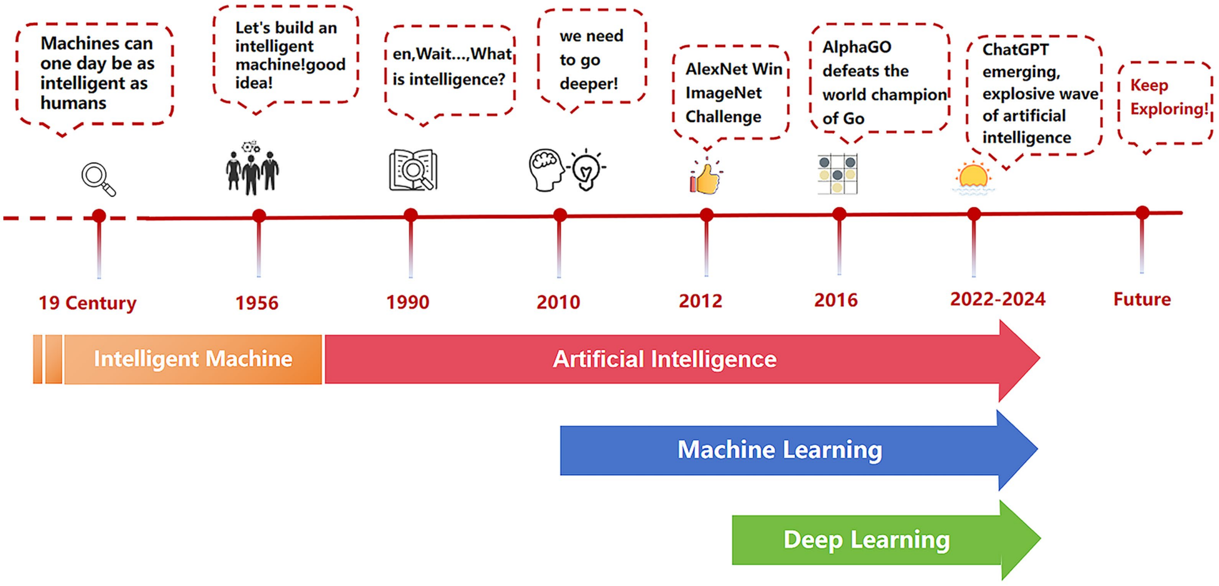


FIGURE 2
Major milestones in the development of artificial intelligence.

The potential of AI in microbiology is yet to be fully realized. Microbial research generates vast amounts of biological image data, and AI has proven crucial in analyzing high-throughput sequencing data and using natural language processing to identify pathogenic microorganisms. Traditional computing methods are slow in processing these data, whereas AI, especially deep learning, excels in both accuracy and speed (Camacho et al., 2018; Ching et al., 2018). Deep learning has introduced new applications to microbial research, significantly advancing microbial identification and diagnosis. The application of deep learning in microbial image recognition and classification has grown rapidly (Wainberg et al., 2018; Cao et al., 2018; Jiang et al., 2022). This article reviews the use of AI in identifying and diagnosing pathogenic microorganisms.

Application of AI in image analysis of pathogenic microorganisms

AI, particularly machine learning and deep learning, has made significant strides in the automatic recognition and classification of pathogenic microorganisms in microscope images. These technologies effectively analyze and classify bacteria, viruses, fungi, and parasites. Deep learning has made microscope image analysis more efficient and universal, enabling accurate cell detection and classification (Figure 3). Compared to traditional methods, deep learning significantly enhances the accuracy and reliability of microorganism detection (Esteva et al., 2021; Chen and Asch, 2017).

To address the challenges of pathogen detection, particularly with large sample sizes and the identification of difficult bacteria, researchers have been exploring intelligent clinical microbial morphology testing. In 2020, Professor Aydogan Ozcan's team at the University of California developed a highly sensitive, precise, timely, and low-cost microbial online monitoring AI platform. This system combines coherent microscopy imaging with deep neural network analysis to enable the intelligent identification and classification of live microorganisms. By analyzing growth delay holograms, the system achieves rapid detection of bacterial growth and species classification, with a detection limit of approximately 1 CFU/L for *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa* within ≤9 h. This significantly reduces the testing time compared to the EPA gold standard method, which takes at least 24 h (Wang et al., 2020). Similarly, a team from the University of Geneva in Switzerland has developed an automated urine culture analysis system. The WASPLab software automatically reads and analyzes bacterial colony images on urine culture plates, quickly reporting urine culture results. Using automated equipment, the turnaround time is reduced by nearly 50%, minimizing manual reading errors and improving detection efficiency and accuracy (Cherkaoui et al., 2020).

Currently, the interpretation of imaging results relies heavily on the subjective clinical experience of professional imaging doctors. Clinically, there is a strong expectation for the testing department to diagnose pathogens rapidly and provide accurate drug sensitivity results. AI is now widely used in medical imaging, particularly in detecting and diagnosing infectious diseases. For instance, during the global COVID-19 outbreak in 2019, AI significantly improved the diagnostic accuracy and efficiency of chest CT scans and X-rays, enabling rapid and precise screening, identification, and characterization of COVID-19 (Hassan et al., 2022). AI also aids in detecting and analyzing secondary pulmonary infections in COVID-19 patients, enhancing diagnostic accuracy and helping to assess disease severity and predict clinical outcomes (Viswanathan et al., 2022). In lymphoma patients, deep learning accurately identifies high metabolic tumor sites in 18F-FDG-PET/CT scans, potentially aiding in excluding metabolically active diseases (Ikeda et al., 1987). These studies highlight AI's potential in enhancing diagnostic efficiency and accuracy for infectious diseases and its broad application prospects in medical imaging.

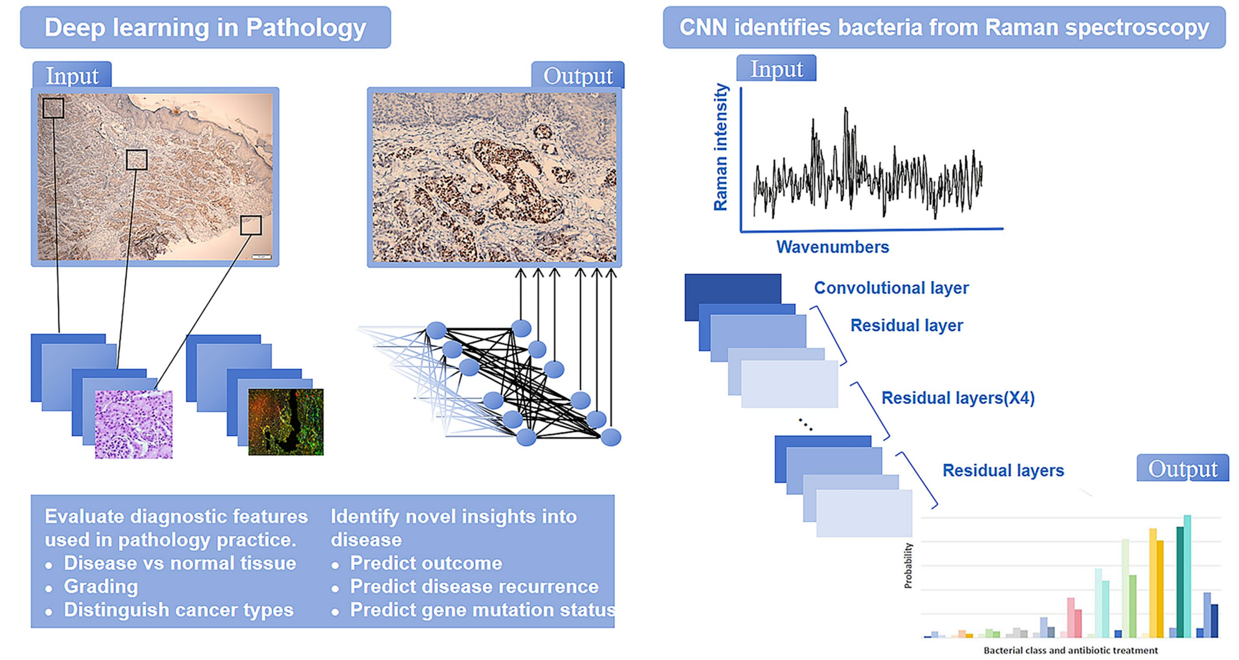


FIGURE 3
The application of AI related algorithms in image analysis.

TABLE 1 Top 10 countries/regions medical imaging in artificial intelligence research from 2014 to 2023.

Rank	Countries/regions	Number of publications/article
1	CHINA	14,338
2	USA	11,309
3	INDIA	5,377
4	ENGLAND	3,291
5	UK	3,184
6	SOUTH KOREA	2,575
7	GERMANY	2,505
8	CANADA	2,390
9	SAUDI ARABIA	1946
10	ITALY	1932

TABLE 2 Top 10 research area in artificial intelligence research from 2014 to 2023.

Rank	Research area	Number of publications/article
1	Mathematical Computational Biology	42,691
2	Radiology Nuclear Medicine Medical Imaging	36,960
3	Engineering	35,466
4	Communication	33,268
5	Mathematics	20,565
6	Neurosciences Neurology	9,151
7	Science Technology Other Topics	9,003
8	Imaging Science Photographic Technology	6,757
9	Oncology	6,608
10	Automation Control Systems	6,347

We conducted a bibliometric analysis using the Web of Science database to search for original research on the application of AI in medical imaging over the past decade, with the keywords "Artificial Intelligence" and "Medical Imaging." We analyzed the retrieved literature and generated a citation report. A total of 50,547 articles were found, with a notable increase in publication volume since 2020. Europe and the United States remain leaders in this field, while Chinese scholars have shown rapid development in the past 2 years, now leading in publication volume. However, the impact of Chinese research is relatively low, indicating an academic quality gap with European and American countries in AI-assisted medical imaging (Tables 1, 2; Figure 4). Most clinical research focus on using deep learning and its derivative algorithms to improve image segmentation accuracy and assist clinical diagnosis. According to our statistical results, AI ranks ninth in the field of infection research. With significant progress in AI-driven microbial microscopy image detection, the application of deep learning in microbial image recognition and classification has immense development potential.

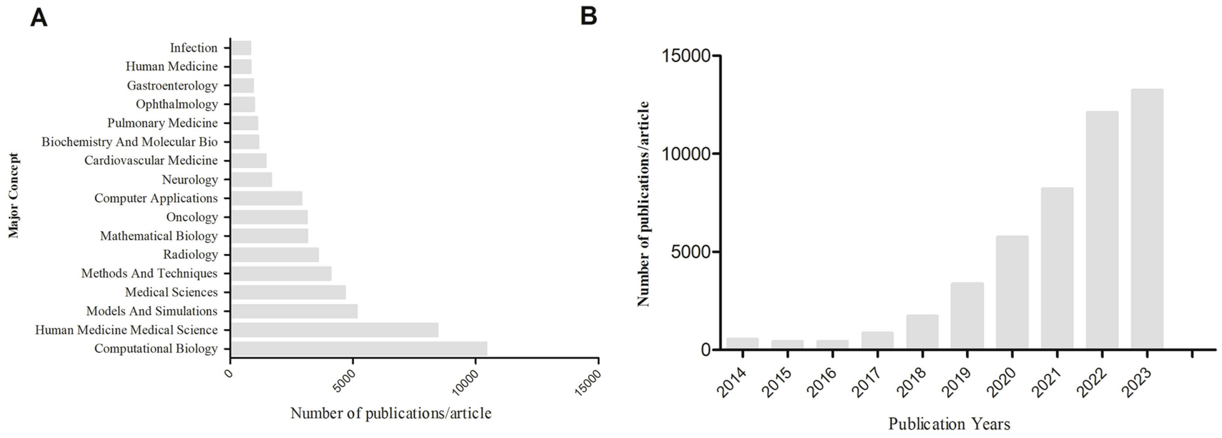


FIGURE 4
(A) Search for keywords "artificial intelligence" and "medical imaging" to rank the top major concepts in terms of article volume. **(B)** The number of articles and publications retrieved using "artificial intelligence" and "medical imaging" as keywords from 2014 to 2023.

The application of AI in genome sequencing data analysis

AI is widely used in analyzing next-generation sequencing (NGS) data, particularly for pathogen identification and classification. AI technology can quickly process NGS data and identify pathogens in samples, which is crucial for the timely diagnosis of infectious diseases. In a study published in Nature Medicine, scientists have developed an AI framework that integrates a multi detection platform for detecting and identifying biomolecules. The system analyzes three representative plasmids with different color signals, which are derived from drug-resistant *Klebsiella pneumoniae* bacteria. Compared with traditional technology, this system demonstrates excellent recall and accuracy, detecting 93.8% of events in real-time and achieving a classification accuracy of 99.8%. This study demonstrates the potential of AI in medical diagnosis, especially in clinical environments that require rapid and accurate analysis (Ganjalizadeh et al., 2023). A study published in Scientific Reports, researchers used AI algorithms combined with NGS data from T cell receptors (TCRs) to diagnose glioma patients. This study explores multidimensional classification and feature selection of TCR sequence diversity index, as well as two-dimensional classification and feature selection analysis of TCR related sequences. The results indicate that through these analyzes, researchers were able to identify two sets of core sequences, each containing three sequences, sufficient to achieve a 96.7% accuracy in glioma detection and diagnosis (Zhou et al., 2024).

The metagenomic high-throughput sequencing technology (mNGS) has shown great potential in pathogen detection. It identifies pathogenic microorganisms by directly sequencing nucleic acids in samples, without the need to pre-set target sequences, thus overcoming the limitations of traditional microbial detection methods. The IDseq platform is a cloud based open-source platform developed by the Chan Zuckerberg Initiative. Based on pathogen metagenomics detection technology, high-throughput sequencing technology is used to analyze microorganisms and host nucleic acids in clinical samples, enabling unbiased detection of various pathogenic microorganisms, including bacteria, fungi, viruses, and parasites. This technology has shown important application value in the detection of infectious diseases pathogens, especially when the traditional etiological diagnosis methods are difficult to meet the clinical needs. This platform has the comprehensiveness to process diverse samples and detect numerous pathogens, high sensitivity to improve pathogen detection sensitivity, and in-depth analysis capabilities for drug resistance and virulence analysis. Its open-source nature and cloud computing foundation make it easy to access and process big data on a global scale, reducing the need for bioinformatics experts and local server level hardware resources through automated processes, thereby lowering costs and time. The platform is user-friendly and supports real-time pathogen detection, including newly emerging pathogens. It also supports the generation of environmental background models and data sharing, promoting scientific research collaboration (Kalantar et al., 2020). In a case of pathogen discovery in childhood meningitis in Bangladesh, researchers used the IDseq platform to reanalyze three meningitis samples with the aim of exploring unknown pathogens. These three samples include one meningitis sample caused by *Streptococcus pneumoniae* (CHRF 0002), one meningitis sample caused by chikungunya virus (CHRF 0094), and one water control sample (CHRF 0000). The IDseq platform has successfully identified pathogens through effective host sequence filtering and quality control. Especially in the CHRF 0094 sample, after host filtering and QC steps, the chikungunya virus accounted for 63% of non-host reads, and through the coverage visualization tool of the IDseq portal, researchers were able to observe the whole genome coverage of the chikungunya virus in the sample. This indicates that the IDseq platform can effectively assist researchers in quickly obtaining in-depth insights into sample quality, microbial content, and cohort trends (Saha et al., 2019).

AI algorithms can accurately classify pathogens based on genomic data, which is crucial for monitoring their evolution and transmission. MetaPhlAn (Metagenomic Phylogenetic Analysis) is a widely used bioinformatics tool that provides species-level analysis of microbial composition from metagenomic shotgun sequencing data. A 2023 Nature article detailed how researchers integrated extensive new microbial genome and metagenomic data into the MetaPhlAn database, defining 26,970 Species-Level Genome Bins (SGBs). This expansion allows MetaPhlAn 4 to analyze metagenomic data more accurately, particularly in identifying uncharacterized species and improving the explanatory power of microbial community composition analysis (Blanco-Miguez et al., 2023). Antibiotic resistance is a pressing global health threat. Rapid whole-genome sequencing offers opportunities to predict antibiotic resistance from genomic data. In 2024, the Helmholtz Center for Infection Research in Braunschweig, Germany, evaluated four advanced machine learning methods (Kofer, PhenotypeSeeker, Seq2Geno2Pheno, and Aytan Aktig), a baseline ML method, and ResFinder. The results showed significant performance differences among these technologies and datasets, with ML methods excelling in closely related strains and ResFinder performing better with more divergent genomes. ResFinder, combining AI technology, can detect and classify antibiotic resistance genes from NGS data, providing crucial data for public health monitoring (Hu et al., 2024).

To address data diversity, break down information silos, meet the demands of big data analysis, enhance research efficiency, support interdisciplinary research, and leverage modern information technology, integrating databases and knowledge bases has become essential. AI algorithms, combined with extensive databases like NCBI[1] and EMBL-EBI[2], and knowledge bases, can significantly improve the accuracy of pathogen identification and classification. For instance, Kraken2 is a highly efficient pathogen classification tool that uses AI technology and a comprehensive reference database to enable rapid analysis of NGS data.

Application of NLP in identification of pathogenic microorganisms

What is Natural Language Processing (NLP)? NLP is a machine learning technology that enables computers to interpret, process, and understand human language. It serves as a crucial bridge for communication between humans and machines.

Medical literature is an essential resource for both medical and clinical research. The vast variety of pathogenic microorganisms and parasites associated with infectious diseases, however, poses significant challenges for doctors and researchers when it comes to consulting and organizing this massive volume of literature. The application of NLP technology facilitates the extraction of valuable insights from medical literature and enhances the accuracy and convenience of laboratory data analysis. NLP technology can process microbial data through structured data techniques, such as standardizing EMR (Electronic Medical Records) and laboratory data, then storing this information in databases. Additionally, deep learning algorithms can denoise, segment, and extract features from imaging data (Ananiadou et al., 2010; Chen et al., 2015; Wang et al., 2018; Lee et al., 2020; Rajkomar et al., 2019). An article published in Scientific Reports in 2024 introduced a MarkerGeneBERT system, an NLP system developed by CapitalBio Technology, which automatically extracts information on species, tissues, cell types, and cell marker genes from single-cell sequencing literature. In a study, the system extracted 8,873 human and 9,064 mouse cell markers from 3,987 studies, demonstrating 76% completeness and 75% accuracy, surpassing the CellMarker2.0 system. In addition, MarkerGeneBERT has discovered 89 new cell types and 183 new marker genes. In terms of gene recognition, the system achieved an F1 score of 87%, with a cell name recognition accuracy of 92%. More than 20,000 genes and 4,000 cell types were identified from literature, with accuracies of 90.8 and 92.7%, respectively. Additionally, 1764 new cell types were added, all of which were not previously recorded in the database (Cheng et al., 2024).

In 2022, David Burstein's team published an article in Nature Communications on using NLP to interpret microbial gene function. They developed a deep learning model that utilized gene embeddings, calculated based on the co-occurrence rate of gene families, as input for a classifier to predict gene function. The word2vec algorithm was employed to calculate the gene embedding space, providing a simple, fast, and direct method. Through scarcity analysis, the study highlighted functional categories with high discovery potential and uncovered hypothetical bacterial membrane-binding mechanisms and microbial defense systems in the human microbiome. Additionally, NLP models can be fine-tuned to explore specific systems or functions, such as training classifiers for particular genes or creating new embeddings using relevant corpora (such as virus genomes, specific microbial communities). This approach is applicable not only for inferring functions of genes without sequence similarity to characteristic proteins but also for exploring diverse functions of homologous genes. This greatly enhances the understanding of microbial gene functions and aids in interpreting unknown microbial gene functions and evolution (Miller et al., 2022). In the same year, another article in Nature Communications introduced a universal "gene semantic" model using NLP. This model employed convolutional neural networks (CNN) to classify peptide sequences and identify potential antimicrobial peptides (AMPs). The deep learning model demonstrated significantly higher accuracy and recall in identifying AMPs compared to traditional methods. A new set of AMPs sequences was identified from the human gut microbiome, showing strong antibacterial activity *in vitro* and validating the model's predictions (Ma et al., 2022).

Representative case

Antibiotics have been used to treat life-threatening infections for nearly a century, but with the increase of drug-resistant bacteria, traditional therapies are no longer effective against these infections. The crisis of antibiotic resistance has become an urgent global health issue that requires the discovery of a new generation of nucleic acid and peptide based antibiotics. However, traditional methods for developing antimicrobial peptides (AMPs) are slow and costly.

In 2023, Nat Commun published an article exploring methods to accelerate the development of AMPs by combining cell-free protein synthesis (CFPS) and deep learning techniques. Researchers use generative deep learning models to learn from a large number of unlabeled natural protein sequences and propose new AMPs sequences. Combined with the CFPS system, this *in vitro* transcription and translation system uses DNA templates for protein synthesis, enabling rapid and small-scale production and screening of hundreds of peptides, overcoming the cytotoxicity issues in traditional cell expression systems. Within 24 h, researchers designed, produced, and screened 500 candidate AMPs, ultimately identifying 30 functional AMPs, of which 6 exhibited high antibacterial activity against

---

[1] https://www.ncbi.nlm.nih.gov/

[2] https://www.ebi.ac.uk/

multidrug-resistant pathogens and low cytotoxicity to human cells. This study demonstrates the potential of deep learning and CFPS technology in accelerating the development of AMPs, providing an efficient and economical new approach to combat microbial resistance (Pandi et al., 2023).

In 2024, Fudan University and a team of Virtue scientists combined AI and biomedical research to predict nearly 1 million new antimicrobial peptides from the global microbiome. They developed a new machine learning algorithm that effectively reduces the false positive rate in AMP recognition. They predicted nearly 1 million novel non redundant antimicrobial peptides from 63,410 environmental and host related metagenomes worldwide, as well as 87,920 high-quality bacterial and archaeal genomes. They also created the AMP comprehensive database AMPSphere, which was published in the main issue of Cell (Santos-Junior et al., 2024).

In May 2023, Professor James Collins and his team published a paper in Nature Chemical Biology, using AI algorithms to discover a novel antibiotic abaucin that can specifically kill the drug-resistant bacterium *Acinetobacter baumannii*. This study is the first to use AI and interpretable deep learning to discover a groundbreaking new class of antibiotics that are effective against multidrug-resistant pathogens, demonstrating the enormous potential of AI in drug discovery and combating antibiotic resistance (Liu et al., 2023).

In a study published in the journal Antibiotics, researchers used a decision tree based machine learning algorithm to predict antibiotic resistance. This study trained 10 machine learning classifiers and generated predictive models for meropenem, ciprofloxacin, and cefotaxime drugs. Research has found that certain models exhibit higher F1 scores, accuracy, precision, and specificity among all machine learning models used. For example, RandomForestClassifier showed moderate F1 score (0.6), accuracy (0.61), and specificity (0.625) for ciprofloxacin. For cefotaxime, RidgeClassifier performed well and displayed F1 score (0.652), accuracy (0.654), and specificity (0.652) values. For meropenem, KNeighborsClassifier showed moderate F1 scores (0.629), accuracy (0.629), and specificity (0.629) (Yasir et al., 2022). In 2022, a collaboration between the Federal Institute of Technology Zurich, Basel University Hospital, and Basel University used mass spectrometry combined with AI algorithms to identify multidrug-resistant pathogens. Researchers collected over 300,000 clinical strains from four diagnostic laboratories in Switzerland between 2016 and 2018, using Bruker's MALDI Biotyper microbial mass spectrometry system. The mass spectrometry data were associated with drug resistance information to create the DRIAMS dataset, which includes data for 803 bacterial strains, over 300,000 clinical strains, and 768,300 antibiotic resistance entries for more than 70 antibiotics. Using this dataset, they trained three machine learning algorithms—logistic regression, gradient-boosted decision trees (LightGBM), and deep neural networks (MLP)—to establish a classification model for drug-resistant bacteria. The prediction model was validated with *Staphylococcus aureus*, *Escherichia coli*, and *Klebsiella pneumoniae*, showing AUROC values of 0.80, 0.74, and 0.74, respectively, indicating accurate predictions of antibiotic resistance. This study highlights the significant impact of AI in the image analysis of pathogenic microorganisms. Automated and intelligent image analysis technologies enable medical institutions to diagnose infectious diseases more quickly and accurately, enhancing overall public health prevention and control capabilities (Weis et al., 2022; Tahir et al., 2018).

Advantages and challenges

The main advantages of AI in diagnosing pathogenic microorganisms are:

1. Rapid Processing and Analysis: AI can quickly process large volumes of microbial data, including genomic and metabolomic information, significantly reducing the time needed for differential diagnosis. AI programs can complete complex data analysis in minutes, saving substantial time compared to traditional methods (Erlich and Narayanan, 2014; He et al., 2010; Topol, 2019).
2. High Accuracy: AI models, through training, achieve high-precision identification and classification, especially with complex microbial communities. Using machine learning and deep learning algorithms, AI can recognize specific microbial features and provide accurate diagnostic results (Knights et al., 2011; Libbrecht and Noble, 2015; Esteva et al., 2017).
3. Automation and Scalability: AI systems automate the microbial identification and diagnosis process, reducing manual operations and improving laboratory efficiency. These models continuously update and optimize with new data, adapting to evolving pathogenic microorganisms (Mamoshina et al., 2016).
4. Data Integration and Knowledge Discovery: AI integrates information from various sources—genomic, metabolite, and clinical data—to offer comprehensive diagnostic insights. Through big data analysis, AI uncovers new characteristics and resistance mechanisms in pathogens, contributing to public health and disease prevention (Marx, 2013; Libbrecht and Noble, 2015; Topol, 2019).

Currently, AI integration in global healthcare is driving a technological revolution. However, AI faces several major challenges:

1. Data Issues: Despite accumulating a large amount of medical data, high-value data is still scarce and scattered. Lack of unified data standards, widespread data silos, and enhanced requirements for personal medical information security (Topol, 2019; Raghupathi and Raghupathi, 2014). The other main challenges faced by AI in processing genomic data include incomplete and noisy data, which may lead to inaccurate analysis results. To overcome these issues, researchers have proposed various strategies, such as using interpolation techniques to fill missing values, using hybrid models to enhance robustness to noise, improving model generalization ability through data augmentation and transfer learning, and applying multi view learning and deep learning techniques to more comprehensively understand and predict genomic data. These methods help improve the accuracy and reliability of genomic data analysis, providing stronger support for researchers and clinical applications (Gupta and Gupta, 2019).
2. Data Interpretability: To prevent errors or inaccuracies in the application of artificial intelligence in healthcare, one can improve the data interpretability of AI models through various strategies. These include the use of transparent and simple algorithms, the application of local and global interpretation techniques, the calculation of SHAP values, the conduct of internal model analyzes, the assurance of model accountability,

the inference of causality, the establishment of clear model boundaries, the implementation of adversarial testing, the practice of continuous evaluation, the development of user-friendly interpretations, and the adoption of multimodal interpretation methods. Such methods aid in enhancing user trust in AI decision-making processes, ensuring model transparency and accountability, and fulfilling regulatory requirements (Finlayson et al., 2019; Chu et al., 2023).

(3) Data Privacy: In order to protect data privacy in artificial intelligence applications that enhance pathogen identification, various technologies and methods can be adopted, including federated learning, group learning, privacy computing technology, PHDtools platform, and differential privacy. These methods can effectively protect data involving personal privacy while improving the accuracy of pathogen identification by means of collaborative training models, combining edge computing and blockchain, applying homomorphic encryption and secure multi-party computing, developing interactive online platforms, and introducing data processing noise. These developments provide new ideas and solutions for privacy protection of medical data (Obermeyer and Emanuel, 2016; Price and Cohen, 2019; Ahuja, 2019; Martin and Zimmermann, 2024; Khalid et al., 2023).

Conclusion

With the advancement of algorithmic computing power, computer hardware, and the advent of the big data era, AI technology has flourished and penetrated the medical field, transforming traditional medical practices. This review discusses the significant role of AI in identifying and diagnosing pathogenic microorganisms. Machine learning and deep learning algorithms enable faster, more accurate pathogen recognition with automation, efficiency, high sensitivity, and specificity. AI-assisted imaging technology allows computers to analyze vast amounts of medical imaging data, helping doctors make quicker and more accurate diagnoses. Natural language processing in AI extracts valuable information from scientific literature and databases, aiding clinical decision-making and research. Additionally, AI algorithms accurately classify pathogens based on genomic data, crucial for monitoring pathogen evolution and transmission. Using machine learning to optimize antibiotic use in healthcare settings is a forward-thinking approach to combating antimicrobial resistance now and in the future. In order to further promote the development of this field, interdisciplinary collaboration between artificial intelligence researchers and microbiologists is particularly important. This will help combine the professional knowledge of microbiology with the powerful analytical capabilities of artificial intelligence to jointly develop more accurate and efficient pathogen identification tools.

Author contributions

YG: Writing – review & editing, Writing – original draft. ML: Supervision, 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 funded and supported by Natural Science Foundation of Shenyang city (No. 23-503-6-13).

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

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* 7:e7702. doi: 10.7717/peerj.7702

Ananiadou, S., Pyysalo, S., Tsujii, J., and Kell, D. B. (2010). Event extraction for systems biology by text mining the literature. *Trends Biotechnol.* 28, 381–390. doi: 10.1016/j.tibtech.2010.04.005

Antimicrobial Resistance Collaborators (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 399, 629–655. doi: 10.1016/S0140-6736(21)02724-0

Blanco-Miguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* 41, 1633–1644. doi: 10.1038/s41587-023-01688-w

Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., and Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell* 173, 1581–1592. doi: 10.1016/j.cell.2018.05.015

Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., et al. (2018). Deep learning and its applications in biomedicine. *Genomics Proteomics Bioinformatics* 16, 17–32. doi: 10.1016/j.gpb.2017.07.003

Chen, J. H., and Asch, S. M. (2017). Machine learning and prediction in medicine – beyond the peak of inflated expectations. *N. Engl. J. Med.* 376, 2507–2509. doi: 10.1056/NEJMp1702071

Chen, Y., Lasko, T. A., Mei, Q., Denny, J. C., and Xu, H. (2015). A study of active learning methods for named entity recognition in clinical text. *J. Biomed. Inform.* 58, 11–18. doi: 10.1016/j.jbi.2015.09.010

Cheng, P., Peng, Y., Zhang, X. L., Chen, S., Fang, B. B., Li, Y. Z., et al. (2024). A natural language processing system for the efficient extraction of cell markers. *Sci. Rep.* 14:21183. doi: 10.1038/s41598-024-72204-6

Cherkaoui, A., Renzi, G., Martischang, R., Harbarth, S., Vuilleumier, N., and Schrenzel, J. (2020). Impact of Total Laboratory automation on turnaround times for urine cultures and screening specimens for MRSA, ESBL, and VRE carriage: retrospective comparison with manual workflow. *Front. Cell. Infect. Microbiol.* 10:552122. doi: 10.3389/fcimb.2020.552122

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 15:20170387. doi: 10.1098/rsif.2017.0387

Chu, W. T., Reza, S. M. S., Anibal, J. T., Landa, A., Crozier, I., Bagci, U., et al. (2023). Artificial intelligence and infectious disease imaging. *J. Infect. Dis.* 228, S322–S336. doi: 10.1093/infdis/jiad158

Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433. doi: 10.1128/MMBR.00016-10

de Kraker, M. E., Stewardson, A. J., and Harbarth, S. (2016). Will 10 million people die a year due to Antimicrobial Resistance by 2050? *PLoS Med.* 13:e1002184. doi: 10.1371/journal.pmed.1002184

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* 47, 20–33. doi: 10.1016/j.jhealeco.2016.01.012

Erlich, Y., and Narayanan, A. (2014). Routes for breaching and protecting genetic privacy. *Nat. Rev. Genet.* 15, 409–421. doi: 10.1038/nrg3723

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., et al. (2021). Deep learning-enabled medical computer vision. *NPJ Digit. Med.* 4:5. doi: 10.1038/s41746-020-00376-2

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056

Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., and Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science* 363, 1287–1289. doi: 10.1126/science.aaw4399

Fleming, A. (2001). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. Influenzae. 1929. *Bull. World Health Organ.* 79, 780–790.

Ganjalizadeh, V., Meena, G. G., Stott, M. A., Hawkins, A. R., and Schmidt, H. (2023). Machine learning at the edge for AI-enabled multiplexed pathogen detection. *Sci. Rep.* 13:4744. doi: 10.1038/s41598-023-31694-6

Gupta, S., and Gupta, A. (2019). Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review. *Procedia Computer Science* 161, 466–474. doi: 10.1016/j.procs.2019.11.146

Hassan, H., Ren, Z., Zhao, H., Huang, S., Li, D., Xiang, S., et al. (2022). Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks. *Comput. Biol. Med.* 141:105123. doi: 10.1016/j.compbiomed.2021.105123

He, Z., Deng, Y., Van Nostrand, J. D., Tu, Q., Xu, M., Hemme, C. L., et al. (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J.* 4, 1167–1179. doi: 10.1038/ismej.2010.46

Hu, K., Meyer, F., Deng, Z. L., Asgari, E., Kuo, T. H., Munch, P. C., et al. (2024). Assessing computational predictions of antimicrobial resistance phenotypes from microbial genomes. *Brief. Bioinform.* 25:bbae206. doi: 10.1093/bib/bbae206

Ikeda, Y., Keese, S. M., Fenton, W. A., and Tanaka, K. (1987). Biosynthesis of four rat liver mitochondrial acyl-CoA dehydrogenases: in vitro synthesis, import into mitochondria, and processing of their precursors in a cell-free system and in cultured cells. *Arch. Biochem. Biophys.* 252, 662–674. doi: 10.1016/0003-9861(87)90072-5

Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17:239. doi: 10.1186/s13059-016-1103-0

Jiang, Y., Luo, J., Huang, D., Liu, Y., and Li, D. D. (2022). Machine learning advances in microbiology: a review of methods and applications. *Front. Microbiol.* 13:925454. doi: 10.3389/fmicb.2022.925454

Kalantar, K. L., Carvalho, T., de Bourcy, C. F. A., Dimitrov, B., Dingle, G., Egger, R., et al. (2020). IDseq-an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* 9:giaa111. doi: 10.1093/gigascience/giaa111

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., and Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: techniques and applications. *Comput. Biol. Med.* 158:106848. doi: 10.1016/j.compbiomed.2023.106848

Knights, D., Costello, E. K., and Knight, R. (2011). Supervised classification of human microbiota. *FEMS Microbiol. Rev.* 35, 343–359. doi: 10.1111/j.1574-6976.2010.00251.x

Kothari, A., Morgan, M., and Haake, D. A. (2014). Emerging technologies for rapid identification of bloodstream pathogens. *Clin. Infect. Dis.* 59, 272–278. doi: 10.1093/cid/ciu292

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 1234–1240. doi: 10.1093/bioinformatics/btz682

Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi: 10.1038/nrg3920

Liu, G., Catacutan, D. B., Rathod, K., Swanson, K., Jin, W., Mohammed, J. C., et al. (2023). Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nat. Chem. Biol.* 19, 1342–1350. doi: 10.1038/s41589-023-01349-8

Ma, Y., Guo, Z., Xia, B., Zhang, Y., Liu, X., Yu, Y., et al. (2022). Identification of antimicrobial peptides from the human gut microbiome using deep learning. *Nat. Biotechnol.* 40, 921–931. doi: 10.1038/s41587-022-01226-0

Mamoshina, P., Vieira, A., Putin, E., and Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Mol. Pharm.* 13, 1445–1454. doi: 10.1021/acs.molpharmaceut.5b00982

Martin, K. D., and Zimmermann, J. (2024). Artificial intelligence and its implications for data privacy. *Curr. Opin. Psychol.* 58:101829. doi: 10.1016/j.copsyc.2024.101829

Marx, V. (2013). Biology: the big challenges of big data. *Nature* 498, 255–260. doi: 10.1038/498255a

Miller, D., Stern, A., and Burstein, D. (2022). Deciphering microbial gene function using natural language processing. *Nat. Commun.* 13:5731. doi: 10.1038/s41467-022-33397-4

Obermeyer, Z., and Emanuel, E. J. (2016). Predicting the future – big data, machine learning, and clinical medicine. *N. Engl. J. Med.* 375, 1216–1219. doi: 10.1056/NEJMp1606181

Pandi, A., Adam, D., Zare, A., Trinh, V. T., Schaefer, S. L., Burt, M., et al. (2023). Cell-free biosynthesis combined with deep learning accelerates de novo-development of antimicrobial peptides. *Nat. Commun.* 14:7197. doi: 10.1038/s41467-023-42434-9

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36, 983–987. doi: 10.1038/nbt.4235

Prestinaci, F., Pezzotti, P., and Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathog. Glob. Health* 109, 309–318. doi: 10.1179/2047773215Y.0000000030

Price, W. N. 2nd, and Cohen, I. G. (2019). Privacy in the age of medical big data. *Nat. Med.* 25, 37–43. doi: 10.1038/s41591-018-0272-7

Raghupathi, W., and Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* 2:3. doi: 10.1186/2047-2501-2-3

Rajkomar, A., Dean, J., and Kohane, I. (2019). Machine learning in medicine. *N. Engl. J. Med.* 380, 1347–1358. doi: 10.1056/NEJMra1814259

Ranjbar, R., and Alam, M. (2023). Antimicrobial Resistance Collaborators (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Evid. Based Nurs.* 27:16.

Saha, S., Ramesh, A., Kalantar, K., Malaker, R., Hasanuzzaman, M., Khan, L. M., et al. (2019). Unbiased metagenomic sequencing for pediatric meningitis in Bangladesh reveals Neuroinvasive chikungunya virus outbreak and other unrealized pathogens. *MBio* 10:10. doi: 10.1128/mBio.02877-19

Santos-Junior, C. D., Torres, M. D. T., Duan, Y., Rodriguez Del Rio, A., Schmidt, T. S. B., Chong, H., et al. (2024). Discovery of antimicrobial peptides in the global microbiome with machine learning. *Cell* 187, 3761–3778.e16. doi: 10.1016/j.cell.2024.05.013

Smith, K. P., and Kirby, J. E. (2020). Image analysis and artificial intelligence in infectious disease diagnostics. *Clin. Microbiol. Infect.* 26, 1318–1323. doi: 10.1016/j.cmi.2020.03.012

Tahir, M. W., Zaidi, N. A., Rao, A. A., Blank, R., Vellekoop, M. J., and Lang, W. (2018). A fungus spores dataset and a convolutional neural network based approach for fungus detection. *IEEE Trans. Nanobioscience* 17, 281–290. doi: 10.1109/TNB.2018.2839585

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 25, 44–56. doi: 10.1038/s41591-018-0300-7

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P. T.* 40, 277–283.

Viswanathan, V. S., Toro, P., Corredor, G., Mukhopadhyay, S., and Madabhushi, A. (2022). The state of the art for artificial intelligence in lung digital pathology. *J. Pathol.* 257, 413–429. doi: 10.1002/path.5966

Wainberg, M., Merico, D., Delong, A., and Frey, B. J. (2018). Deep learning in biomedicine. *Nat. Biotechnol.* 36, 829–838. doi: 10.1038/nbt.4233

Wang, H., Ceylan Koydemir, H., Qiu, Y., Bai, B., Zhang, Y., Jin, Y., et al. (2020). Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light Sci. Appl.* 9:118. doi: 10.1038/s41377-020-00358-9

Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., et al. (2018). Clinical information extraction applications: a literature review. *J. Biomed. Inform.* 77, 34–49. doi: 10.1016/j.jbi.2017.11.011

Weis, C., Cuenod, A., Rieck, B., Dubuis, O., Graf, S., Lang, C., et al. (2022). Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat. Med.* 28, 164–174. doi: 10.1038/s41591-021-01619-9

Whiley, H., and Taylor, M. (2016). Legionella detection by culture and qPCR: comparing apples and oranges. *Crit. Rev. Microbiol.* 42, 65–74. doi: 10.3109/1040841X.2014.885930

Wouters, O. J., McKee, M., and Luyten, J. (2020). Estimated Research and Development investment needed to bring a new medicine to market, 2009-2018. *JAMA* 323, 844–853. doi: 10.1001/jama.2020.1166

Yasir, M., Karim, A. M., Malik, S. K., Bajaffer, A. A., and Azhar, E. I. (2022). Application of decision-tree-based machine learning algorithms for prediction of Antimicrobial Resistance. *Antibiotics (Basel)* 11:11. doi: 10.3390/antibiotics11111593

Zhang, D., Bi, H., Liu, B., and Qiao, L. (2018). Detection of pathogenic microorganisms by microfluidics based analytical methods. *Anal. Chem.* 90, 5512–5520. doi: 10.1021/acs.analchem.8b00399

Zhang, J., Li, C., Yin, Y., Zhang, J., and Grzegorzek, M. (2023). Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. *Artif. Intell. Rev.* 56, 1013–1070. doi: 10.1007/s10462-022-10192-7

Zhou, K., Xiao, Z., Liu, Q., Wang, X., Huo, J., Wu, X., et al. (2024). Comprehensive application of AI algorithms with TCR NGS data for glioma diagnosis. *Sci. Rep.* 14:15361. doi: 10.1038/s41598-024-65305-9

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 03 October 2024
DOI 10.3389/fmicb.2024.1483052

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Changyu Wu,
Xuzhou Medical University, China
Jin Gu,
Southwest Jiaotong University, China
Yina Wang,
Nanjing Forestry University, China

*CORRESPONDENCE
Yanyun Jiang
✉ yanyun.jiang@qq.com

RECEIVED 19 August 2024
ACCEPTED 18 September 2024
PUBLISHED 03 October 2024

CITATION
Pan N, Mi X, Li H, Ge X, Sui X and Jiang Y (2024) WSSS-CRAM: precise segmentation of histopathological images via class region activation mapping.
Front. Microbiol. 15:1483052.
doi: 10.3389/fmicb.2024.1483052

WSSS-CRAM: precise segmentation of histopathological images via class region activation mapping

Ningning Pan, Xiangyue Mi, Hongzhuang Li, Xinting Ge, Xiaodan Sui and Yanyun Jiang*

Shandong Normal University, Jinan, China

Introduction: Fast, accurate, and automatic analysis of histopathological images using digital image processing and deep learning technology is a necessary task. Conventional histopathological image analysis algorithms require the manual design of features, while deep learning methods can achieve fast prediction and accurate analysis, but rely on the drive of a large amount of labeled data.

Methods: In this work, we introduce WSSS-CRAM a weakly-supervised semantic segmentation that can obtain detailed pixel-level labels from image-level annotated data. Specifically, we use a discriminative activation strategy to generate category-specific image activation maps via class labels. The category-specific activation maps are then post-processed using conditional random fields to obtain reliable regions that are directly used as ground-truth labels for the segmentation branch. Critically, the two steps of the pseudo-label acquisition and training segmentation model are integrated into an end-to-end model for joint training in this method.

Results: Through quantitative evaluation and visualization results, we demonstrate that the framework can predict pixel-level labels from image-level labels, and also perform well when testing images without image-level annotations.

Discussion: Future, we consider extending the algorithm to different pathological datasets and types of tissue images to validate its generalization capability.

KEYWORDS

histopathological image, precise semantic segmentation, weakly-supervised method, category-specific image activation maps, deep learning

1 Introduction

Cancer is a leading cause of death worldwide, with increasing incidence and mortality rates, and high treatment costs that impose a heavy burden on families and society (Sung et al., 2021; Ferlay et al., 2021). Histopathological slides are the gold standard for cancer diagnosis, providing not only basic information on tumor grading and subtype classification but also a wealth of information about the tumor microenvironment (TME). This not only plays a crucial role in explaining tumor development and metastasis but also in influencing the treatment outcomes and prognosis of cancer patients. Recent studies have found that the spatial organization of different tissues and cells is highly correlated with tumor progression, and TME features can reveal gene expression in biological pathways (Wang et al., 2020). Therefore, there is an urgent need for detailed segmentation of different tissues for further clinical research.

Clinically, histopathological slides are visually inspected by pathologists and evaluated semi-quantitatively, and the diagnostic results are reflected in the pathology report. Quantitative assessment for research purposes requires manual annotation by pathologists.

However, the reproducibility and consistency of manual segmentation have been questioned due to inter-observer annotation differences and inter-observer variability (Wang et al., 2020). Due to the specific data storage format and large size of histopathological slides, specific tools need to be used for viewing and labeling, such as QuPath (Bankhead et al., 2017), which makes data annotation work difficult. In addition, manual annotation is very time-consuming and labor-intensive, requiring several days for detailed segmentation of each histopathological slide. Therefore, public research on histopathological image segmentation is usually limited to partial areas of pathological slides, or uses classification methods to achieve segmentation-like effects on whole-slice histopathological images (Lu et al., 2021; Yan et al., 2022; Pan et al., 2023), with very few studies focusing on tissue segmentation in whole-slide histopathological images (Cardenas et al., 2019; Amgad et al., 2022; Chan et al., 2019).

Therefore, it is imperative to develop fast and efficient methods for the rapid, accurate, and consistent delineation of target tissue areas. Semantic segmentation is a fundamental task in computer vision, and deep learning-based automatic segmentation frameworks have shown remarkable performance in medical image segmentation tasks (Hesamian et al., 2019; Xun et al., 2022), achieving outstanding results in various competitions. Popular models for this task include FCN (Long et al., 2015), U-Net (Ronneberger et al., 2015), V-Net (Milletari et al., 2016), nnU-Net (Isensee et al., 2021), among others. Furthermore, other hybrid models have also demonstrated excellent performance in medical image segmentation (Jin et al., 2021; Leube et al., 2023; He et al., 2023).

However, there are two major challenges in using deep-learning-based segmentation algorithms for histopathological image analysis tasks: (1) the performance of deep learning models heavily relies on the quality and quantity of annotated data, and histopathological image data is difficult to annotate, with pixel-level annotation being even more challenging; (2) tumors from different regions exhibit specificity, resulting in high costs for the transfer learning of trained networks.

Although high-quality pixel-level annotation data is scarce, coarse-grained or image-level annotation data is readily available. In fact, for the problem of analyzing histopathological images, there are publicly available datasets that can be downloaded and used for research, such as TCGA,[1] which contains tumor and normal tissues from over 11,000 patients. The database provides image-level descriptions of entire tissue pathology slides and corresponding genomic sequencing results. To reduce the need for pixel-level annotated images during model training, researchers have proposed semi-supervised and weakly supervised learning models, which attempt to improve the model's performance by providing unlabeled or image-level annotated data and hoping to improve the model's generalization ability.

Drawing inspiration from weakly-supervised deep learning methods, we propose a weakly-supervised segmentation algorithm based on Class Region Activation Maps (CRAM) for tissue region segmentation in histopathological images. The framework utilizes image-level annotations to obtain Class Activation Maps (CAM) as pseudo-labels for semantic segmentation. The algorithm can be summarized into two main steps: (1) Obtain the CRAM: using a deep learning classification model, high-quality pixel-level pseudo-labels are generated based on image-level labels. (2) Train a segmentation model: the pixel-level pseudo-labels generated in step (1) are used as ground truth for model training. However, salient region activation can exhibit a higher response to a single class, while typically, multiple classes are present in one region of a pathological image. Therefore, this paper uses a Discriminative Activation (DA) layer to generate specific category masks for foreground and background, which serve as initial segmentation responses. To further increase the reliability of the pseudo-labels, this paper introduces a joint training method by merging the two steps into an end-to-end model. Furthermore, a joint loss function is adopted to optimize both branches and then improves the pseudo-labels' quality. Furthermore, an additional Conditional Random Field (CRF) operation is performed on the activation regions, which are modified into more reliable regions as pseudo-labels.

This approach primarily focuses on whole-slide images (WSI) of lung adenocarcinoma stained with H&E. The research dataset is sourced from the WSSS4LUAD[2] challenge dataset, with the goal of achieving pixel-level segmentation for normal tissue, tumor epithelium, and tumor-associated stroma within the histopathological sections. Figure 1 presents image patches extracted from whole-slide pathology images of lung adenocarcinoma, scanned at a resolution of $0.2517\mu m/pixel$ and 40× magnification. Corresponding segmentation labels for the three prevalent tissue types are also provided. As depicted, these three tissue types may simultaneously appear within a single image patch, particularly tumor epithelium and tumor-associated stroma, since tumor cells often adhere to the stroma. Thus, tumors and stroma frequently coexist in the same image patch. Figure 2 displays examples from the training dataset, where each image patch is annotated with image-level labels indicating the presence of tumor, stroma, and normal tissue. The training dataset encompasses a total of 10,091 image patches. A comprehensive description of the dataset is presented in Section 4.1 of this paper.

Our main contributions are illustrated as follows: (1) Proposing a WSSS-CRAM that improves the traditional CAM method by activating corresponding regions for each class in the image, effectively utilizing the supervisory information of image-level labels. (2) Integrating the steps of obtaining pseudo-labels and training the segmentation model into an end-to-end model for joint training. (3) Performing additional post-processing on the activation regions, using a CRF operation to modify the activation regions into more reliable pseudo-label regions.

2 Related work

This paper centers on the main research subject of semantically segmenting tissue in lung adenocarcinoma. The pertinent techniques predominantly center on semi-supervised segmentation methods based on CAM. Therefore,

---

[1] https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

[2] https://wsss4luad.grand-challenge.org/

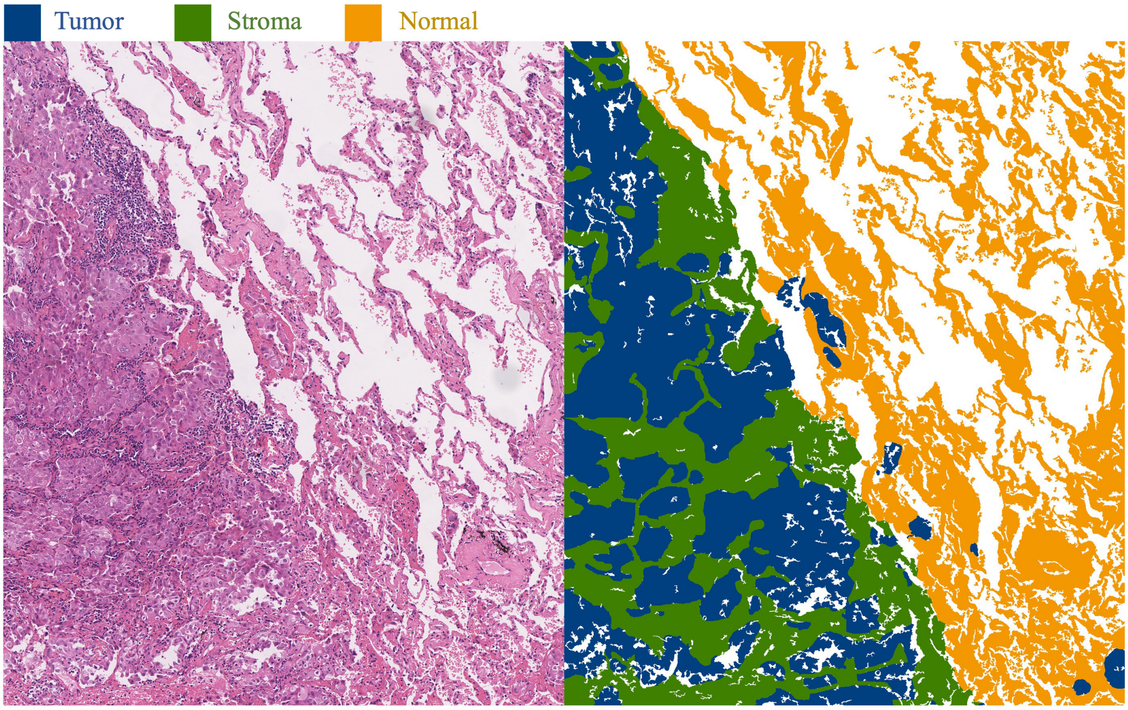


FIGURE 1
Histopathological images of lung adenocarcinoma tissue and their segmentation illustration. The blue area in the image represents the tumor region, the green area represents the stroma region, and the yellow area represents the normal region.

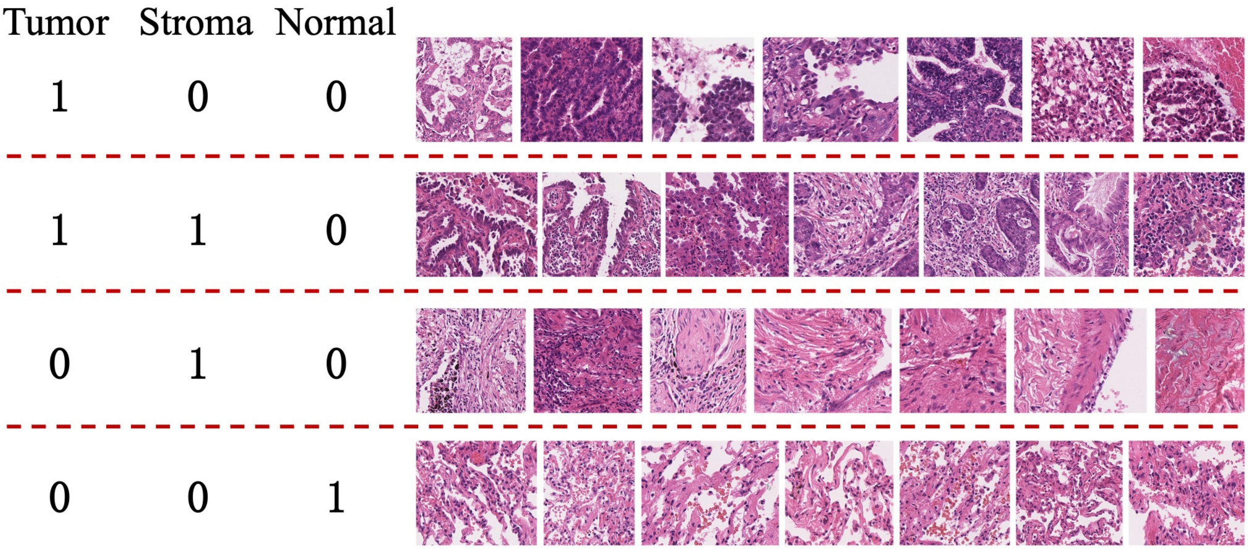


FIGURE 2
Examples from the training set of the WSSS4LUAD Challenge. 1 indicates the presence of the tissue in the image, while 0 indicates the absence of the tissue in the image. **Top row**: Tumor region; **Second row**: Tumor and stroma region; **Third row**: Stroma region; **Fourth row**: Normal region.

before delving into the specifics of the methods, we initially introduce the task of region segmentation in histopathological images of tissues. Following that, we offer a concise analysis of pertinent research concerning semi-supervised segmentation methods.

2.1 Histopathological image segmentation

Since the emergence of whole-slide pathology scanning techniques, the utility of whole-slide tissue pathology imaging has been confirmed across various applications within the realm of pathology. Digitized tissue pathology images have facilitated tasks including remote expert consultations, prognostic analysis, and tumor biomarker assessment (Kumar et al., 2020). As scanning technologies and computational capacities have advanced, significant strides have also been made in the domain of tissue pathology image segmentation. Early approaches entailed manual feature extraction, employing models such as support vector machines and Bayesian models for the segmentation of tissue pathology images. For example, Hiary et al. (2013) employed a Bayesian model to automatically segment stromal tissue in breast tissue pathology images, leveraging color and texture attributes. With the advancement of deep learning techniques, the remarkable performance exhibited by deep learning in image segmentation has prompted its application in the segmentation of tissue pathology images. Among these techniques, FCN and U-Net have emerged as the most frequently employed foundational architectures. For instance, Chen et al. (2017a) introduced the utilization of a Deep Contour-Aware Network (DCAN) for the segmentation of colonic glands. This model incorporated auxiliary supervision mechanisms to tackle the challenge of gradient vanishing during training (Chen et al., 2017a). This approach secured the first rank in the 2015 MICCAI Gland Segmentation Challenge and the 2015 MICCAI Nuclei Segmentation Challenge. Oskal et al. (2019) employed a U-Net-based architecture to achieve a positive predictive value of 0.89 ± 0.16 and sensitivity of 0.92 ± 0.1 in epidermal or non-epidermal pixel classification tasks. In recent years, semi-supervised methods have also gradually been employed in tissue pathology image segmentation tasks to address the issue of limited annotated data (Jin et al., 2022).

Moreover, in recent years, various international competitions have introduced challenges related to the analysis of tissue pathology regions. For instance, the Digestive-System Pathological Detection and Segmentation Challenge (DigestPath 2019) held within MICCAI 2019 (Da et al., 2022; Li et al., 2019) was centered around automating the segmentation of benign and malignant regions within complete tissues. The Multi-organ Nuclei Segmentation and Classification Challenge (MoNuSAC) (Verma et al., 2021) in ISBI 2020 encompassed the identification and segmentation of multiple cell types across four organs. Additionally, the AGGC 2022 (Automated Gleason Grading Challenge) within MICCAI 2022 addressed the automatic segmentation of five tissue types in prostate cancer whole-slide pathology images.

2.2 Weakly-supervised semantic segmentation utilizing CAM

Instance segmentation, one of the most challenging problems in computer vision, has undergone extensive research (He et al., 2017; Arnab and Torr, 2017; Liu et al., 2018). However, many of these studies necessitate manual annotation of instance masks to provide strong supervision, thereby constraining their utility on datasets with sparsely annotated structures. Semi-supervised and weakly supervised instance segmentation strategies strive to transcend this constraint. In scenarios involving solely image-level categories, synthetic labels extracted from class response maps are harnessed to train networks for paired semantic segmentation (Ahn and Kwak, 2018). Employing a classification model to derive CAM stands as a standardized process for generating pseudo masks in the realm of Weakly Supervised Semantic Segmentation (WSSS).

2.2.1 Class activation maps

The Vanilla CAM approach initially scales the feature map using fully connected weights learned for each individual class. Subsequently, seed masks are generated through channel averaging, spatial normalization, and thresholding (Zhou et al., 2016). The GAIN model applies CAM to the original image for mask generation, minimizing model prediction scores to capture features beyond the prior step's activation map in successive training rounds. This gradually refines the activated regions, ensuring complete coverage of the target area (Li et al., 2018). Recently emerged erase-based approaches also embrace similar principles (Zhang et al., 2018; Kweon et al., 2021). The distinction lies in their direct erasure of seed regions in CAM, followed by inputting the erased image into the model to generate the next round's CAM, expected to capture new regions. Moreover, certain schemes have been proposed to optimize CAM. For instance, in Qin et al. (2022), Activation Modulation and Recalibration Scheme (AMR) employs channel/spatial attention mechanisms for fine-tuning activation area calibration, thereby achieving adaptive modulation for segmentation-oriented activation responses. The ReCAM strategy reactivates CAM activation regions using Softmax Cross-Entropy Loss (SCL), resulting in ReCAM with Binary Cross-Entropy (BCE) constraints (Chen et al., 2022). Embedded Discriminative Attention Mechanism (EDAM) is a recent endeavor that employs CAM-based perturbations to optimize an additional classifier. It employs an extra DA layer to generate class-specific masks (Wu et al., 2021).

2.2.2 Generation of pseudo-labels

The seed masks generated from CAM or its variations can undergo refinement steps to enhance the quality of pseudo-labels, employing both non-learning-based and learning-based methods. SEC introduced the principles of Seed, Expand, and Constrain for refining CAM, which have been widely adopted by subsequent works (Kolesnikov and Lampert, 2016). Among these, CRF is an earlier post-processing method that is user-friendly, independent of features extracted by the trained model, and relies solely on the original image features. DSRG, inspired by Seeded Region Growing (SRG), employs CAM as seeds to expand regions of interest (Huang et al., 2018). This approach integrates the SRG process into the deep segmentation network, deviating from the previous strategy of training segmentation models using pseudo-labels generated through SRG.

Learning-based methods introduce additional network modules. For example, AffinityNet employs a deep neural network to predict semantic affinities between adjacent image coordinates, achieving semantic propagation through random walks (Ahn and Kwak, 2018). IRNet estimates rough regions of individual instances and detects boundaries between different object classes. It focuses on pixel relations on the graph and computes affinities based on these relations (Ahn et al., 2019). Furthermore, incorporating confidence regions from saliency maps into CAM for pseudo-label refinement has become a common practice in recent methodologies (Chen et al., 2022; Wu et al., 2021). Approaches like OOA (Jiang et al., 2019) and CONTA (Zhang et al., 2020b) integrate CAM inferences generated through multiple training iterations, directing attention accumulation toward various parts of objects.

3 Methodology

In this section, the main focus is on introducing the CRAM algorithm framework. We provide a comprehensive explanation of the CNN-based pseudo-label acquisition module, the target semantic segmentation module, and the employed loss functions in the algorithm.

3.1 Framework

The foundational model for the CAM-based semi-supervised segmentation algorithm used in this paper is divided into two distinct steps: pseudo-label acquisition and independent segmentation model training modules, as depicted in Figure 3. The pseudo-label acquisition module utilizes a standard image classification network supervised by image-level labels. By accentuating response areas of image-level labels through CAM, it generates pixel-level masks corresponding to each image, serving as pseudo-labels for the semantic segmentation module. The semantic segmentation module can be any end-to-end segmentation network, using the pixel-level pseudo-labels generated by the pseudo-label acquisition module as actual labels for training the model. During inference, segmentation predictions can be achieved solely by utilizing the semantic segmentation module.

The model presented in this paper is based on the algorithm outlined in Figure 3 and is divided into two primary modules: the pseudo-label acquisition module and the semantic segmentation module. Differing from the majority of previous methodologies that adopt independent two-step procedures, this paper amalgamates pseudo-label acquisition and semantic segmentation into a cohesive end-to-end model for joint training. As illustrated in Figure 4, following feature extraction by a backbone network, the image is directed to both the pseudo-label acquisition module and the semantic segmentation module. The integrated model is subject to joint training via a full loss function.

The pseudo-label acquisition module: Within this module, the model incorporates a DA layer to extract category-specific activation regions. Unlike CAM, which employs a single activation map for classification, the DA layer generates category-specific activation maps for each category. These category-specific activation maps are fused with the original feature layer to derive category-specific feature maps. The self-supervised layer explores collaborative information within and across images in a batch. Ultimately, classification predictions are made based on the collaborative information corresponding to each image. Given that all images in the training set are associated with image-level labels, a binary cross-entropy (BCE) loss function is employed independently for each category.

The semantic segmentation module: This module initially refines the feature maps extracted from the backbone network through a series of convolutional layers. Subsequently, an independent CRF is employed to enhance the category-specific activation maps obtained from the pseudo-label acquisition module. This refinement process helps eliminate mislabeled pixels, resulting in comparatively reliable pseudo-labels. The target semantic segmentation module applies cross-entropy loss and energy loss to the confident and non-confident regions of the pseudo-labels, respectively.

Joint loss function: The loss function is used to supervise the optimization of parameters within the model. In the presented algorithm, the classification and segmentation models are integrated into an end-to-end framework for joint training. As a result, the overall loss function comprises a binary cross-entropy loss for classification, as well as cross-entropy loss and energy loss for segmentation.

3.2 Pseudo-label acquisition based on CNN classification model

3.2.1 Discriminative activation layer

For a given batch of data $X = \left\{\left(x_n, l_n\right)\right\}_N$, where N represents the number of mini-batches, x_n represents the n-th image in this batch, and l_n represents the corresponding class label. It should be noted that l_n is represented as $\{0, 1\}^K$, indicating image-level labels corresponding to K categories. Backbone network extracts the feature map $F_n \in \mathbf{R}^{C\times H\times W}$ corresponding to image l_n, where C represents the number of channels in the feature map, and H and W represent the height and width of the feature map. Connect the DA layer to generate activation maps $M_n \in \mathbf{R}^{(K+1)\times H\times W}$ corresponding to K target categories. To explicitly represent the background region, in addition to generating activation maps for each category, the DA layer also generates activation maps corresponding to the background.

Applying L2-norm regularization to the activation map M_n can generate pixel-level probabilities for the corresponding class or background:

$$\widehat{M}_n(i,j) = L2 - \text{norm}\left(\left|M_n(i,j)\right|\right). \tag{1}$$

After the L2-norm regularization operation, $\widehat{M}_n(i,j)$ represents the pixel-level class probability distribution at position (i,j), and $\widehat{M}_n^k(i,j)$ represents the probability corresponding to class k at position (i,j). Through the above operations, activation maps corresponding to each category in the image are obtained.

3.2.2 Self-supervised layer

Combining the feature map $F_n \in \mathbf{R}^{C\times H\times W}$ corresponding to image l_n with the activation map $\widehat{M}_n^k(i,j)$ corresponding to K target categories, generates feature maps for each class:

$$F_n^k = F_n \cdot \widehat{M}_n^k, \tag{2}$$

where F_n^k is the feature map corresponding to category k in the image l_n.

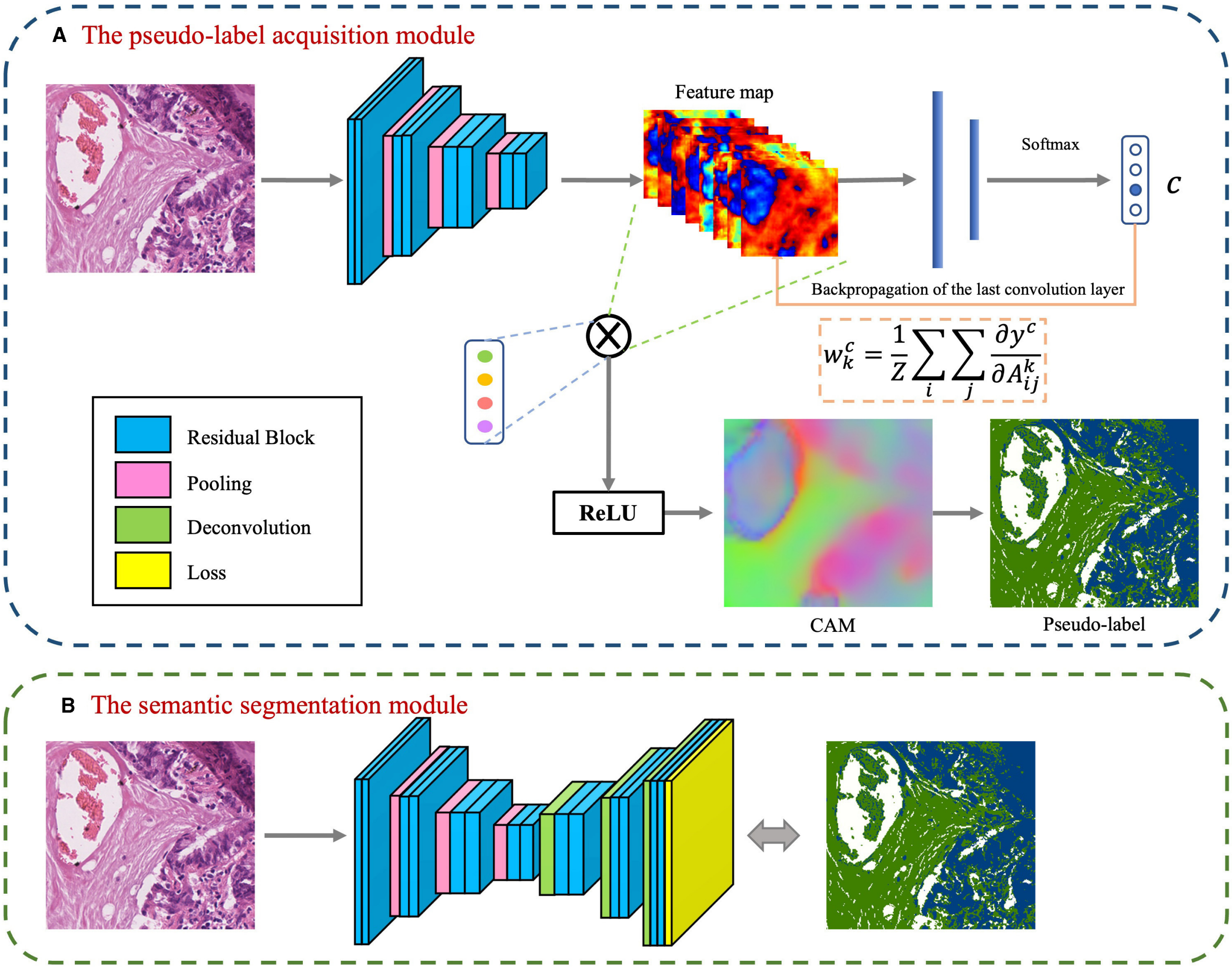


FIGURE 3
Basic conceptual diagram of CAM-based semi-supervised segmentation algorithm, which contains two main modules: **(A)** The pseudo-label acquisition module; **(B)** The semantic segmentation module.

For a batch of B images, the corresponding feature maps are represented as $\mathcal{F}^k = \left[F_1^k, F_2^k, \ldots F_B^k\right] \in \mathbf{R}^{B\times C\times H\times W}$. After a 1×1 convolution, the feature maps are transformed into activation features $\hat{\mathcal{F}}^k \in \mathbf{R}^{1\times(B\times C\times H)\times d}$ corresponding to each category. The combination of activation maps with the initial feature maps is used to explore collaborative information specific to category activation maps. The self-supervised layer simultaneously considers feature attention within and between images in a batch, making the exploration of collaborative information more effective. The model generates category-specific feature maps for each category, using global average pooling and employing a specific classifier for label prediction of the given category. Since in histopathological images, one image often corresponds to multiple image categories, to make the activation regions corresponding to categories more effective, this paper transforms the multi-class problem into multiple binary classification problems.

The purpose of the self-supervised layer is to highlight similar regions in the activation maps corresponding to images in a batch through self-attention mechanisms, to obtain better activation maps for each category.

3.2.3 Classification loss function

The category-specific features output by the self-supervised layer are mapped to categories through a fully connected layer, with image-level labels corresponding to the image as supervision. The classification loss function is represented as:

$$\mathcal{L}_{class} = \frac{1}{B\times K}\sum_{n=1}^{B}\sum_{k=1}^{K}\mathcal{L}_{BCE}\left(Linear\left(GAP\left(A_n^k\right)\right), l_n^k\right), \quad (3)$$

where

$$\left[A_1^k, A_2^k, \ldots, A_B^k\right] = \mathrm{SelfAttention}\left(\hat{\mathcal{F}}^k\right), \quad (4)$$

where A_n^k is the activation map corresponding to input image x_n after the self-supervised layer for the k-th category, $l_n^k \in [0,1]$ represents the true label of input image x_n corresponding to the k-th category. Since the input to the self-supervised layer is a combination of category-specific activation maps and initial feature maps, the loss function of the self-supervised layer will, through

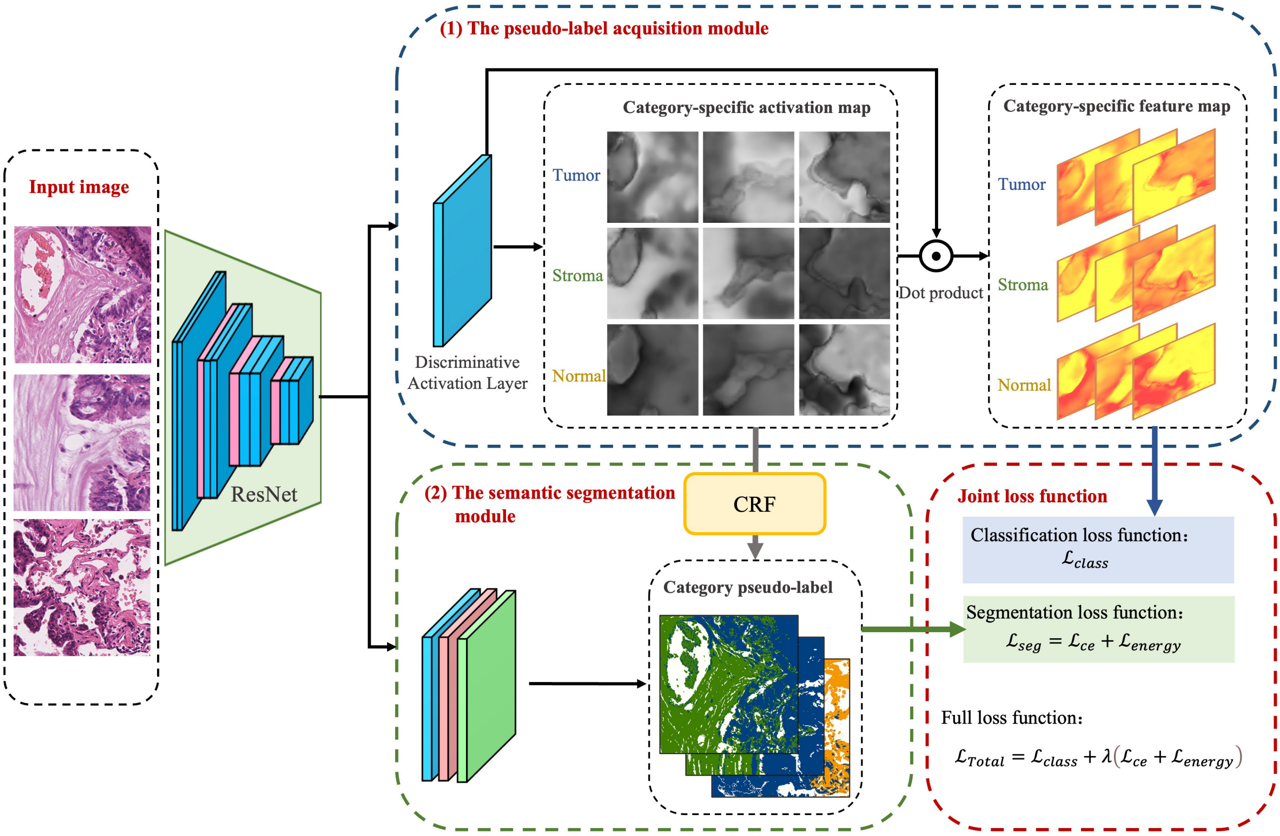


FIGURE 4
Weakly-supervised segmentation algorithm based on category activation regions.

backpropagation, affect the distinguishing activation layers of all foreground categories, thereby influencing the parameter training of the backbone network.

3.3 Target semantic segmentation model

3.3.1 Reliable semantic segmentation labels

From the pseudo-label acquisition module, activation maps corresponding to each category can be obtained, which highlight the regions where each category plays a role in classification. In this activation map, select the high-confidence foreground and background regions as reliable regions, and the remaining regions as unreliable regions. High-confidence maps are represented as:

$$p^r(i,j) = \begin{cases} \widehat{M}^k(i,j), & \text{if } \widehat{M}^k(i,j) < \alpha \text{ or } \widehat{M}^k(i,j) > \beta \\ 255, & \text{else} \end{cases} \quad (5)$$

where α and β represent pre-established thresholds. When the threshold falls below α, it signifies the region as a dependable background area; conversely, when the threshold surpasses β, the region is retained as a foreground area.

We employ CRF for post-processing the activation maps, removing incorrectly labeled pixels, and enhancing the probability maps associated with each category:

$$p^{crf} = CRF(x, \widehat{M}). \quad (6)$$

Taking into account the constraints imposed by CRF on the activation maps, the ultimate pixel-level pseudo-labels are as follows:

$$p^{pseudo}(i,j) = \begin{cases} p^r(i,j), & \text{if } p^r(i,j) = p^{crf}(i,j) \\ 255, & else. \end{cases} \quad (7)$$

If $p^r(i,j) = p^{crf}(i,j)$, signifying alignment between the high-confidence map and the CRF activation map, we retain this region as the confident pseudo-label area, with the rest designated as non-confident pseudo-label areas.

3.3.2 Segmentation loss function

The pseudo-labels generated by the model serve as the ground truth labels for training the semantic segmentation module, encompassing both areas with high-confidence pseudo-labels and areas with low-confidence pseudo-labels.

In the case of confident pseudo-label regions, the model utilizes the standard cross-entropy loss function, denoted as:

$$\mathcal{L}_{ce} = -\sum_{(i,j)\in\varphi} B(i,j)\log\left(P_{net}^{k}(i,j)\right), \tag{8}$$

where $B(i,j)$ is a binary label indicating whether the label belongs to class k. φ represents the confident pseudo-label region, i.e., when $p^{pseudo}(i,j) \neq 255$. $P_{net}^{k}(i,j)$ represents the prediction of the segmentation model.

The model utilizes the dense energy loss function (Zhang et al., 2020a), applied to both confident and non-confident regions, and it is represented as:

$$\mathcal{L}_{energy} = \sum_{i=0,j=0}^{H,W}\ \sum_{\substack{a=0,b=0\\(i,j)\neq(a,b)}}^{H,W} S(i,j)E((i,j),(a,b)), \tag{9}$$

where $S(i,j)$ represents a soft filter. For regions with confident pseudo-labels, soft filter weights are determined based on the model's predicted class probabilities. In contrast, for regions with non-confident pseudo-labels, a dense energy loss is employed. Cross-entropy loss functions are designed for hard labels, while the pseudo-labels used in this study are not guaranteed to be 100% accurate. Therefore, applying the cross-entropy loss directly to confident regions could introduce errors during model training. The dense energy loss function, using a soft labeling strategy for confident regions, allows for further refinement of the confident regions generated in the preceding step. $S(i,j)$ is defined as:

$$S(i,j) = \begin{cases} 1-\max_{k\in K}\left(P_{net}^{k}(i,j)\right), & (i,j)\in\varphi \\ 1, & \text{else} \end{cases} \tag{10}$$

Here, $E((i,j),(a,b))$ represents the energy formula that characterizes the relationship between pixel (i,j) and pixel (a,b):

$$E((i,j),(a,b)) = \sum_{\substack{k_1,k_2\in K\\k_1\neq k_2}} G((i,j),(a,b))P_{net}^{k_1}(i,j)P_{net}^{k_2}(a,b), \tag{11}$$

where $G((i,j),(a,b))$ is a Gaussian filter.

The total loss function associated with the target semantic segmentation network comprises both cross-entropy loss and energy loss:

$$\mathcal{L}_{seg} = \mathcal{L}_{ce} + \mathcal{L}_{energy}. \tag{12}$$

3.4 Joint loss function

The approach presented in this paper integrates classification and segmentation models into an end-to-end framework. The overall loss function comprises the $\mathcal{L}_{class}$ loss function from the pseudo-label acquisition module and the $\mathcal{L}_{seg}$ loss function from the semantic segmentation network. The combined loss function is shown below:

$$\mathcal{L}_{Total} = \mathcal{L}_{class} + \lambda\mathcal{L}_{seg} = \mathcal{L}_{class} + \lambda(\mathcal{L}_{ce} + \mathcal{L}_{energy}), \tag{13}$$

where λ is a weighting coefficient that controls the balance between the pseudo-label acquisition module and the target segmentation module.

3.5 Independent semantic segmentation model

After weakly-supervised training, the combination of the backbone network and the target semantic segmentation network can serve as an independent inference module for generating semantic segmentation results during the testing phase. Alternatively, the model proposed in this paper can be used as a whole for pseudo-label acquisition. During the training phase, the segmentation model outputs optimized region segmentation results, which are used as artificial pseudo-labels for an independent semantic segmentation model.

Define an independent semantic segmentation module: This semantic segmentation module is designed as a standalone component, utilizing pseudo-labels obtained from the previous step's image classification model as training labels for the training model. During the final inference phase, running inference is as simple as using this trained model. The standalone segmentation model can employ any end-to-end semantic segmentation model as its backbone network, such as FCN (Long et al., 2015), U-Net (Ronneberger et al., 2015), DeepLab v3 (Chen et al., 2017b), and so on. In this paper, we draw inspiration from previous research in weakly-supervised segmentation, where the semantic segmentation module combines the ResNet model and the DeepLab v3 model. This network model consists of two parts: an Encoder based on the ResNet model and a Decoder based on the DeepLab v3.

4 Algorithm validation and evaluation

4.1 Datasets

The dataset used in this paper is publicly available data from the WSSS4LUAD challenge (Han et al., 2022a,b), which includes 67 H&E (Hematoxylin and eosin)-stained WSI (Whole Slide Images) from the Guangdong Provincial People's Hospital (GDPH) and 20 WSI images from the TCGA public dataset. These images have annotations for three common and meaningful tissue types: tumor epithelial tissue, stromal tissue, and normal tissue.

The training dataset in this dataset consists of 63 WSI (49 from GDPH and 14 from TCGA), from which 10,091 image patches were cropped and selected. The image size ranges from 150×150 to 300×300. Each image in the training set has image-level annotations in the form of a three-digit label [tumor, stroma, normal]. It includes 6,579 images of tumor tissue, 7,076 images of stromal tissue, and 1,832 images of normal tissue.

The most common label is [1,1,0], indicating images containing both tumor and stroma, with a total of 5,393 images. This is followed by 1,832 images with the [0,0,1] label (indicating normal tissue), 1,680 images with the [0,1,0] label (indicating stromal tissue), and 1,181 images with the [1,0,0] label (indicating tumor tissue).

The validation set comprises 12 WSI (9 from GDPH and 3 from TCGA), from which 40 image patches are cropped. These include 9 large image patches ranging in size from 1,500×1,500 to 5,000×5,000 and 31 small image patches ranging in size from 200×200 to 500×500. The validation dataset has pixel-level labels and is used to validate the trained models.

The test set also consists of 12 WSI (9 from GDPH and 3 from TCGA), from which 80 image patches are cropped. These include 14 large image patches ranging in size from 1,500 × 1,500 to 5,000 × 5,000 and 66 small image patches ranging in size from 200 × 200 to 500 × 500. The test dataset has pixel-level labels and is used for the final model testing.

4.2 Experimental settings

This experiment was conducted in a PyTorch environment, utilizing NVIDIA CUDA (version 11.4) and cuDNN library (version 8.2.2). All experiments were performed on a computer running Ubuntu 20.04 LTS, using 4 NVIDIA Tesla A100 GPUs with 40GB of VRAM each. The model's backbone network was pre-trained on the ImageNet dataset and further fine-tuned on the target dataset used in this paper.

The model used an SGD optimizer with a batch size of 8, an initial learning rate of 0.001, weight decay set to 0.0002, and momentum set to 0.9. Two hyperparameters, α and β, were set to 0.3 and 0.9, respectively.

During both training and testing, a CRF operation was used to generate refined labels, with parameters following the default values as described in Huang et al. (2018). During training, the loss functions computed by the classification and segmentation modules were updated through backpropagation to update the backbone network. During testing, only the segmentation module was used to generate region segmentation corresponding to the images.

Considering the irregular sizes of image patches in this dataset, they were standardized through resizing before being fed into the model. During the training phase, the image dimensions were initially randomly increased to two to three times their original size. Subsequently, these enlarged images were uniformly cropped to a size of 513 × 513 pixels, serving as the input images for the model. In the testing phase, the image dimensions were enlarged to 2.5 times their original size, and the model made predictions and generated segmentation results based on the enlarged images. Due to limitations in GPU VRAM, particularly with extremely large pixel images, they were proactively cropped to a fixed size (ranging from 400 × 400 to 500 × 500 in this paper). The model's predicted results were then combined for visualization purposes.

4.3 Performance evaluation metrics

In the experiments, model evaluation is performed using the mean Intersection over Union (mIoU), which is expressed as follows:

$$mIoU = \frac{1}{k+1}\sum_{k=0}^{k}\frac{TP}{FN+FP+TP} \tag{14}$$

where TP stands for true positives (correctly predicted positive instances), while FN and FP represent false negatives (positive instances incorrectly predicted as negative) and false positives (negative instances incorrectly predicted as positive), respectively. The variable k denotes the number of classes. In our experiments, the test dataset includes a background label. Therefore, when computing the final mIoU, the background region is excluded and not included in the calculation area.

4.4 Model analysis

4.4.1 Comparison with state-of-the-art methods

Table 1 presents a comparison between our proposed method with the existing fully supervised baseline segmentation methods

TABLE 1 Comparison with the state-of-the-art methods.

Model		mIoU	Tumor	Stroma	Normal
Supervised	U-Net (Ronneberger et al., 2015)	0.5362	0.4158	0.7075	0.4854
	ResNet101 (He et al., 2016)	0.5992	0.5312	0.7323	0.5342
	DeepLab v3 (Chen et al., 2017b)	0.6222	0.5859	0.7318	0.5489
Weakly-supervised	ChunhuiLin	**0.8413**	0.8389	0.8919	0.7931
	baseline0412	0.8222	**0.8402**	0.8343	0.7921
	Vison307	0.8058	0.8165	0.8554	0.7456
	WSSS-CRAM1	0.7265	0.7074	0.8125	0.6597
	WSSS-CRAM2	0.7618	0.7493	0.8237	0.7125
	WSSS-CRAM3	0.8401	0.8293	**0.8923**	**0.7987**

Bold values indicates best result obtained for predictions.

and the top three performers in the WSSS challenge, including ChunhuiLin, baseline0412, and Vison307, with the best result highlighted in bold. The fully supervised approach was trained using training data containing only one tissue category, with [1,0,0], [0,0,1], and [0,1,0] corresponding to 1,181, 1,832, and 1,680 images, respectively. Among the comparison weakly-supervised methods, including ChunhuiLin, baseline0412, and Vison307 are semi-supervised methods. Training details can be found in the paper (Han et al., 2022a). WSSS-CRAM1 entails training a model exclusively using image-level labels from the training set, without any reference to pixel-level labels from the validation set throughout the training process. Building upon jointly optimized pseudo-labels, WSSS-CRAM2 establishes a separate segmentation module to learn pixel-level pseudo-labels, the model is shown in Figure 3. In contrast, WSSS-CRAM3 incorporates pixel-level labels from the validation set as a supervisory condition when training a separate segmentation model with pseudo-labels. Notably, our proposed approach, when training a dedicated semantic segmentation module and incorporating pixel-level labels from the validation set into the model training, achieves results differing by a mere 0.0012 from the competition's top performance, indicating a remarkable quantitative proximity. This outcome may be attributed to the omission of weight consideration for pseudo-labels compared to the known labels from the validation set during the model training process.

4.4.2 Ablation experiment

Table 2 presents the results of ablation experiments aimed at demonstrating the effectiveness of our method's design. To maintain control over the variables in these experiments, we focused solely on the acquisition of pseudo-labels. In this process, pseudo-labels obtained from the training dataset were combined with pixel-level annotated labels from the validation data to train

TABLE 2 Ablation experiments for each module in the network.

Joint optimization	CAM	DA	CRF	mIoU
	✓			0.6925
	✓	✓		0.7680
	✓	✓	✓	0.7912
✓	✓			0.7684
✓	✓	✓		0.8059
✓	✓	✓	✓	**0.8401**

Bold values indicates best result obtained for predictions.

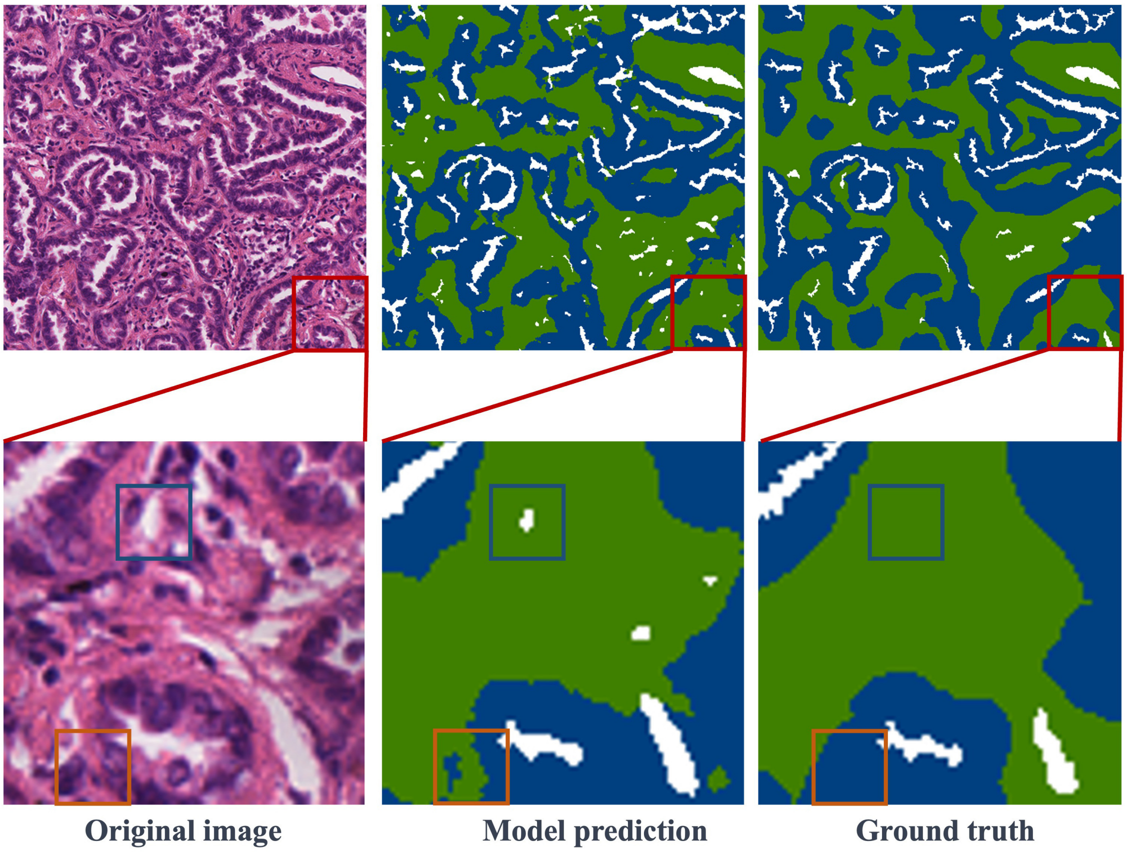


FIGURE 5
Visualization of segmentation results: the first column features the original images, the second column showcases the model's predictions, and the third column reveals the ground truth labels. Notably, red boxes highlight representative regions, which are further magnified in the second row.

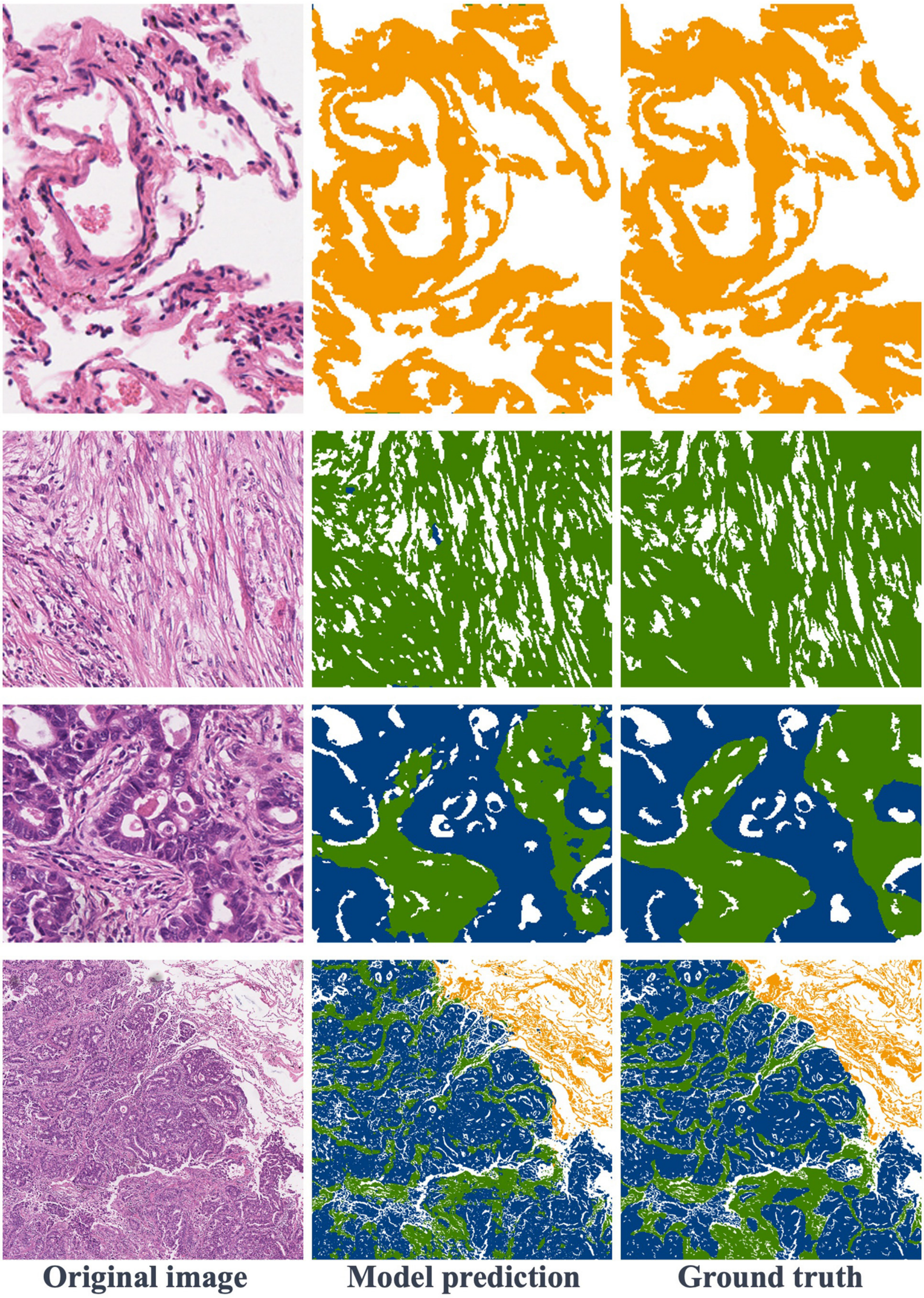


FIGURE 6
Visualization of segmentation results: the first row is normal regions, the second row is stroma regions, the third row includes stroma and tumor, and the fourth row features a large image containing normal, stroma, and tumor areas.

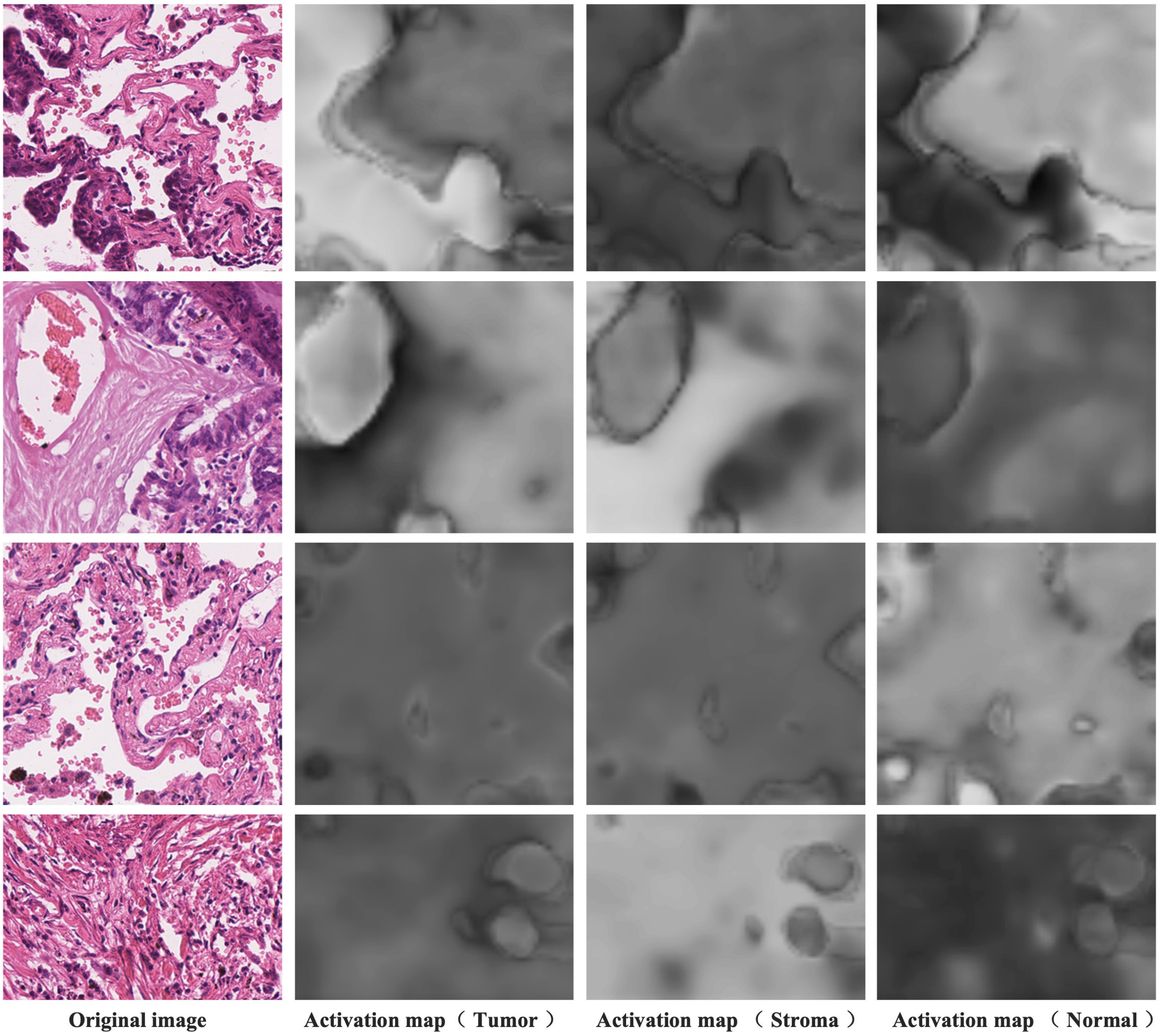


FIGURE 7
Category-specific activation maps corresponding to the tumor, stroma, and normal regions.

separate semantic segmentation modules. It's worth noting that CAM, which serves as the foundational strategy for obtaining pixel-level labels from image-level labels, was included in all ablation models. As observed in the table, the joint optimization of segmentation and classification modules yields a significant improvement in segmentation performance. Furthermore, the strategy of DA layer and CRF also contributes to enhancing segmentation performance.

4.5 Visualized results

4.5.1 Visual presentation of results

Figure 5 presents the segmentation results obtained in the test dataset. The first column contains image blocks extracted from the overall histopathological image, the second column showcases the model's predictions, and the third column displays the ground truth labels. In the second row, specific details from the first-row images have been selectively magnified for closer inspection. The result images clearly demonstrate a close alignment between the model's predictions and the ground truth labels. The trained model exhibits the capability to accurately segment regions within histopathological images of lung adenocarcinoma. In the second row of enlarged images, regions, where the model's predictions deviate from the ground truth labels, are enclosed within blue and orange rectangles. Upon a closer examination of the original images, it becomes apparent that the region inside the blue rectangle corresponds to a blank area in the original image, whereas the region within the orange rectangle should indeed be labeled as stroma, consistent with the ground truth. While this result may differ from the manually annotated ground truth, it may offer a

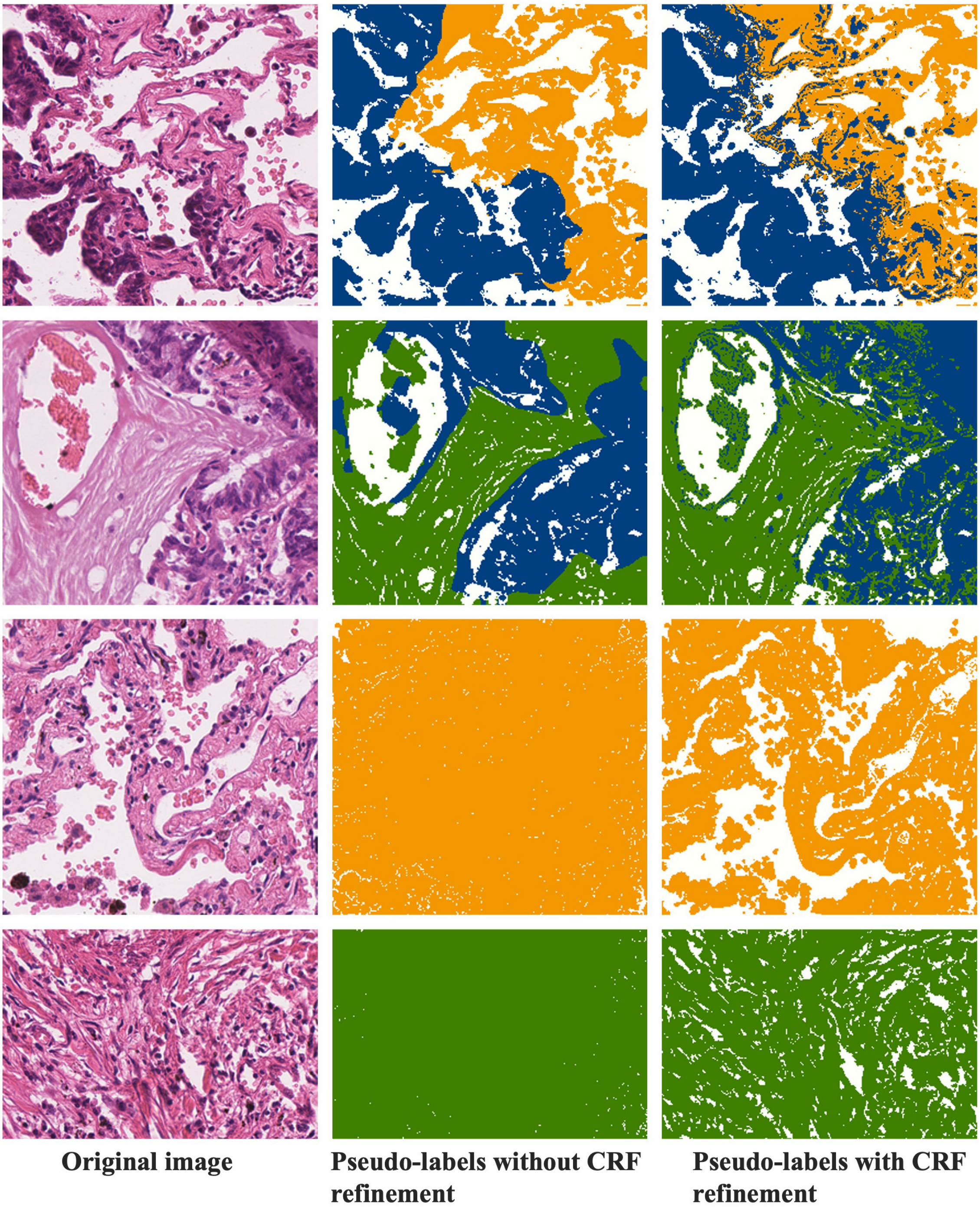


FIGURE 8
Comparison of pseudo-labels before and after CRF operations.

more precise representation of the intricate details in comparison to the human-labeled labels. This comparative analysis indicates that the model not only learns pixel-level annotations from image-level labels but also excels in accurately predicting tissue boundaries and intricate details.

Figure 6 showcases various segmentation examples from the test dataset. In the first column, you'll find the original images, while the second column reveals the model's predictions, and the third column displays the ground truth labels. The result images clearly depict that the first and second rows represent image blocks from normal and stroma regions, respectively. In these cases, the model excels in delivering remarkably accurate predictions that closely align with the ground truth labels. Moving to the third row, we encounter images featuring the coexistence of tumors and stroma. Upon close examination, it becomes apparent that the model also produces relatively precise predictions, with minor boundary prediction errors occurring solely at the edges of the tumor and stroma regions. Finally, in the last row of Figure 6, this is a typical example of a large image block, encompassing tumor, stroma, and normal areas. The model's prediction results affirm that the proposed method consistently yields precise segmentation results, even for intricate histopathological images.

4.5.2 Category-specific activation maps from discriminative activation layer

Figure 7 presents the activation maps generated by the model after differentiating the activation maps from the activation layer output. The data showcased here is sourced from the training dataset and, therefore, lacks corresponding pixel-level annotations. The four examples shown correspond to image-level labels [1,0,1], [1,1,0], [0,0,1], and [0,1,0], representing tumor and normal, tumor and stroma, normal, and stroma, respectively. In the first column, you can see the original images, while the second column displays the activation maps for tumor regions, the third column displays the activation maps for stroma regions, and the fourth column reveals the activation maps for normal regions. Higher brightness in the activation maps indicates a higher probability of the corresponding region belonging to that class. From these images, it's evident that distinguishing the activation layer enables the generation of activation regions corresponding to each class. Remarkably, even without the explicit use of pixel-level annotations during training to inform the model about specific regions as the tumor, stroma, or normal, weakly supervised learning using only image-level labels demonstrates the ability to produce pixel-level activations, showcasing a crucial feature of CAM.

4.5.3 CRF refinement of pseudo-labels

Figure 8 demonstrates the refinement of pseudo-labels through CRF operations. The first column showcases the original images, the second column displays the pseudo-labels before CRF refinement, and the third column reveals the pseudo-labels after CRF refinement. Let's compare the state of the labels before and after CRF operations based on these results. From the examples in the first row, it's evident that the pseudo-labels before CRF refinement exhibit distinct boundaries between tumor and normal regions but overlook individual tumor cells present in the finer details. CRF operations, guided by the original image, rectify these boundaries, resulting in a more precise demarcation between tumor and normal regions. In the second row of examples, it becomes apparent that CRF not only refines details but also corrects more extensive areas of segmentation error. The third and fourth rows represent normal and stromal tissues, and a comparison with Figures 4, 5 reveals that activation maps can emphasize specific classes without clearly defined activation boundaries for the image's boundary details. Consequently, in the pseudo-labels of the second column, only the categories are nearly discernible. After undergoing CRF operations, the distinctions between foreground and background become much clearer.

5 Conclusion

This paper proposes a novel weakly-supervised segmentation method based on class region activation mapping, effectively achieving the segmentation of tissue regions in lung adenocarcinoma pathological images. The paper incorporates distinguishing activation layers and self-supervised layers into the classification network to predict activation maps corresponding to each category in the image and explore inter-image collaborative information. Subsequently, pseudo-labels generated from the activation maps are used as training labels for the target semantic segmentation module. The fusion of the pseudo-label prediction module and the target segmentation module allows for better utilization of pixel-level segmentation of target regions with image-level labels. Experimental results on the test set of a publicly available lung adenocarcinoma dataset validate the performance of the weakly-supervised segmentation algorithm based on category-specific activation. Compared to traditional weakly-supervised semantic segmentation methods based on category activation maps, this algorithm exhibits a significant improvement in segmentation accuracy in the literature.

The algorithm has only been validated on a lung adenocarcinoma dataset. Although the algorithm performs well on the lung adenocarcinoma dataset, its generalization ability to other diseases or types of tissue images has not been verified. Therefore, the method's performance on other image datasets may not be as expected. Future, we consider extending the algorithm to different pathological datasets and types of tissue images to validate its generalization capability. Consider integrating pathological images with other types of medical imaging (e.g., CT, MRI) for multimodal analysis to enhance diagnostic accuracy and the applicability of the model.

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

NP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. XM: Data curation, Formal analysis, Investigation, Project administration, Resources, Supervision, Validation, Writing – original draft. HL: Conceptualization, Data curation, Formal analysis, Project administration, Writing – original draft. XG: Conceptualization, Data curation, Funding acquisition, Resources, Writing – original draft. XS: Conceptualization, Data curation, Formal analysis, Writing – original draft. YJ: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Software, 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. This work was supported by the National Nature Science Foundation of China (62272283); the Natural Science Foundation of Shandong Province (ZR2021QF135); and the "Young Innovation Team Program" of Shandong Provincial University (2022KJ250).

Acknowledgments

The authors gratefully acknowledge the participants who volunteered to help with the present 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

References

Ahn, J., Cho, S., and Kwak, S. (2019). "Weakly supervised learning of instance segmentation with inter-pixel relations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2209–2218. doi: 10.1109/CVPR.2019.00231

Ahn, J., and Kwak, S. (2018). "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4981–4990. doi: 10.1109/CVPR.2018.00523

Amgad, M., Atteya, L. A., Hussein, H., Mohammed, K. H., Hafiz, E., Elsebaie, M. A., et al. (2022). Nucls: a scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer. *GigaScience* 11:giac037. doi: 10.1093/gigascience/giac037

Arnab, A., and Torr, P. H. (2017). "Pixelwise instance segmentation with a dynamically instantiated network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 441–450. doi: 10.1109/CVPR.2017.100

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., et al. (2017). Qupath: open source software for digital pathology image analysis. *Sci. Rep.* 7, 1–7. doi: 10.1038/s41598-017-17204-5

Cardenas, C. E., Yang, J., Anderson, B. M., Court, L. E., and Brock, K. B. (2019). "Advances in auto-segmentation," in *Seminars in Radiation Oncology* (Elsevier), 185–197. doi: 10.1016/j.semradonc.2019.02.001

Chan, L., Hosseini, M. S., Rowsell, C., Plataniotis, K. N., and Damaskinos, S. (2019). "Histosegnet: semantic segmentation of histological tissue type in whole slide images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10662–10671. doi: 10.1109/ICCV.2019.01076

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., and Heng, P.-A. (2017a). Dcan: deep contour-aware networks for object instance segmentation from histology images. *Med. Image Anal.* 36:135–146. doi: 10.1016/j.media.2016.11.004

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017b). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587.*

Chen, Z., Wang, T., Wu, X., Hua, X.-S., Zhang, H., and Sun, Q. (2022). "Class re-activation maps for weakly-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 969–978. doi: 10.1109/CVPR52688.2022.00104

Da, Q., Huang, X., Li, Z., Zuo, Y., Zhang, C., Liu, J., et al. (2022). Digestpath: a benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system. *Med. Image Anal.* 80:102485. doi: 10.1016/j.media.2022.102485

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Pi neros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: an overview. *Int. J. Cancer* 149, 778–789. doi: 10.1002/ijc.33588

Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., et al. (2022a). Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. *Med. Image Anal.* 80:102487. doi: 10.1016/j.media.2022.102487

Han, C., Pan, X., Yan, L., Lin, H., Li, B., Yao, S., et al. (2022b). Wsss4luad: Grand challenge on weakly-supervised tissue semantic segmentation for lung adenocarcinoma. *arXiv preprint arXiv:2204.06455.*

He, A., Wang, K., Li, T., Du, C., Xia, S., and Fu, H. (2023). H2former: an efficient hierarchical hybrid transformer for medical image segmentation. *IEEE Trans. Med. Imaging.* 42, 2763–2775. doi: 10.1109/TMI.2023.3264513

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969. doi: 10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Identity mappings in deep residual networks," in *European Conference on Computer Vision* (Springer), 630–645. doi: 10.1007/978-3-319-46493-0_38

Hesamian, M. H., Jia, W., He, X., and Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *J. Digit. Imaging* 32:582–596. doi: 10.1007/s10278-019-00227-x

Hiary, H., Alomari, R. S., Saadah, M., and Chaudhary, V. (2013). Automated segmentation of stromal tissue in histology images using a voting bayesian model. *Signal, Image Video Proc.* 7, 1229–1237. doi: 10.1007/s11760-012-0393-2

Huang, Z., Wang, X., Wang, J., Liu, W., and Wang, J. (2018). "Weakly-supervised semantic segmentation network with deep seeded region growing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7014–7023. doi: 10.1109/CVPR.2018.00733

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). NNU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18, 203–211. doi: 10.1038/s41592-020-01008-z

Jiang, P.-T., Hou, Q., Cao, Y., Cheng, M.-M., Wei, Y., and Xiong, H.-K. (2019). "Integral object mining via online attention accumulation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2070–2079. doi: 10.1109/ICCV.2019.00216

Jin, G., Liu, C., and Chen, X. (2021). Adversarial network integrating dual attention and sparse representation for semi-supervised semantic segmentation. *Inf. Proc. Manag.* 58:102680. doi: 10.1016/j.ipm.2021.102680

Jin, Q., Cui, H., Sun, C., Zheng, J., Wei, L., Fang, Z., et al. (2022). "Semi-supervised histological image segmentation via hierarchical consistency enforcement," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), 3–13. doi: 10.1007/978-3-031-16434-7_1

Kolesnikov, A., and Lampert, C. H. (2016). "Seed, expand and constrain: three principles for weakly-supervised image segmentation," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14* (Springer), 695–711. doi: 10.1007/978-3-319-46493-0_42

Kumar, N., Gupta, R., and Gupta, S. (2020). Whole slide imaging (WSI) in pathology: current perspectives and future directions. *J. Digit. Imaging* 33, 1034–1040. doi: 10.1007/s10278-020-00351-z

Kweon, H., Yoon, S.-H., Kim, H., Park, D., and Yoon, K.-J. (2021). "Unlocking the potential of ordinary classifier: class-specific adversarial erasing framework for weakly supervised semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6994–7003. doi: 10.1109/ICCV48922.2021.00691

Leube, J., Horn, M., Hartrampf, P. E., Buck, A. K., Lassmann, M., and Tran-Gia, J. (2023). PSMA-PET improves deep learning-based automated CT kidney segmentation. *Zeitschrift Med. Physik.* 34, 231–241. doi: 10.1016/j.zemedi.2023.08.006

Li, J., Yang, S., Huang, X., Da, Q., Yang, X., Hu, Z., et al. (2019). "Signet ring cell detection with a semi-supervised learning framework," in *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2-7, 2019, Proceedings 26* (Springer), 842–854. doi: 10.1007/978-3-030-20351-1_66

Li, K., Wu, Z., Peng, K.-C., Ernst, J., and Fu, Y. (2018). "Tell me where to look: Guided attention inference network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9215–9223. doi: 10.1109/CVPR.2018.00960

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8759–8768. doi: 10.1109/CVPR.2018.00913

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. doi: 10.1109/CVPR.2015.7298965

Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., and Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng* 5, 555–570. doi: 10.1038/s41551-020-00682-w

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)* (IEEE), 565–571. doi: 10.1109/3DV.2016.79

Oskal, K. R., Risdal, M., Janssen, E. A., Undersrud, E. S., and Gulsrud, T. O. (2019). A u-net based approach to epidermal tissue segmentation in whole slide histopathological images. *SN Appl. Sci.* 1, 1–12. doi: 10.1007/s42452-019-0694-y

Pan, W., Yan, J., Chen, H., Yang, J., Xu, Z., Li, X., et al. (2023). "Human-machine interactive tissue prototype learning for label-efficient histopathology image segmentation," in *International Conference on Information Processing in Medical Imaging* (Springer), 679–691. doi: 10.1007/978-3-031-34048-2_52

Qin, J., Wu, J., Xiao, X., Li, L., and Wang, X. (2022). "Activation modulation and recalibration scheme for weakly supervised semantic segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2117–2125. doi: 10.1609/aaai.v36i2.20108

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Verma, R., Kumar, N., Patil, A., Kurian, N. C., Rane, S., Graham, S., et al. (2021). Monusac2020: a multi-organ nuclei segmentation and classification challenge. *IEEE Trans. Med. Imaging* 40, 3413–3423. doi: 10.1109/TMI.2021.3085712

Wang, S., Rong, R., Yang, D. M., Fujimoto, J., Yan, S., Cai, L., et al. (2020). Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer Res.* 80, 2056–2066. doi: 10.1158/0008-5472.CAN-19-1629

Wu, T., Huang, J., Gao, G., Wei, X., Wei, X., Luo, X., et al. (2021). Embedded discriminative attention mechanism for weakly supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16765–16774. doi: 10.1109/CVPR46437.2021.01649

Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., et al. (2022). Generative adversarial networks in medical image segmentation: a review. *Comput. Biol. Med.* 140:105063. doi: 10.1016/j.compbiomed.2021.105063

Yan, J., Chen, H., Li, X., and Yao, J. (2022). Deep contrastive learning based tissue clustering for annotation-free histopathology image analysis. *Computerized Medical Imaging and Graphics* 97:102053. doi: 10.1016/j.compmedimag.2022.102053

Zhang, B., Xiao, J., Wei, Y., Sun, M., and Huang, K. (2020a). "Reliability does matter: an end-to-end weakly supervised semantic segmentation approach," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 12765–12772. doi: 10.1609/aaai.v34i07.6971

Zhang, D., Zhang, H., Tang, J., Hua, X.-S., and Sun, Q. (2020b). "Causal intervention for weakly-supervised semantic segmentation," in *Advances in Neural Information Processing Systems*, 655–666.

Zhang, X., Wei, Y., Feng, J., Yang, Y., and Huang, T. S. (2018). "Adversarial complementary learning for weakly supervised object localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1325–1334. doi: 10.1109/CVPR.2018.00144

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929. doi: 10.1109/CVPR.2016.319

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 08 October 2024
DOI 10.3389/fmicb.2024.1467113

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Zhang Jianglin,
Jinan University, Southern University of Science and Technology, China
Qian Yu,
Tongji University, China
Xia Tianbao,
PLA Strategic Support Force Characteristic Medical Center, China

*CORRESPONDENCE
Xueli Niu
✉ m18842419480@163.com

RECEIVED 19 July 2024
ACCEPTED 27 September 2024
PUBLISHED 08 October 2024

CITATION
Han R, Fan X, Ren S and Niu X (2024) Artificial intelligence in assisting pathogenic microorganism diagnosis and treatment: a review of infectious skin diseases.
Front. Microbiol. 15:1467113.
doi: 10.3389/fmicb.2024.1467113

Artificial intelligence in assisting pathogenic microorganism diagnosis and treatment: a review of infectious skin diseases

Renjie Han[1,2], Xinyun Fan[1,2], Shuyan Ren[1,2] and Xueli Niu[1,2]*

[1]Department of Dermatology, The First Hospital of China Medical University, Shenyang, China, [2]Key Laboratory of Immunodermatology, Ministry of Education and NHC, National Joint Engineering Research Center for Theranostics of Immunological Skin Diseases, Shenyang, China

The skin, the largest organ of the human body, covers the body surface and serves as a crucial barrier for maintaining internal environmental stability. Various microorganisms such as bacteria, fungi, and viruses reside on the skin surface, and densely arranged keratinocytes exhibit inhibitory effects on pathogenic microorganisms. The skin is an essential barrier against pathogenic microbial infections, many of which manifest as skin lesions. Therefore, the rapid diagnosis of related skin lesions is of utmost importance for early treatment and intervention of infectious diseases. With the continuous rapid development of artificial intelligence, significant progress has been made in healthcare, transforming healthcare services, disease diagnosis, and management, including a significant impact in the field of dermatology. In this review, we provide a detailed overview of the application of artificial intelligence in skin and sexually transmitted diseases caused by pathogenic microorganisms, including auxiliary diagnosis, treatment decisions, and analysis and prediction of epidemiological characteristics.

KEYWORDS

artificial intelligence, pathogenic microorganisms, infectious skin diseases, auxiliary diagnosis, treatment decisions

1 Introduction

Infectious skin diseases caused by pathogenic microorganisms are diverse, and many clinical manifestations appear similar. Diagnosis often requires assistance from dermatopathology, and is a complex process that necessitates experienced skin pathology specialists. These intricate procedures make the diagnosis of many skin diseases challenging, particularly infectious skin diseases, where prolonged diagnostic processes can lead to treatment delays. Furthermore, there are a relatively limited number of dermatologists, leading to many diseases being diagnosed and treated by non-specialists, resulting in lower diagnostic accuracy and the likelihood of improper or delayed treatment (Liu et al., 2020). Therefore, there is an urgent need to introduce artificial intelligence algorithms to assist physicians in rapid diagnosis and treatment.

Artificial intelligence (AI) simulates human intelligence using computer systems. This is a new technological science that studies and develops theories, methods, technologies, and application systems to simulate, extend, and expand human intelligence. Machine learning is a subset of AI that enables machines to learn tasks automatically by inferring data patterns. Neural networks are flexible mathematical models that employ various algorithms to identify complex relationships in large databases. Neural networks are currently the most popular machine learning technology, particularly with subtypes such as deep learning and convolutional neural networks (CNNs). We input data in the input layer, process it in a hidden multilayer algorithm, and the processed data is displayed in the output layer. Deep learning can be understood as a computational process with very many hidden layers, rather than a simple neural network with only one or a few layers of nodes between the input and output layers. As the computational power grows, the number of hidden layers can even be stacked indefinitely, resulting in a machine with higher sensitivity and specificity (Jartarkar, 2023). Convolutional Neural Networks are a deep learning model with great success in the computer field, inspired by the biological visual system and designed to mimic the processing of human vision. It uses convolutional operations to capture localized features in an image without being affected by their positions. And unlike traditional feature extraction methods, it does not require manual extraction of features (Li Z. et al., 2022). AI can not only process large amounts of data quickly but also access infinite sources of information, with capabilities for perpetual learning and rapid processing (Lillicrap and Morrissey, 2023). With its powerful functions, AI is widely used in the medical field. AI is often used for medical image recognition and interpretation, such as combining with histopathology to identify specific cells in pathological images, and combining with imaging to identify key features in the images to achieve further assisted diagnosis of various diseases, such as cardiovascular diseases, endocrine diseases, and tumors (Hutchinson et al., 2023; Giorgini et al., 2024; Makimoto and Kohro, 2024). AI is also able to analyze multi-parameter data to develop personalized treatment and care plans that can be referenced, such as combining with radiology to assist in the treatment of diseases such as tumors (Alabi et al., 2024; Bo et al., 2024), and can also be used to improve the efficiency of cardiovascular disease care (Jain et al., 2024). In surgery, AI can automate robotic surgeries, provide computer vision, perform pre-operative risk assessment, and post-operative monitoring (Mirshahvalad et al., 2024). AI is used not only for the classification and recognition of skin diseases but also for epidemiological analysis and predictions, as well as for drug and vaccine development (Russo et al., 2020; Paul et al., 2021; Sung and Hopper, 2023). In dermatology, AI is often used for diagnostic recognition of various skin tumors (Brancaccio et al., 2024; Hartmann et al., 2024), as well as for assisted diagnosis, management, and evaluation of various inflammatory and autoimmune diseases (Doolan and Thomas, 2024; Li Pomi et al., 2024). We carefully searched the literature on AI in the field of infectious dermatology to organize and analyze the applications of AI in this field.

2 AI-assisted diagnosis of skin diseases related to pathogenic microorganisms

The development of AI can assist both professional and nonprofessional individuals in diagnosing and distinguishing diagnoses. By utilizing a large number of clinical case images for training and testing AI algorithms and continuously adjusting and updating them to enhance their sensitivity and specificity, rapid and efficient identification of new case images can be achieved (Figure 1). Currently, there have been some advancements in AI-assisted diagnosis of skin lesions caused by pathogenic microbial infections. We have organized and summarized AI-assisted diagnosis as follows:

Deep CNNs can classify images based on their unique features and are widely used in skin disease classification and diagnostic identification. CNN can be developed and integrated into applications that assist individuals with the diagnosis of monkeypox skin lesions.

Monkeypox, caused by the monkeypox virus (MPXV), is a zoonotic disease (Elsayed et al., 2022). It is characterized by skin lesions that initially present as progressive macules and papules, and later progress to vesicles, pustules, or pseudo-pustules. Thieme et al. (2023) utilized a large dataset of images depicting both monkeypox and non-monkeypox skin lesions to train a CNN algorithm for detecting MPXV skin lesions (MPXV-CNN). The SHapley Additive exPlanations (SHAP) algorithm was employed to identify the regions of high feature importance in the images. In the validation and testing sets, MPXV-CNN exhibited sensitivities of 0.83 and 0.91, and specificities of 0.965 and 0.898, respectively. The effects of factors such as the number of lesions, duration, and site of occurrence on the algorithm have been evaluated (Thieme et al., 2023). An algorithm based on the Al-Biruni Earth radius optimization-based stochastic fractal search was used to fine-tune the CNN, improving its performance from 0.9337 to 0.9883 (Khafaga et al., 2022). Enhanced residual CNNs based on λ function and context transformer (LaCTResNet) (Chen and Han, 2023) and Chaos game optimization algorithm-based fusion of deep neural networks (CGO-ensemble) (Asif et al., 2024) were also employed for monkeypox image recognition, enhancing the efficiency of monkeypox identification. McNeil et al. (2023) developed an AI algorithm based on a ubiquitous U-Net deep learning architecture to calculate the number of monkeypox lesions in patient photographs, which employs a segmentation method that categorizes each pixel in each photograph as either belonging to a monkeypox lesion or not belonging to a monkeypox lesion, aiding in monitoring the staging and severity of monkeypox.

Deep CNNs can also be used for the auxiliary diagnosis of skin fungal infections, especially onychomycosis (Han et al., 2018; Kim et al., 2020; Lim et al., 2021; Gupta and Hall, 2022; Zhu et al., 2022; Fang et al., 2023).

Fungal nail disease is caused by fungal infections and leads to discoloration, thickening, and separation of the nail bed (Westerberg and Voyack, 2013). Microscopy and fungal culture are the gold standard techniques for diagnosing onychomycosis; however, they have a relatively high false negative rate (Gupta et al., 2020). The combination of the YOLO v4 deep convolutional network with microscopy enables automation of fungal identification and detection (Koo et al., 2021). The developers trained the target detection convolutional neural network YOLO v4 on microscope images with magnifications of 100×, 40×, and (100 + 40)×. Its sensitivity and specificity were, respectively, 0.952 and 1.0 in the 100× data model, and 0.99 and 0.866 in the 40× data model; the sensitivity and specificity in the combined (100 + 40) × data model were 0.932 and 0.89, respectively, indicating that mycelium was detected with reliable accuracy. Additionally, the integration of the VGG16 and InceptionV3 models with deep CNNs (Yilmaz et al., 2022), as well as image

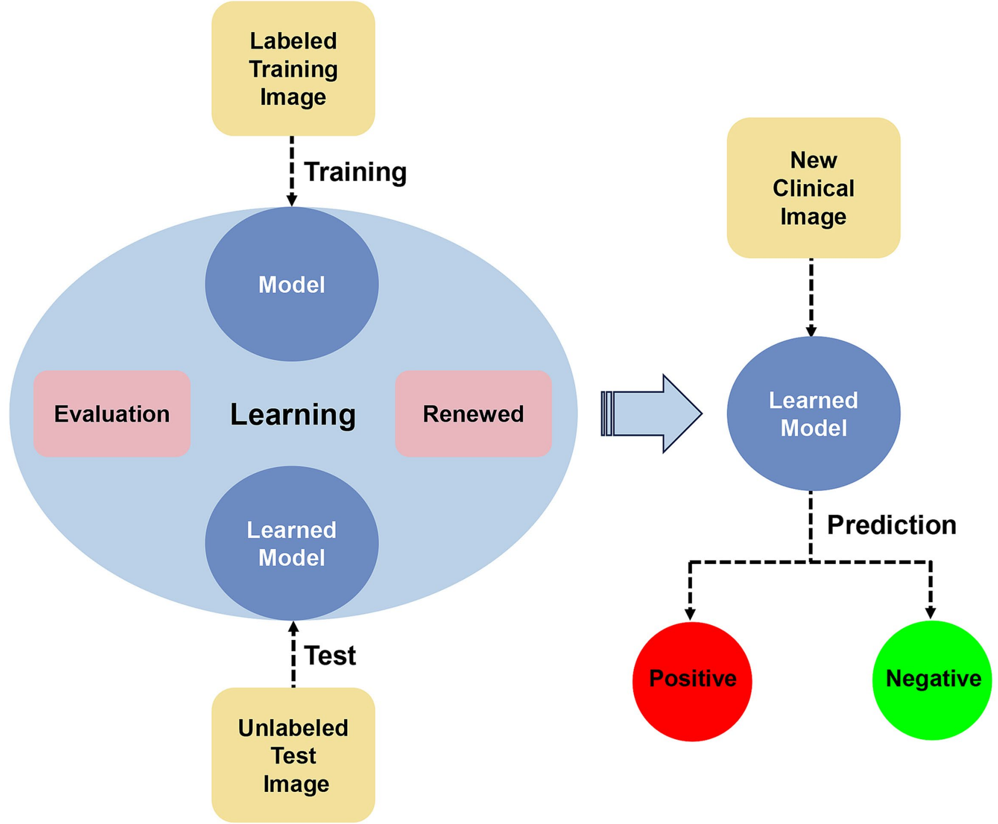


FIGURE 1
Artificial intelligence-assisted diagnostic model diagram.

processing models based on residual neural networks (ResNet) (Gao et al., 2021), allows for automatic detection of skin fungi using microscopy. Furthermore, employing AI to assist single-cell Raman spectroscopy technology not only distinguishes between bacterial and fungal skin infections but also identifies fungal species, with researchers reporting a 100% accuracy rate at the strain level (Xu et al., 2023). AI deep learning combined with histopathology can serve as a screening tool to highlight suspicious mycelial areas for rapid confirmation by dermatopathologists (Decroos et al., 2021). Moreover, AI deep learning has been applied in the diagnosis of cryptococcosis (Wei et al., 2023).

For bacterial skin diseases, AI has been widely used in the auxiliary diagnosis of acne (Yang et al., 2021) and leprosy (Barbieri et al., 2022; Fernandes et al., 2023), achieving good results in monitoring, preventing, and guiding patient medication. We made a table to make it easier to see (Table 1). The Inception-v3 network, a deep learning-based classification model, was trained by the researchers using common clinical photographs of acne of varying severity to model the assessment of acne severity and classify the type of lesion based on the image; Inception-v4 and ResNet-50 were also used to train the assessment of leprosy images.

In conclusion, intelligent AI diagnosis can be applied to viral, bacterial, and fungal skin diseases, especially for the rapid identification of monkeypox skin lesions and the differential diagnosis of skin fungi, demonstrating high sensitivity and specificity, and can serve as an important screening tool.

3 The application of AI in predicting and monitoring infectious skin diseases and sexually transmitted diseases

Infectious diseases are associated with the presence of pathogenic microorganisms. Understanding their epidemiological characteristics, early prediction, and monitoring of at-risk populations can help in disease prevention and control.

The EPIWATCH system can analyze the epidemiological characteristics of febrile exanthematous diseases such as monkeypox. It uses automated technology to scan large amounts of open-source data from media reports, news releases, official reports, and social media for early warning of emerging infectious diseases (Hutchinson et al., 2023). Additionally, researchers have compared the accuracy of nine AI models in predicting monkeypox outbreaks and provided the details of each model (Chadaga et al., 2023).

Early detection of *Mycobacterium leprae* and its infections is a key factor in breaking the leprosy transmission chain. An AI molecular and serological comprehensive analysis method based on the random forest algorithm can be used to better diagnose and predict new leprosy cases among contacts. Its sensitivity in the diagnosis of polymicrobial leprosy was 90.5%, better than traditional anti-LID-1 (0.632), anti-ND-O-LID (0.579), and especially in oligomicrobial leprosy (70.6%), which also showed a significant increase in sensitivity, with a total specificity of 92.5% (Gama et al., 2019).

TABLE 1 Application of AI algorithm model in infectious skin diseases.

Pathogen	Disease	AI Algorithm model and features		Application
Virus infection	Monkeypox	CNN + SHAP (Thieme et al., 2023) CNN + BERSFS (Khafaga et al., 2022) CGO-ensemble (Asif et al., 2024) ResNet (Chen and Han, 2023)	Identify the image features; better than clinical CNN	Auxiliary diagnosis
		EPIWATCH system (Hutchinson et al., 2023)	Analyze the epidemiological characteristics	Risk prediction
	Warts	IAPSO for AIRS (Abdar et al., 2019) Fuzzy rule-based system (Khozeimeh et al., 2017)	Predict and evaluate treatment response	Clinical Decision
	AIDS/HIV Infection	GBM, RF, DL, XGBoost (Bao et al., 2021)	Can be used for high-risk populations and individuals	Risk prediction
		ChatGPT 3.5 (Koh et al., 2024)	Provide treatment advice for patients	Clinical Decision
		Logistic regression, RF, AdaBoost (Maskew et al., 2022)	Guide the adjustment of interventions	
		RF, SVM, MLP (Li B. et al., 2022)	Predicted changes in immune function after 9.9 years	Prognosis prediction
		Bayesian Additive Regression Trees (Elder et al., 2021)	Recurrence risk prediction	
		Graph neural network (GNN) (Wang et al., 2023) Simplified Molecular Input Line System (Chavez-Hernandez et al., 2021) Decision trees, Logistic regression, Artificial neural networks (Singh and Su, 2016)	For HIV-1 protease inhibitors	Treatment target prediction
		Artificial neural networks (Conti and Karplus, 2019)	Estimate the breadth of antibodies	Vaccine development
		IDEPI (Hepler et al., 2014)	Predicted antibody epitope	
Fungal infection	Onychomycosis	DCNN+YoLov4 (Koo et al., 2021) DCNN+VGG16 + InceptionV3 (Yilmaz et al., 2022) ResNet (Gao et al., 2021)	Combined with a microscope for automatic detection	Auxiliary diagnosis
	Cryptococcosis	VGG19, MobileNet, InceptionV3, Incept ResNetV2, DenseNet201 (Wei et al., 2023)	–	Auxiliary diagnosis
Bacterial infection	Acne	Inception-v3 (Yang et al., 2021)	–	Auxiliary diagnosis
	Lepra	Inception-v4, ResNet-50 (Barbieri et al., 2022)	–	Auxiliary diagnosis
		RF (Gama et al., 2019)	New case prediction	Risk prediction
		Bayesian networks (de Andrade Rodrigues et al., 2023)	Predict LR probability	Clinical Decision
	Pyemia	Bayesian networks (Komorowski et al., 2018)	–	Clinical Decision

AI can also be employed to predict HIV infection risk in high-risk populations, showing significant improvements over traditional prediction methods (Bao et al., 2021). Pre-exposure prophylaxis (PrEP) involves the use of specific antiviral drugs by individuals not infected with HIV before engaging in HIV-susceptible behaviors. Machine learning algorithms have substantial potential to optimize PrEP by enhancing the identification of high-risk HIV populations (Marcus et al., 2020). Risk scores generated by the AI-based risk assessment tool MySTIRisk, in conjunction with the Jördan index, exhibited 86% sensitivity and 65.6% specificity for identifying populations at high risk for HIV/sexually transmitted diseases (Latt et al., 2024). Researchers have utilized the MySTIRisk tool to develop an online self-assessment questionnaire for predicting individual HIV and sexually transmitted infection risks (Xu et al., 2022). Using machine learning, researchers designed an algorithm model based on electronic health records to swiftly identify individuals at higher risk of HIV infection (Burns et al., 2023), ultimately contributing to increased PrEP utilization. AI plays a crucial role in revolutionizing healthcare, demonstrating significant potential for HIV prevention and intervention strategies (Khatami and Gopalappa, 2021; Xiang et al., 2022).

Many studies have utilized new AI algorithms to predict the occurrence of syphilis in high-risk populations, which could potentially serve as tools for controlling and monitoring its spread (Albuquerque et al., 2023).

The application of AI algorithms aids in the early detection of infectious diseases, such as smallpox and leprosy, helping to break the chain of transmission. Predicting high-risk populations for sexually transmitted diseases such as HIV and syphilis is beneficial for the prevention and control of related diseases and for guiding the rational distribution of health resources.

4 AI aids in developing better treatment plans for infectious skin diseases and sexually transmitted diseases

In addition to assisting in the diagnosis and monitoring of infectious skin diseases, AI also aids in the development of optimal treatment plans.

These include classifying leprosy cases, ensuring patient compliance with drug therapy, monitoring geographical treatment coverage, and facilitating the early detection of adverse drug reactions and antimicrobial resistance. AI can also help in the early detection of nerve damage in patients with leprosy, thereby aiding disability prevention and rehabilitation planning (Deps et al., 2024). An AI-based leprosy screening cross-platform application can classify cases as paucibacillary leprosy or multibacillary leprosy, assisting professionals in accurate disease classification and determining appropriate treatment methods (De Souza et al., 2021). The leprosy reaction (LR) is a severe inflammatory response in patients with leprosy and is a major cause of permanent nerve damage. Assessing the risk factors for LR in patients is challenging, but AI can be used to predict LR. An AI system developed based on Bayesian networks and utilizing the NETICA software can assess LR risk based on clinical, demographic, and genetic data, thereby effectively guiding clinical decisions. It has an accuracy of 0.827, a sensitivity of 0.793 and a specificity of 0.862 (de Andrade Rodrigues et al., 2023). AI models can also provide personalized and clinically interpretable treatment decisions for sepsis, thereby improving patient outcomes (Komorowski et al., 2018).

Abdar et al. (2019) proposed a novel evolutionary computer-aided diagnosis (CAD) system, whose main architecture is a combination of improved adaptive particle swarm optimization (IAPSO) and an Artificial Immune Recognition System (AIRS). The CAD system can be used to evaluate the response of warts to immunotherapy and cryotherapy. AIRS is a classification algorithm modeled after the human immune system, and IAPSO has improved the treatment response performance of AIRS by improving the algorithm. Other scholars have utilized Fuzzy rule-based system to predict and assess treatment responses to these two therapies for warts by using information gain to identify the factors that characterize the effective treatment, and then the Fuzzy rule-based system to predict the treatment effect, aiding physicians in selecting the optimal treatment method (Khozeimeh et al., 2017; Singh, 2021).

Trained ChatGPT can provide professional and scientific answers to common treatment queries from HIV-infected individuals, offering consultations and advice on antiretroviral therapy to guide patients through the treatment process (Koh et al., 2024). Machine learning algorithms can also predict and identify HIV patients at risk of treatment interruption and unsuppressed viral load, allowing targeted interventions through differentiated care models to improve cost-effectiveness and prognosis before patients deviate from treatment plans (Maskew et al., 2022). Domínguez-Rodríguez et al. (2022) compared seven machine learning algorithms and found accurate predictions of the prognosis of children with perinatally acquired HIV infection (Domínguez-Rodríguez et al., 2022). AI machine learning models can utilize clinical monitoring indicators to predict changes in the immune function of AIDS patients after 9.9 years of antiretroviral therapy, aiding in patient prognosis assessment (Li B. et al., 2022). Machine learning linked to electronic medical records can be used to predict the risk of recurrent infectious diseases and provide valuable insights (Elder et al., 2021).

In summary, AI can integrate and analyze large amounts clinical, demographic, genetic, and epidemiological data to provide personalized clinical diagnosis and treatment decisions for patients with high-risk infectious diseases such as leprosy and AIDS. It has achieved favorable results in clinical indicator monitoring, disease progression prediction, and cost-effectiveness improvement, thereby providing a more comprehensive perspective on the diagnosis and treatment of infectious skin diseases.

5 AI assists in drug development and vaccine research

With the development of computer-aided drug design technology, AI has been successfully utilized for rapid innovation in the virtual screening of candidate drugs (Wang et al., 2023). The application of graph neural networks to predict the antibiotic activity and cytotoxicity of 12,076,365 compounds aids in the selection of molecules with antibiotic activity (Wong et al., 2024). Graph neural networks have also been employed to accurately predict potential therapeutic drugs for HIV-1/HBV coinfection, showing potential applications in multi-target drug virtual screening (Wang et al., 2023).

The development of a virtual HIV-1 protease inhibitor library based on natural compound fragments using AI can facilitate the discovery of effective HIV-1 protease inhibitors (Chavez-Hernandez et al., 2021). AI algorithms can also be utilized to predict HIV-1 protease cleavage sites, contributing to the development of HIV-1 protease inhibitors (Singh and Su, 2016; Hu et al., 2022).

Artificial neural networks (ANNs) are a powerful tool that can be used to predict multiple resistances to HIV-1 protease and reverse transcriptase inhibitors (Sheik Amamuddy et al., 2017; Tunc et al., 2023). The EuResist engine was used to forecast responses to anti-HIV treatments, effectively assisting virology experts in selecting effective target drugs for patients carrying drug-resistant HIV strains (Zazzi et al., 2012). A combination of chemoinformatics and artificial neural network methods can be employed to predict and score the activity of ligands that bind to the catalytic core domain of the HIV-1 integrase enzyme (Thangsunan et al., 2016). Regularized ANNs have also been employed to simulate the activity of cyclic urea (a type of HIV-1 protease inhibitor) (Fernandez and Caballero, 2006).

The traditional production of vaccines requires several years and involves high costs. By utilizing AI to assist vaccine development and design, significant time and economic costs can be reduced. Machine learning also plays a role in HIV and antibody research, with AI computational methods used to predict applications in antibodies, neutralizing breadth against multiple viral strains, detecting antibody-virus binding sites, enhancing antibody design, and studying antibody-induced immune responses (Danaila et al., 2022). Machine learning and molecular modeling can also estimate the breadth of CD4bs-targeting HIV antibodies, a method that holds promise for use in the design of HIV antibodies (Conti and Karplus, 2019). By leveraging open-source general machine learning algorithms and libraries, Hepler et al. (2014) developed a software package called IDEPI (IDentify EPItopes) for learning genotype-to-phenotype prediction models from sequences with known phenotypes with the aim of rapidly predicting HIV-1 antibody epitopes and other phenotypic characteristics (Hepler et al., 2014).

With its powerful learning ability, AI can process large amounts of data in a short period, greatly enhancing the efficiency of screening and predicting molecules with pharmaceutical activity and drug therapeutic targets. They can also help to predict the binding sites of viruses and antibodies, evaluate the neutralization potency of antibodies, and play important roles in drug development and vaccine design.

6 Discussion

AI, with its powerful computing and learning capabilities, has become a technological direction with huge potential that profoundly impacts social development and human civilization. AI is widely applied in the fields of infectious skin diseases and sexually transmitted diseases, not only for assisting in diagnosis but also for helping in disease treatment, epidemic prevention and control decision-making, prediction of drug treatment targets, and vaccine development (Figure 2). There are more and more researches on the combination of AI with dermoscopy and dermatopathology in the recognition and diagnosis of infectious skin diseases, however, there are many kinds of skin diseases, and many diseases have similar skin lesion manifestations, which makes it easy for even experienced dermatologists to make mistakes, therefore, AI still needs to be improved in terms of accuracy in the assisted diagnosis of dermatological diseases, and perhaps more dimensional parameters other than pictures can be added to improve the diagnostic efficiency and lead toward more precise diagnosis and treatment. Additionally, the application of AI is mostly in the research stage at present, and the types of skin diseases involved are not comprehensive; the algorithms of AI rely on the selection of reasonable parameters, and their learning mode is limited by the quality of the received information, with various factors affecting whether the

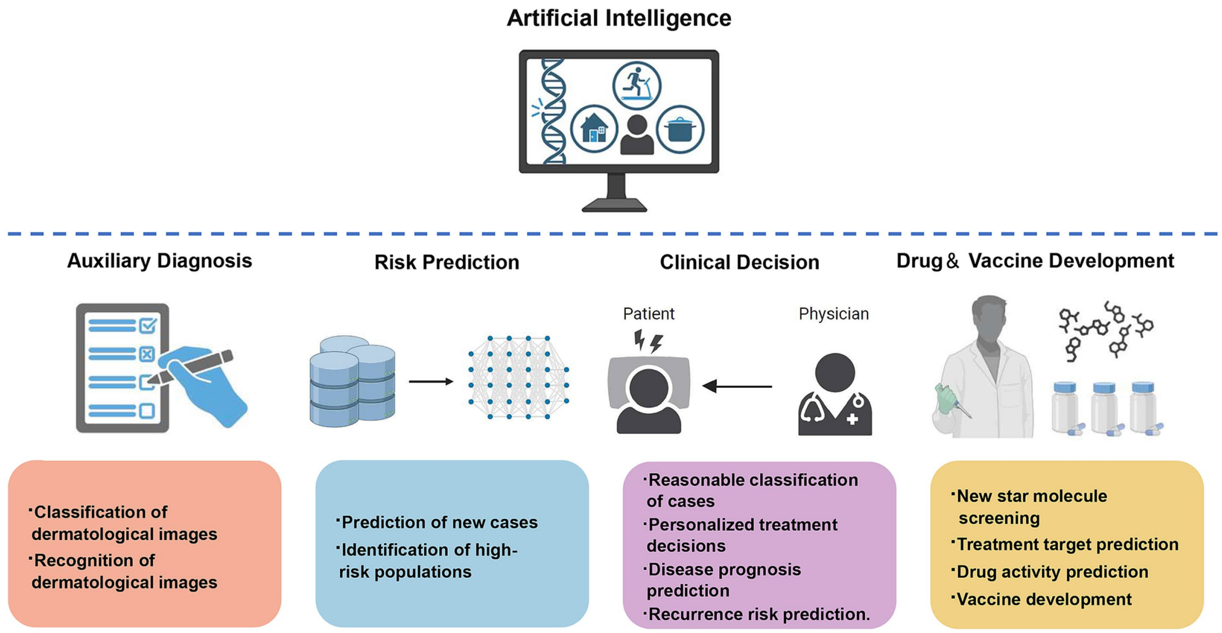


FIGURE 2
The application of artificial intelligence in infectious skin diseases and sexually transmitted diseases in dermatology.

algorithm produces an impact, all of which need to be carefully and rigorously tested. In addition, ethical issues in the application of AI need to be pondered, which may involve issues such as data security, privacy invasion, and the lack of standardized regulations (Goldust and Grant-Kels, 2024; Gordon et al., 2024), which may require safeguards to ensure the sound application of AI. Overall, AI is constantly progressing, and these limitations will receive more attention and discussion in the future. It is expected that in the future, AI algorithms and computing power will continue to improve; be applied to more skin diseases; computer science, biology, and medicine more cross-field cooperation, and joint participation in the research and application of AI in dermatology and venereology, improve the effectiveness of disease diagnosis and treatment; effectively reduce the health, psychological, and economic burden on patients; and make greater contributions to human health.

Author contributions

RH: Writing – original draft. XF: Writing – review & editing, Supervision, Methodology. SR: Writing – review & editing, Supervision, Methodology. XN: Supervision, 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 funded and supported by the Liaoning Provincial Natural Science Foundation—Joint Fund for Intersection of Medicine and Engineering (2021-YGJC-11).

Acknowledgments

We sincerely appreciate the support and funding provided by the Liaoning Provincial Natural Science Foundation - Joint Fund for Intersection of Medicine and Engineering, which has enabled the smooth progress of our research.

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

Abdar, M., Wijayaningrum, V. N., Hussain, S., Alizadehsani, R., Plawiak, P., Acharya, U. R., et al. (2019). IAPSO-AIRS: a novel improved machine learning-based system for wart disease treatment. *J. Med. Syst.* 43:220. doi: 10.1007/s10916-019-1343-0

Alabi, R. O., Elmusrati, M., Leivo, I., Almangush, A., and Makitie, A. A. (2024). Artificial intelligence-driven Radiomics in head and neck Cancer: current status and future prospects. *Int. J. Med. Inform.* 188:105464. doi: 10.1016/j.ijmedinf.2024.105464

Albuquerque, G., Fernandes, F., Barbalho, I. M. P., Barros, D. M. S., Morais, P. S. G., Morais, A. H. F., et al. (2023). Computational methods applied to syphilis: where are we, and where are we going? *Front. Public Health* 11:1201725. doi: 10.3389/fpubh.2023.1201725

Asif, S., Zhao, M., Li, Y., Tang, F., and Zhu, Y. (2024). CGO-ensemble: Chaos game optimization algorithm-based fusion of deep neural networks for accurate Mpox detection. *Neural Netw.* 173:106183. doi: 10.1016/j.neunet.2024.106183

Bao, Y., Medland, N. A., Fairley, C. K., Wu, J., Shang, X., Chow, E. P. F., et al. (2021). Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. *J. Infect.* 82, 48–59. doi: 10.1016/j.jinf.2020.11.007

Barbieri, R. R., Xu, Y., Setian, L., Souza-Santos, P. T., Trivedi, A., Cristofono, J., et al. (2022). Reimagining leprosy elimination with AI analysis of a combination of skin lesion images with demographic and clinical data. *Lancet Reg Health Am* 9:100192. doi: 10.1016/j.lana.2022.100192

Bo, Z., Song, J., He, Q., Chen, B., Chen, Z., Xie, X., et al. (2024). Application of artificial intelligence radiomics in the diagnosis, treatment, and prognosis of hepatocellular carcinoma. *Comput. Biol. Med.* 173:108337. doi: 10.1016/j.compbiomed.2024.108337

Brancaccio, G., Balato, A., Malvehy, J., Puig, S., Argenziano, G., and Kittler, H. (2024). Artificial intelligence in skin Cancer diagnosis: a reality check. *J. Invest. Dermatol.* 144, 492–499. doi: 10.1016/j.jid.2023.10.004

Burns, C. M., Pung, L., Witt, D., Gao, M., Sendak, M., Balu, S., et al. (2023). Development of a human immunodeficiency virus risk prediction model using electronic health record data from an academic health system in the southern United States. *Clin. Infect. Dis.* 76, 299–306. doi: 10.1093/cid/ciac775

Chadaga, K., Prabhu, S., Sampathila, N., Nireshwalya, S., Katta, S. S., Tan, R. S., et al. (2023). Application of artificial intelligence techniques for Monkeypox: a systematic review. *Diagnostics* 13:824. doi: 10.3390/diagnostics13050824

Chavez-Hernandez, A. L., Juarez-Mercado, K. E., Saldivar-Gonzalez, F. I., and Medina-Franco, J. L. (2021). Towards the De novo design of HIV-1 protease inhibitors based on natural products. *Biomol. Ther.* 11:1805. doi: 10.3390/biom11121805

Chen, J., and Han, J. (2023). A study on the recognition of Monkeypox infection based on deep convolutional neural networks. *Front. Immunol.* 14:1225557. doi: 10.3389/fimmu.2023.1225557

Conti, S., and Karplus, M. (2019). Estimation of the breadth of CD4bs targeting HIV antibodies by molecular modeling and machine learning. *PLoS Comput. Biol.* 15:e1006954. doi: 10.1371/journal.pcbi.1006954

Danaila, V. R., Avram, S., and Buiu, C. (2022). The applications of machine learning in HIV neutralizing antibodies research-a systematic review. *Artif. Intell. Med.* 134:102429. doi: 10.1016/j.artmed.2022.102429

de Andrade Rodrigues, R. S., Heise, E. F. J., Hartmann, L. F., Rocha, G. E., Olandoski, M., de Araujo Stefani, M. M., et al. (2023). Prediction of the occurrence of leprosy reactions based on Bayesian networks. *Front. Med.* 10:1233220. doi: 10.3389/fmed.2023.1233220

De Souza, M. L. M., Lopes, G. A., Branco, A. C., Fairley, J. K., and Fraga, L. A. O. (2021). Leprosy screening based on artificial intelligence: development of a cross-platform app. *JMIR Mhealth Uhealth* 9:e23718. doi: 10.2196/23718

Decroos, F., Springenberg, S., Lang, T., Papper, M., Zapf, A., Metze, D., et al. (2021). A deep learning approach for histopathological diagnosis of onychomycosis: not inferior to analogue diagnosis by Histopathologists. *Acta Derm. Venereol.* 101:adv00532. doi: 10.2340/00015555-3893

Deps, P. D., Yotsu, R., Furriel, B., de Oliveira, B. D., de Lima, S. L., and Loureiro, R. M. (2024). The potential role of artificial intelligence in the clinical management of Hansen's disease (leprosy). *Front. Med.* 11:1338598. doi: 10.3389/fmed.2024.1338598

Domínguez-Rodríguez, S., Serna-Pascual, M., Oletto, A., Barnabas, S., Zuidewind, P., Dobbels, E., et al. (2022). Machine learning outperformed logistic regression classification even with limit sample size: a model to predict pediatric HIV mortality and clinical progression to AIDS. *PLoS One* 17:e0276116. doi: 10.1371/journal.pone.0276116

Doolan, B. J., and Thomas, B. R. (2024). Bursting the bubble on diagnostics: artificial intelligence in autoimmune bullous disease. *Br. J. Dermatol.* 191, 160–161. doi: 10.1093/bjd/ljae197

Elder, H. R., Gruber, S., Willis, S. J., Cocoros, N., Callahan, M., Flagg, E. W., et al. (2021). Can machine learning help identify patients at risk for recurrent sexually transmitted infections? *Sex. Transm. Dis.* 48, 56–62. doi: 10.1097/OLQ.0000000000001264

Elsayed, S., Bondy, L., and Hanage, W. P. (2022). Monkeypox virus infections in humans. *Clin. Microbiol. Rev.* 35:e0009222. doi: 10.1128/cmr.00092-22

Fang, W., Wu, J., Cheng, M., Zhu, X., Du, M., Chen, C., et al. (2023). Diagnosis of invasive fungal infections: challenges and recent developments. *J. Biomed. Sci.* 30:42. doi: 10.1186/s12929-023-00926-2

Fernandes, J. R. N., Teles, A. S., Fernandes, T. R. S., Lima, L. D. B., Balhara, S., Gupta, N., et al. (2023). Artificial intelligence on diagnostic aid of leprosy: a systematic literature review. *J. Clin. Med.* 13:180. doi: 10.3390/jcm13010180

Fernandez, M., and Caballero, J. (2006). Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks. *Bioorg. Med. Chem.* 14, 280–294. doi: 10.1016/j.bmc.2005.08.022

Gama, R. S., Souza, M. L. M., Sarno, E. N., Moraes, M. O., Goncalves, A., Stefani, M. M. A., et al. (2019). A novel integrated molecular and serological analysis method to predict new cases of leprosy amongst household contacts. *PLoS Negl. Trop. Dis.* 13:e0007400. doi: 10.1371/journal.pntd.0007400

Gao, W., Li, M., Wu, R., Du, W., Zhang, S., Yin, S., et al. (2021). The design and application of an automated microscope developed based on deep learning for fungal detection in dermatology. *Mycoses* 64, 245–251. doi: 10.1111/myc.13209

Giorgini, F., Di Dalmazi, G., and Diciotti, S. (2024). Artificial intelligence in endocrinology: a comprehensive review. *J. Endocrinol. Investig.* 47, 1067–1082. doi: 10.1007/s40618-023-02235-9

Goldust, M., and Grant-Kels, J. M. (2024). Professional standards and regulations for the use of artificial intelligence in dermatology. *Int. J. Dermatol.* doi: 10.1111/ijd.17235

Gordon, E. R., Trager, M. H., Kontos, D., Weng, C., Geskin, L. J., Dugdale, L. S., et al. (2024). Ethical considerations for artificial intelligence in dermatology: a scoping review. *Br. J. Dermatol.* 190, 789–797. doi: 10.1093/bjd/ljae040

Gupta, A. K., and Hall, D. C. (2022). Diagnosing onychomycosis: a step forward? *J. Cosmet. Dermatol.* 21, 530–535. doi: 10.1111/jocd.14681

Gupta, A. K., Stec, N., Summerbell, R. C., Shear, N. H., Piguet, V., Tosti, A., et al. (2020). Onychomycosis: a review. *J. Eur. Acad. Dermatol. Venereol.* 34, 1972–1990. doi: 10.1111/jdv.16394

Han, S. S., Park, G. H., Lim, W., Kim, M. S., Na, J. I., Park, I., et al. (2018). Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS One* 13:e0191493. doi: 10.1371/journal.pone.0191493

Hartmann, T., Passauer, J., Hartmann, J., Schmidberger, L., Kneilling, M., and Volc, S. (2024). Basic principles of artificial intelligence in dermatology explained using melanoma. *J. Dtsch. Dermatol. Ges.* 22, 339–347. doi: 10.1111/ddg.15322

Hepler, N. L., Scheffler, K., Weaver, S., Murrell, B., Richman, D. D., Burton, D. R., et al. (2014). IDEPI: rapid prediction of HIV-1 antibody epitopes and other phenotypic features from sequence data using a flexible machine learning platform. *PLoS Comput. Biol.* 10:e1003842. doi: 10.1371/journal.pcbi.1003842

Hu, L., Li, Z., Tang, Z., Zhao, C., Zhou, X., and Hu, P. (2022). Effectively predicting HIV-1 protease cleavage sites by using an ensemble learning approach. *BMC Bioinf.* 23:447. doi: 10.1186/s12859-022-04999-y

Hutchinson, D., Kunasekaran, M., Quigley, A., Moa, A., and MacIntyre, C. R. (2023). Could it be Monkeypox? Use of an AI-based epidemic early warning system to monitor rash and fever illness. *Public Health* 220, 142–147. doi: 10.1016/j.puhe.2023.05.010

Jain, S. S., Elias, P., Poterucha, T., Randazzo, M., Lopez Jimenez, F., Khera, R., et al. (2024). Artificial intelligence in cardiovascular care-part 2: applications: JACC review topic of the week. *J. Am. Coll. Cardiol.* 83, 2487–2496. doi: 10.1016/j.jacc.2024.03.401

Jartarkar, S. R. (2023). Artificial intelligence: its role in dermatopathology. *Indian J. Dermatol. Venereol. Leprol.* 89, 549–552. doi: 10.25259/IJDVL_725_2021

Khafaga, D. S., Ibrahim, A., El-Kenawy, E. M., Abdelhamid, A. A., Karim, F. K., Mirjalili, S., et al. (2022). An Al-Biruni earth radius optimization-based deep convolutional neural network for classifying Monkeypox disease. *Diagnostics* 12:2892. doi: 10.3390/diagnostics12112892

Khatami, S. N., and Gopalappa, C. (2021). A reinforcement learning model to inform optimal decision paths for HIV elimination. *Math. Biosci. Eng.* 18, 7666–7684. doi: 10.3934/mbe.2021380

Khozeimeh, F., Alizadehsani, R., Roshanzamir, M., Khosravi, A., Layegh, P., and Nahavandi, S. (2017). An expert system for selecting wart treatment method. *Comput. Biol. Med.* 81, 167–175. doi: 10.1016/j.compbiomed.2017.01.001

Kim, Y. J., Han, S. S., Yang, H. J., and Chang, S. E. (2020). Prospective, comparative evaluation of a deep neural network and dermoscopy in the diagnosis of onychomycosis. *PLoS One* 15:e0234334. doi: 10.1371/journal.pone.0234334

Koh, M. C. Y., Ngiam, J. N., Yong, J., Tambyah, P. A., and Archuleta, S. (2024). The role of an artificial intelligence model in antiretroviral therapy counselling and advice for people living with HIV. *HIV Med.* 25, 504–508. doi: 10.1111/hiv.13604

Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., and Faisal, A. A. (2018). The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* 24, 1716–1720. doi: 10.1038/s41591-018-0213-5

Koo, T., Kim, M. H., and Jue, M. S. (2021). Automated detection of superficial fungal infections from microscopic images through a regional convolutional neural network. *PLoS One* 16:e0256290. doi: 10.1371/journal.pone.0256290

Latt, P. M., Soe, N. N., Xu, X., Ong, J. J., Chow, E. P. F., Fairley, C. K., et al. (2024). Identifying individuals at high risk for HIV and sexually transmitted infections with an artificial intelligence-based risk assessment tool. *Open Forum Infect. Dis.* 11:ofae011. doi: 10.1093/ofid/ofae011

Li, B., Li, M., Song, Y., Lu, X., Liu, D., He, C., et al. (2022). Construction of machine learning models to predict changes in immune function using clinical monitoring indices in HIV/AIDS patients after 9.9-years of antiretroviral therapy in Yunnan, China. *Front. Cell. Infect. Microbiol.* 12:867737. doi: 10.3389/fcimb.2022.867737

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. Neural. Netw. Learn. Syst.* 33, 6999–7019. doi: 10.1109/TNNLS.2021.3084827

Li Pomi, F., Papa, V., Borgia, F., Vaccaro, M., Pioggia, G., and Gangemi, S. (2024). Artificial intelligence: a snapshot of its application in chronic inflammatory and autoimmune skin diseases. *Life* 14:40516. doi: 10.3390/life14040516

Lillicrap, D., and Morrissey, J. H. (2023). Artificial intelligence, science, and learning. *J. Thromb. Haemost.* 21:709. doi: 10.1016/j.jtha.2023.01.026

Lim, S. S., Ohn, J., and Mun, J. H. (2021). Diagnosis of onychomycosis: from conventional techniques and dermoscopy to artificial intelligence. *Front. Med.* 8:637216. doi: 10.3389/fmed.2021.637216

Liu, Y., Jain, A., Eng, C., Way, D. H., Lee, K., Bui, P., et al. (2020). A deep learning system for differential diagnosis of skin diseases. *Nat. Med.* 26, 900–908. doi: 10.1038/s41591-020-0842-3

Makimoto, H., and Kohro, T. (2024). Adopting artificial intelligence in cardiovascular medicine: a scoping review. *Hypertens. Res.* 47, 685–699. doi: 10.1038/s41440-023-01469-7

Marcus, J. L., Sewell, W. C., Balzer, L. B., and Krakower, D. S. (2020). Artificial intelligence and machine learning for HIV prevention: emerging approaches to ending the epidemic. *Curr. HIV/AIDS Rep.* 17, 171–179. doi: 10.1007/s11904-020-00490-6

Maskew, M., Sharpey-Schafer, K., De Voux, L., Crompton, T., Bor, J., Rennick, M., et al. (2022). Applying machine learning and predictive modeling to retention and viral suppression in south African HIV treatment cohorts. *Sci. Rep.* 12:12715. doi: 10.1038/s41598-022-16062-0

McNeil, A. J., House, D. W., Mbala-Kingebeni, P., Mbaya, O. T., Dodd, L. E., Cowen, E. W., et al. (2023). Counting Monkeypox lesions in patient photographs: limits of agreement of manual counts and artificial intelligence. *J. Invest. Dermatol.* 143, 347–350. doi: 10.1016/j.jid.2022.08.044

Mirshahvalad, S. A., Eisazadeh, R., Shahbazi-Akbari, M., Pirich, C., and Beheshti, M. (2024). Application of artificial intelligence in oncologic molecular PET-imaging: a narrative review on beyond [(18)F]F-FDG tracers - part I. PSMA, choline, and DOTA radiotracers. *Semin. Nucl. Med.* 54, 171–180. doi: 10.1053/j.semnuclmed.2023.08.004

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., and Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discov. Today* 26, 80–93. doi: 10.1016/j.drudis.2020.10.010

Russo, G., Reche, P., Pennisi, M., and Pappalardo, F. (2020). The combination of artificial intelligence and systems biology for intelligent vaccine design. *Expert Opin. Drug Discov.* 15, 1267–1281. doi: 10.1080/17460441.2020.1791076

Sheik Amamuddy, O., Bishop, N. T., and Tastan Bishop, O. (2017). Improving fold resistance prediction of HIV-1 against protease and reverse transcriptase inhibitors using artificial neural networks. *BMC Bioinf.* 18:369. doi: 10.1186/s12859-017-1782-x

Singh, Y. (2021). The application of machine learning in predicting outcome of cryotherapy and immunotherapy for wart removal. *Ann. Dermatol.* 33, 345–350. doi: 10.5021/ad.2021.33.4.345

Singh, O., and Su, E. C. (2016). Prediction of HIV-1 protease cleavage site using a combination of sequence, structural, and physicochemical features. *BMC Bioinf.* 17:478. doi: 10.1186/s12859-016-1337-6

Sung, J., and Hopper, J. L. (2023). Co-evolution of epidemiology and artificial intelligence: challenges and opportunities. *Int. J. Epidemiol.* 52, 969–973. doi: 10.1093/ije/dyad089

Thangsunan, P., Kittiwachana, S., Meepowpan, P., Kungwan, N., Prangkio, P., Hannongbua, S., et al. (2016). Rapid activity prediction of HIV-1 integrase inhibitors: harnessing docking energetic components for empirical scoring by chemometric and artificial neural network approaches. *J. Comput. Aided Mol. Des.* 30, 471–488. doi: 10.1007/s10822-016-9917-0

Thieme, A. H., Zheng, Y., Machiraju, G., Sadee, C., Mittermaier, M., Gertler, M., et al. (2023). A deep-learning algorithm to classify skin lesions from mpox virus infection. *Nat. Med.* 29, 738–747. doi: 10.1038/s41591-023-02225-7

Tunc, H., Dogan, B., Darendeli Kiraz, B. N., Sari, M., Durdagi, S., and Kotil, S. (2023). Prediction of HIV-1 protease resistance using genotypic, phenotypic, and molecular information with artificial neural networks. *PeerJ* 11:e14987. doi: 10.7717/peerj.14987

Wang, Y., Li, Y., Chen, X., and Zhao, L. (2023). HIV-1/HBV coinfection accurate multitarget prediction using a graph neural network-based ensemble predicting model. *Int. J. Mol. Sci.* 24:7139. doi: 10.3390/ijms24087139

Wei, W., He, X., Bao, X., Wang, G., Luo, Q., Chen, L., et al. (2023). Application of deep learning algorithm in the recognition of cryptococcosis and talaromycosis skin lesions. *Mycoses* 66, 671–679. doi: 10.1111/myc.13598

Westerberg, D. P., and Voyack, M. J. (2013). Onychomycosis: current trends in diagnosis and treatment. *Am. Fam. Physician* 88, 762–770

Wong, F., Zheng, E. J., Valeri, J. A., Donghia, N. M., Anahtar, M. N., Omori, S., et al. (2024). Discovery of a structural class of antibiotics with explainable deep learning. *Nature* 626, 177–185. doi: 10.1038/s41586-023-06887-8

Xiang, Y., Du, J., Fujimoto, K., Li, F., Schneider, J., and Tao, C. (2022). Application of artificial intelligence and machine learning for HIV prevention interventions. *Lancet HIV* 9, e54–e62. doi: 10.1016/S2352-3018(21)00247-2

Xu, J., Luo, Y., Wang, J., Tu, W., Yi, X., Xu, X., et al. (2023). Artificial intelligence-aided rapid and accurate identification of clinical fungal infections by single-cell Raman spectroscopy. *Front. Microbiol.* 14:1125676. doi: 10.3389/fmicb.2023.1125676

Xu, X., Yu, Z., Ge, Z., Chow, E. P. F., Bao, Y., Ong, J. J., et al. (2022). Web-based risk prediction tool for an Individual's risk of HIV and sexually transmitted infections using machine learning algorithms: development and external validation study. *J. Med. Internet Res.* 24:e37850. doi: 10.2196/37850

Yang, Y., Guo, L., Wu, Q., Zhang, M., Zeng, R., Ding, H., et al. (2021). Construction and evaluation of a deep learning model for assessing acne vulgaris using clinical images. *Dermatol. Ther.* 11, 1239–1248. doi: 10.1007/s13555-021-00541-9

Yilmaz, A., Goktay, F., Varol, R., Gencoglan, G., and Uvet, H. (2022). Deep convolutional neural networks for onychomycosis detection using microscopic images with KOH examination. *Mycoses* 65, 1119–1126. doi: 10.1111/myc.13498

Zazzi, M., Incardona, F., Rosen-Zvi, M., Prosperi, M., Lengauer, T., Altmann, A., et al. (2012). Predicting response to antiretroviral treatment by machine learning: the EuResist project. *Intervirology* 55, 123–127. doi: 10.1159/000332008

Zhu, X., Zheng, B., Cai, W., Zhang, J., Lu, S., Li, X., et al. (2022). Deep learning-based diagnosis models for onychomycosis in dermoscopy. *Mycoses* 65, 466–472. doi: 10.1111/myc.13427

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 November 2024
DOI 10.3389/fmicb.2024.1495432

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Yongguo Li,
First Affiliated Hospital of Chongqing Medical University, China
Hong Mei Gao,
Tianjin First Central Hospital, China
Chunli Liu,
Shandong Public Health Clinical Center, China

*CORRESPONDENCE
Guangyu Jiao
✉ jiaogy@sj-hospital.org

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

RECEIVED 12 September 2024
ACCEPTED 22 October 2024
PUBLISHED 06 November 2024

CITATION
Kong W, Liu Y, Li W, Yang K, Yu L and Jiao G (2024) Correlation between oxygenation function and laboratory indicators in COVID-19 patients based on non-enhanced chest CT images and construction of an artificial intelligence prediction model.
Front. Microbiol. 15:1495432.
doi: 10.3389/fmicb.2024.1495432

Correlation between oxygenation function and laboratory indicators in COVID-19 patients based on non-enhanced chest CT images and construction of an artificial intelligence prediction model

Weiheng Kong[1†], Yujia Liu[2†], Wang Li[3], Keyi Yang[1], Lixin Yu[1] and Guangyu Jiao[1*]

[1]Department of Pulmonary and Critical Care Medicine, Shengjing Hospital of China Medical University, Shenyang, China, [2]College of Traditional Chinese Medicine, Liaoning University of Traditional Chinese Medicine, Shenyang, China, [3]Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China

Objective: By extracting early chest CT radiomic features of COVID-19 patients, we explored their correlation with laboratory indicators and oxygenation index (PaO_2/FiO_2), thereby developed an Artificial Intelligence (AI) model based on radiomic features to predict the deterioration of oxygenation function in COVID-19 patients.

Methods: This retrospective study included 384 patients with COVID-19, whose baseline information, laboratory indicators, oxygenation-related parameters, and non-enhanced chest CT images were collected. Utilizing the PaO_2/FiO_2 stratification proposed by the Berlin criteria, patients were divided into 4 groups, and differences in laboratory indicators among these groups were compared. Radiomic features were extracted, and their correlations with laboratory indicators and the PaO_2/FiO_2 were analyzed, respectively. Finally, an AI model was developed using the PaO_2/FiO_2 threshold of less than 200 mmHg as the label, and the model's performance was assessed using the area under the receiver operating characteristic curve (AUC), sensitivity and specificity. Group datas comparison was analyzed using SPSS software, and radiomic features were extracted using Python-based Pyradiomics.

Results: There were no statistically significant differences in baseline characteristics among the groups. Radiomic features showed differences in all 4 groups, while the differences in laboratory indicators were inconsistent, with some PaO_2/FiO_2 groups showed differences and others not. Regardless of whether laboratory indicators demonstrated differences across different PaO_2/FiO_2 groups, they could all be captured by radiomic features. Consequently, we chose radiomic features as variables to establish an AI model based on chest CT radiomic features. On the training set, the model achieved an AUC of 0.8137 (95% CI [0.7631–0.8612]), accuracy of 0.7249, sensitivity of 0.6626 and specificity of 0.8208. On the validation set, the model achieved an AUC of 0.8273 (95% CI [0.7475–0.9005]), accuracy of 0.7739, sensitivity of 0.7429 and specificity of 0.8222.

Conclusion: This study found that the early chest CT radiomic features of COVID-19 patients are strongly associated not only with early laboratory indicators but also with the lowest PaO_2/FiO_2. Consequently, we developed an

AI model based on CT radiomic features to predict deterioration in oxygenation function, which can provide a reliable basis for further clinical management and treatment.

KEYWORDS

SARS-CoV-2, COVID-19, artificial intelligence, machine learning, chest CT radiomic features, PaO_2/FiO_2, laboratory indicators

1 Introduction

Coronavirus Disease 2019 (COVID-19) is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and has become a global pandemic threatening worldwide health (Sudre et al., 2021). SARS-CoV-2 infection can affect multiple organs and presents a variety of clinical symptoms (Wang et al., 2023). In the pathogenesis of COVID-19, a key factor is the dysregulation of immune inflammation (Xu et al., 2020). SARS-CoV-2 primarily enters respiratory epithelial cells by binding to angiotensin-converting enzyme 2 (ACE-2), triggering an immune inflammatory responses that results in varying degrees of damage to the alveolar epithelium, formation of hyaline membranes, and lung consolidation (Camporota et al., 2022; Caso et al., 2020; Qin et al., 2023). Therefore, the clinical symptoms of patients infected with COVID-19 exhibit significant heterogeneity; some patients are asymptomatic or exhibit only mild upper respiratory symptoms, while others may develop respiratory distress, potentially progressing to Acute Respiratory Distress Syndrome (ARDS) (Poston et al., 2020). The lungs are the organs most affected early and severely by COVID-19, and the rapid deterioration in respiratory function due to lung damage is a major cause of the high mortality rate in COVID-19 patients (Torres-Castro et al., 2021; Huang et al., 2020).

Clinically, the PaO_2/FiO_2 is used to represent oxygenation function and serves as a reliable predictor of acute lung injury (Matsubara et al., 2024). Since oxygenation dysfunction is an independent risk factor for progression to severe/critical COVID-19, deterioration in the PaO_2/FiO_2 provides an important basis for early clinical identification of worsening conditions in COVID-19 patients (Zhang et al., 2021). However, some critically ill patients may have mild clinical manifestations early in the disease, which do not correspond to the degree of oxygenation dysfunction due to severe lung damage (Tobin et al., 2020). Several laboratory indicators, such as lymphocytes, neutrophils, and pro-inflammatory cytokines, have been studied for predicting disease worsening and severe outcomes in COVID-19 patients (Del Valle et al., 2020; Zhao et al., 2020). Although these indicators reflect the immune-inflammatory status after SARS-CoV-2 infection, they are not directly indicative of oxygenation function and the extent of lung damage. Research by Fatima N et al. suggested a good correlation between early chest CT images and the PaO_2/FiO_2 in COVID-19 patients, indicating that chest CT can effectively assess the extent of lung damage and has potential for predicting severe cases of COVID-19 (Fatima et al., 2023; Liu F. et al., 2020).

Currently, semi-quantitative chest CT scoring systems have been developed to predict the severity and clinical outcomes of COVID-19 patients. However, these systems require radiologists to visually assess all chest CT images, which introduces considerable human error and prevents precise assessment (Wasilewski et al., 2020). Additionally, manual annotation of all infected areas for training leads to a substantial workload, making routine application challenging (Arian et al., 2023). To improve the sensitivity of COVID-19 assessment, AI-assisted quantitative analysis of chest CT is emerging as a new trend (Shaikh et al., 2021). Limited existing AI studies have extracted features such as lung lesion volume, inflammation area, and lesion density from chest CT images, with sample sizes generally around 100 cases, which limits comprehensive assessment of lung damage (Zhang et al., 2020; Pu et al., 2021; Pang et al., 2021). There is a pressing need to extract more lung features from larger samples to develop AI models that meet clinical needs for predicting severe lung damage in COVID-19 patients. Currently, researches based on AI primarily focus on employing AI techniques to analyze the different imaging findings presented in chest CT images of COVID-19 patients in order to predict disease severity and prognosis (Arian et al., 2023; Cai et al., 2020). There is a lack of comparative studies regarding oxygenation function and chest CT images using AI.

Therefore, this study will analyze the early chest CT radiomic features of COVID-19 patients using the PaO_2/FiO_2 as a stratification standard, exploring the correlation between early laboratory indicators, early chest CT radiomic features, and the PaO_2/FiO_2. We aim to establish an AI model to predict the extent of lung injury and deterioration in oxygenation function, providing a reliable basis for the early clinical management and treatment of COVID-19 patients.

2 Methods

2.1 Study subjects and clinical data

This retrospective study included patients admitted to our hospital from January 1, 2023, to June 1, 2024, with a diagnosis of novel coronavirus infection.

Inclusion criteria:

1. Diagnosed with novel coronavirus infection upon admission (National Health Commission, 2023).
2. Underwent CT examination on the day of admission and multiple blood gas analyses during the hospital stay.
3. Aged ≥18 years.

Cases that may interfere with this study or where obtaining imaging data is challenging will be excluded, including:

1. Patients requiring mechanical ventilation.
2. Pregnant patients or those with end-stage cancer.
3. Patients with concurrent pulmonary diseases such as pneumothorax, pulmonary edema, or mediastinal emphysema.

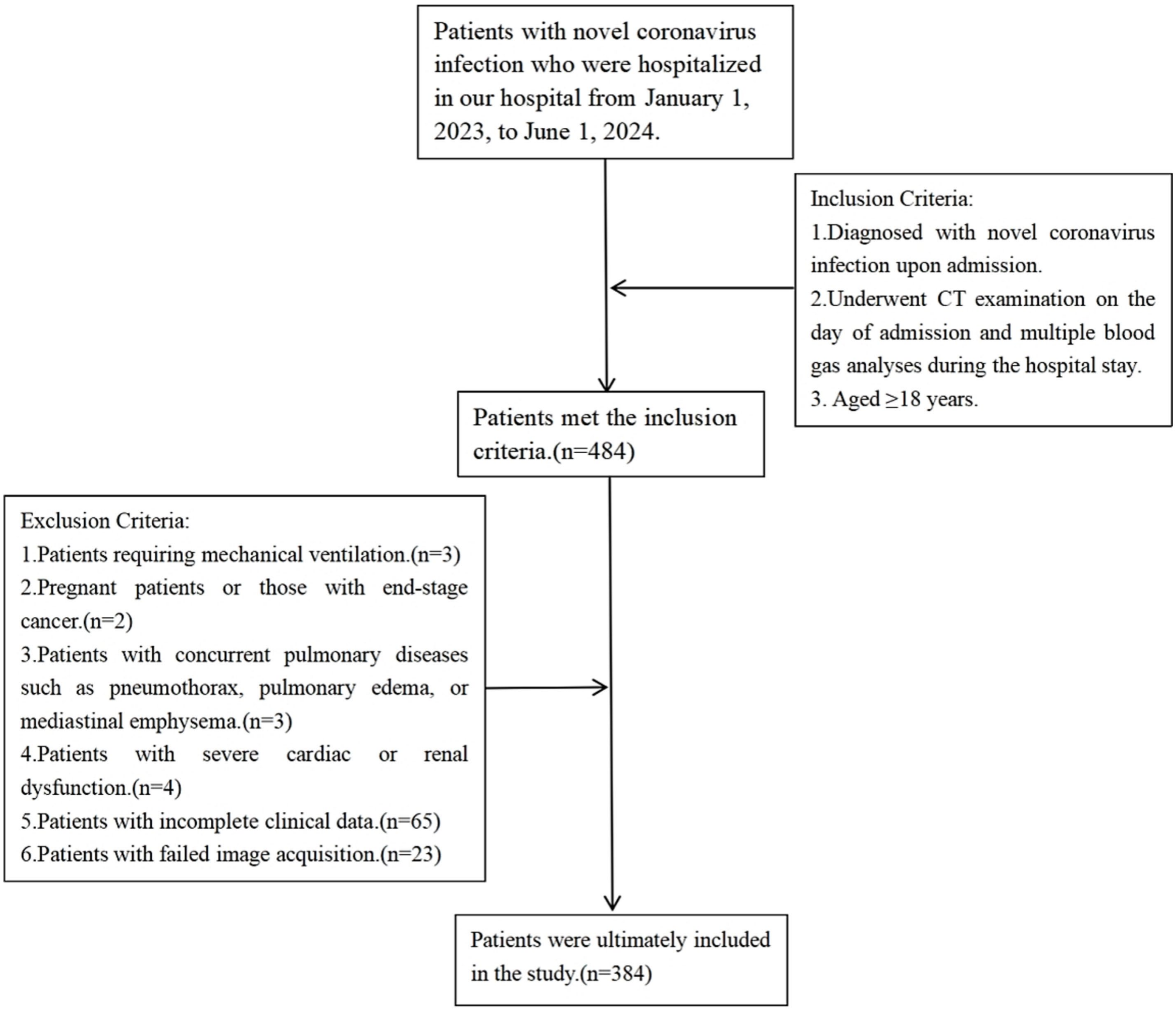


FIGURE 1
Flowchart for inclusion and exclusion of patients.

4 Patients with severe cardiac or renal dysfunction.
5 Patients with incomplete clinical data (Patients with incomplete laboratory indicators relevant to this study).
6 Patients with failed image acquisition (Cases with poor image quality or missing key frames during the CT imaging process).

A total of 384 patients were ultimately included in the study. The flowchart for the inclusion and exclusion of patients is shown in Figure 1.

Relevant clinical and laboratory data from the included patients will be collected, including: Baseline Characteristics: Age, sex, BMI, smoking history, and comorbidities. Blood Gas Analysis and PaO_2/FiO_2: Blood gas analysis results (partial pressure of oxygen, PaO_2), oxygen concentration (FiO_2), and calculation of the PaO_2/FiO_2, with the lowest PaO_2/FiO_2 during hospitalization recorded. Laboratory Indicators on Admission Day: White blood cell count, neutrophil count and percentage, lymphocyte count and percentage, platelet count, C-reactive protein (CRP), D-dimer, lactate dehydrogenase (LDH), interleukin-6 (IL-6), ferritin, liver function indicators (AST, ALT), and cardiac indicators (B-type natriuretic peptide (BNP), troponin). Composite indicators such as the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII) = platelet count × NLR were also calculated.

Quality control measures for laboratory indicators include: all operators complied with operational procedures, with no human-induced errors. The experimental instruments were all within their calibration periods. Reagents, quality control materials, and calibration standards for each indicator were all within their expiration dates and were properly stored. The laboratory environment's temperature and humidity were maintained within acceptable ranges. During the experiments, all indicators passed quality control, with no random or systematic errors observed.

2.2 CT imaging protocol

Chest non-enhanced CT imaging was performed on the day of admission. All scans were conducted in the supine position with the patient in the inspiratory phase. The CT scans were performed using a

Philips Brilliance ICT 256-slice spiral CT scanner, with the scanning range extending from the lung apex to the level of the costophrenic angle. Scanning parameters commonly used in our center included: tube voltage of 120 kV, tube current adjusted automatically, matrix of 512 × 512, pitch of 1, conventional image thickness of 3.0 mm, and thin-slice images with 1.0 mm intervals for 3D reconstruction.

2.3 Lung segmentation and features extraction

To reduce the interference of extrathoracic factors on the model, we developed a machine learning segmentation algorithm for lung segmentation. The image matrix values were first converted to attenuation values for CT images, and pixels with attenuation values less than −700 were used as a mask. After image erosion, only the largest connected domain was retained, and the mask was then expanded again to determine it as the region of interest (ROI) for the lungs, as shown in Figure 2.

Radiomic features were extracted using Python-based Pyradiomics. Prior to feature extraction, the segmented images were preprocessed to minimize the impact of contrast and brightness variations on the radiomic features. A total of 944 radiomic features were generated for each patient, based on first-order ($n = 18$), shape ($n = 14$), texture ($n = 75$), Gaussian Laplacian filters ($n = 93$), and wavelet filters ($n = 744$).

2.4 Machine learning

In clinical practice, patients with PaO_2/FiO_2 less than 200 mmHg are considered to have moderate to severe ARDS and usually require mechanical ventilation (Ranieri et al., 2012; Qadir et al., 2024). Therefore, we used PaO_2/FiO_2 200 mmHg as the grouping criterion, dividing patients into a mechanical ventilation group ($PaO_2/FiO_2 \leq 200$ mmHg) and a non-mechanical ventilation group ($PaO_2/FiO_2 > 200$ mmHg). The machine learning models for this study, developed using the Python sklearn library, employed various machine learning methods to predict the aforementioned labels. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), sensitivity and specificity. Internal validation was used to assess the machine learning models. During model development, the entire dataset was randomly divided into training and validation sets, and five-fold cross-validation was used for model validation.

In this study, we employed the Linear Discriminant Analysis (LDA) algorithm, a form of supervised learning, for dimensionality reduction and essential feature extraction. We extracted over 900 radiomic features for each patient in the study. By utilizing this algorithm, we aimed to reduce the number of features in the input data, enabling the representation of the output affecting labels with a minimal set of features. The fundamental concept is to project the training sample set onto a single line in such a way that the projection points of samples from the same class are as close together as possible, while the centers of the projection points from different classes are as far apart as possible (Xu et al., 2022).

2.5 Statistical analysis

Statistical analysis was performed using IBM SPSS 27.0 software. All data were tested for normality; normally distributed quantitative data were described as mean ± standard deviation, while non-normally distributed quantitative data were described as

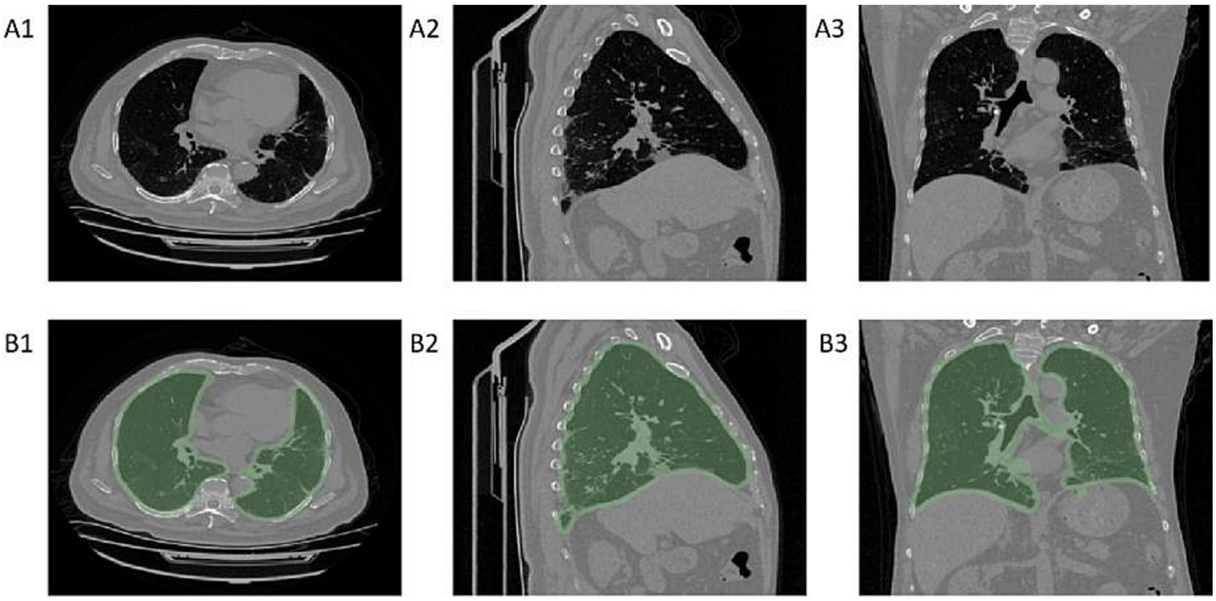


FIGURE 2
Diagram of lung segmentation. Panels **A1–A3** show chest CT cross-sectional images, while panels **B1–B3** display the regions of interest (ROI) for the lungs identified by the machine learning model on chest CT. **A1–B1** are axial CT images, **A2–B2** are sagittal CT images, and **A3–B3** are coronal CT images.

median (interquartile range). The Kruskal-Wallis test was used for comparing multiple groups, and the Mann–Whitney U test was used for multiple comparisons among groups. Categorical data were described using frequencies (percentages) and compared using the chi-square ($\chi 2$) test. A p-value of <0.05 was considered statistically significant. Statistical plots were generated using Python-based matplotlib.

TABLE 1 The clinical characteristics of the included patients.

	Characteristics	Statistical value
Cases number		384
Sex		
	Male	227 (59.1%)
	Female	157 (40.9%)
Age		71.00 (62.25, 78.00)
BMI		24.29 (22.19, 26.88)
Smoking history		
	Yes	105 (27.3%)
	No	279 (72.7%)
Comorbidity		
	Hypertension	184 (47.9%)
	Diabetes	97 (25.3%)
	Cardiovascular disease	76 (19.8%)
	COPD	5 (1.3%)
	Chronic kidney disease	23 (6.0%)
PaO_2/FiO_2 Grouping		
	>300 mmHg	106 (27.6%)
	200-300 mmHg	127 (33.1%)
	100-200 mmHg	119 (31.0%)
	≤100 mmHg	32 (8.3%)

3 Results

3.1 Basic characteristics

A total of 384 patients were included in this study. Table 1 presents the clinical characteristics of the included patients. The median age of all patients was 71.00 (62.25, 78.00) years, and the median BMI was 24.29 (22.19, 26.88). Among the patients, 227 (59.1%) were male and 157 (40.9%) were female. A total of 105 patients (27.3%) had a history of smoking. The most common comorbidities among the included patients were hypertension, diabetes, and cardiovascular diseases. Based on the Berlin definition guidelines for ARDS (Ranieri et al., 2012), patients were divided into four groups according to their PaO_2/FiO_2: $PaO_2/FiO_2 > 300$ mmHg, PaO_2/FiO_2 200-300 mmHg, PaO_2/FiO_2 100-200 mmHg, and $PaO_2/FiO_2 \leq 100$ mmHg.

There were no statistically significant differences in sex, age, BMI, smoking history, or comorbidities among the four patient groups ($p > 0.05$). See Table 2.

As the PaO_2/FiO_2 decreases, the range and density of lung lesions in the chest CT images increase. In the $PaO_2/FiO_2 > 300$ mmHg group, patients exhibit a few scattered exudative lesions in the lungs (see Figures 3A1–A3). In the PaO_2/FiO_2 200-300 mmHg group, patients show fewer lung lesions, primarily ground-glass opacities (GGOs) with limited extent (see Figures 3B1–B3). In the PaO_2/FiO_2 100-200 mmHg group, patients have a larger number of lung lesions, including GGOs and some consolidation, with a more extensive distribution (see Figures 3C1–C3). In the $PaO_2/FiO_2 \leq 100$ mmHg group, patients present with dense lung lesions, including diffuse consolidation, with widespread distribution throughout the lungs (see Figures 3D1–D3).

3.2 Analysis of differences in radiomic features across different PaO_2/FiO_2 groups

As shown in Figure 4, we compared the differences in radiomic feature expressions among different groups. There are significant

TABLE 2 Comparison of baseline characteristics among four groups of patients.

Characteristics	>300 mmHg $n = 106$	200-300 mmHg $n = 127$	100-200 mmHg $n = 119$	≤100 mmH $n = 32$	Statistical value	p value
Sex					$\chi 2 = 3.927$	0.269
Male	61 (59.1%)	68 (53.5%)	78 (65.5%)	20 (62.5%)		
Female	45 (42.5%)	59 (46.1%)	41 (34.5%)	12 (37.5%)		
Age	68.50 (60.00, 74.25)	71.00 (64.00, 78.00)	71.00 (63.00, 79.00)	72.00 (67.25, 76.75)	$H = 7.676$	0.053
BMI	23.75 (22.15, 26.83)	24.77 (21.87, 27.17)	24.22 (22.22, 26.89)	25.23 (22.71, 26.73)	$H = 1.246$	0.742
Smoking history					$\chi 2 = 4.201$	0.241
Yes	25 (23.6%)	30 (23.6%)	39 (32.8%)	11 (34.4%)		
No	84 (76.4%)	97 (76.4%)	80 (67.2%)	21 (65.6%)		
Comorbidities						
Hypertension	46 (43.4%)	58 (45.7%)	59 (49.6%)	21 (65.6%)	$\chi 2 = 5.278$	0.153
Diabetes	20 (18.9%)	32 (25.2%)	32 (26.9%)	13 (40.6%)	$\chi 2 = 6.463$	0.091
Cardiovascular disease	13 (12.3%)	27 (21.3%)	26 (21.8%)	10 (31.3%)	$\chi 2 = 6.920$	0.074
COPD	1 (0.9%)	2 (1.6%)	2 (1.7%)	0 (0.0%)	$\chi 2 = 0.735$	0.865
Chronic kidney disease	6 (5.7%)	7 (5.5%)	7 (5.9%)	3 (9.4%)	$\chi 2 = 0.727$	0.867

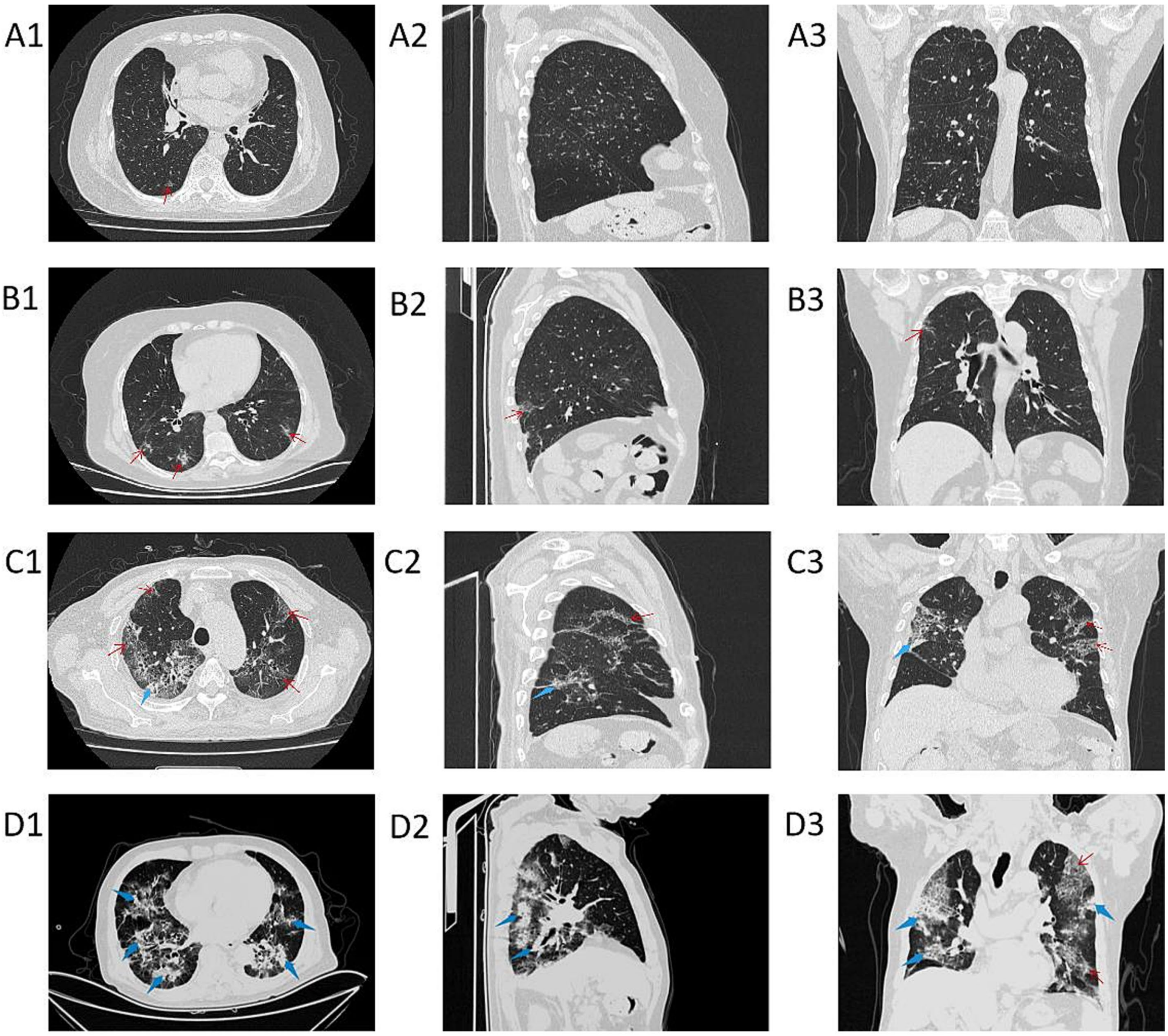


FIGURE 3
Chest CT images of patients in different PaO_2/FiO_2 groups. **A1–A3** are chest CT images of patients with $PaO_2/FiO_2 > 300$ mmHg; **B1–B3** are chest CT images of patients with PaO_2/FiO_2 200-300 mmHg; **C1–C3** are chest CT images of patients with PaO_2/FiO_2 100-200 mmHg; **D1–D3** are chest CT images of patients with $PaO_2/FiO_2 \leq 100$ mmHg. **A1–D1** show axial CT; **A2–D2** show sagittal CT; **A3–D3** show coronal CT. The red thin arrows indicate ground-glass opacities (GGOs) and interlobular septal thickening; the blue thick arrows indicate consolidation.

differences in the radiomic features among patients in different PaO_2/FiO_2 groups ($p < 0.05$). Specifically, patients in the $PaO_2/FiO_2 \leq 100$ mmHg group show the most pronounced differences in radiomic features compared to the other three groups (Figures 4A–C). As the PaO_2/FiO_2 increases, the differences in radiomic features gradually decrease (Figures 4D–F).

3.3 Analysis of laboratory indicators across different PaO_2/FiO_2 groups

Comparing laboratory indicators across different PaO_2/FiO_2 groups, we observed statistical differences in immune-inflammatory indicators, coagulation indicators, and cardiac-related indicators among the four groups ($p < 0.05$). However, no statistical differences were found in platelet counts and liver-related indicators (AST, ALT) ($p > 0.05$) (see Table 3).

To clarify the specific differences between groups, we performed pairwise post-hoc comparisons (see Figure 5). We found statistically significant differences in neutrophil percentage, lymphocyte percentage, LDH, NLR, SII, and troponin across different PaO_2/FiO_2 groups ($p < 0.05$).

However, the differences in laboratory indicators such as white blood cells, neutrophils, lymphocytes, CRP, IL-6, ferritin, D-dimer, BNP, and PLR varied inconsistently among the PaO_2/FiO_2 groups. Specifically, differences in these indicators between the $PaO_2/FiO_2 \leq 100$ mmHg group and the PaO_2/FiO_2 200-300 mmHg group, as well as the $PaO_2/FiO_2 > 300$ mmHg group, were statistically significant ($p < 0.05$). Conversely, white blood cells, D-dimer, IL-6, and ferritin showed no significant differences between the PaO_2/FiO_2 200-300 mmHg group and the $PaO_2/FiO_2 > 300$ mmHg group, or between the PaO_2/FiO_2 200-300 mmHg group and the PaO_2/FiO_2 100-200 mmHg group ($p > 0.05$). Therefore, the direct correlation between laboratory indicators and PaO_2/FiO_2 is not

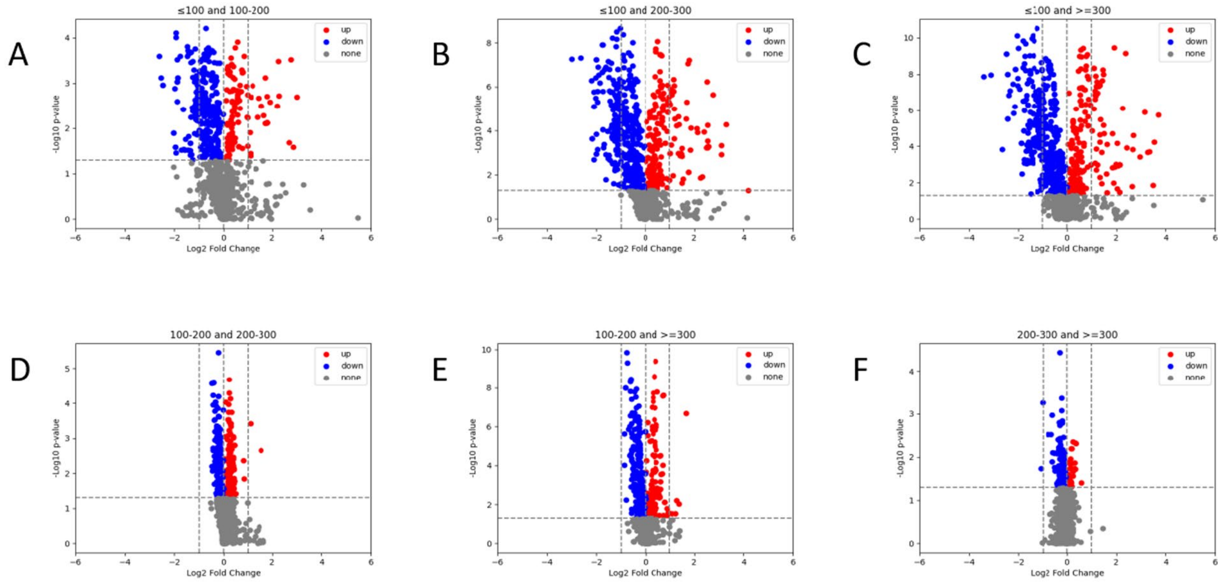


FIGURE 4
Comparison of radiomic features between different PaO_2/FiO_2 groups via volcano plot. There are significant differences in radiomic feature expressions between different PaO_2/FiO_2 groups **(A–F)**. Patients in the $PaO_2/FiO_2 \leq 100$ mmHg group show the most pronounced differences in radiomic features compared to the other three groups **(A–C)**. The differences in radiomic features between the PaO_2/FiO_2 100-200 mmHg group and the PaO_2/FiO_2 200-300 mmHg group, as well as the $PaO_2/FiO_2 \geq 300$ mmHg group, are notable **(D,E)**. There are differences in radiomic features between the PaO_2/FiO_2 200-300 mmHg group and the $PaO_2/FiO_2 \geq 300$ mmHg group, but the differences are small **(F)**. In the figure, blue points and red points represent significant differences, while gray points indicate no difference. A higher number of points indicates a greater degree of difference.

clear, and these indicators cannot accurately reflect the PaO_2/FiO_2 status.

As shown in Figure 6A, correlation analysis between the aforementioned laboratory indicators and radiomic features reveals that changes in laboratory indicators are directly reflected in the patients' chest CT images and are sharply captured by radiomic features. Among these radiomic features that can capture laboratory indicators, the majority show significant differences between different PaO_2/FiO_2 groups, regardless of whether the laboratory indicators themselves differ between the PaO_2/FiO_2 groups, as illustrated in Figure 6B.

3.4 Chest CT radiomic features model

In different PaO_2/FiO_2 groups, there are statistically significant differences in radiomic features. Additionally, regardless of whether laboratory indicators have differences between PaO_2/FiO_2 groups, they can be captured by radiomic features. Therefore, we selected only radiomic features as variables and established a chest CT radiomic features AI model. To efficiently and accurately predict whether COVID-19 patients require mechanical ventilation due to decreased PaO_2/FiO_2, we combined patients with $PaO_2/FiO_2 \leq 100$ mmHg and 100-200 mmHg into the mechanical ventilation group, and those with PaO_2/FiO_2 200-300 mmHg and > 300 mmHg into the non-mechanical ventilation group.

On the training set, the model's AUC was 0.8137 (95% CI [0.7631–0.8612]), with an accuracy of 0.7249, sensitivity of 0.6626, and specificity of 0.8208. On the validation set, the model's AUC was 0.8273 (95% CI [0.7475–0.9005]), with an accuracy of 0.7739, sensitivity of 0.7429, and specificity of 0.8222, as shown in Figure 7.

4 Discussion

This study is the first to use machine learning methods to segment lung ROIs and extract radiomic features from early chest CT images of over 380 COVID-19 patients. We utilized these radiomic features as intermediate variables to explore the direct and indirect correlations between laboratory indicators and PaO_2/FiO_2, thereby validating the assessment capability of CT—one of the most commonly used imaging modalities for COVID-19—of the overall physiological and pathological state represented by laboratory indicators. Finally, we developed an AI model based on early chest CT radiomic features of COVID-19 patients to predict whether mechanical ventilation would be required due to a decrease in PaO_2/FiO_2.

SARS-CoV-2 infection can trigger a robust immune response (Gallais et al., 2021). Early immune response in COVID-19 plays a protective role in viral clearance, whereas an excessive immune response can release an overabundance of pro-inflammatory cytokines and chemokines, leading to cytokine storms and systemic immune cascade reactions, which in turn alter laboratory immune-inflammatory indicators and coagulation indicators (Alzaabi et al., 2021; Chen, R. et al., 2020). Additionally, exacerbated and dysregulated immune responses can cause multi-organ damage, with the lungs being among the earliest and most severely affected organs (Chen N. et al., 2020). Researches by Liu and Fu et al. have demonstrated that laboratory indicators can be used to predict the overall deterioration

TABLE 3 Differential analysis of laboratory indicators among four groups of patients.

Laboratory indicators	>300 mmHg *n* = 106	200-300 mmHg *n* = 127	100-200 mmHg *n* = 119	≤100 mmHg *n* = 32	Statistical value	*p* value
Immune-inflammatory indicators						
White blood cell count (10^9/L)	6.52 (5.14, 8.69)	7.00 (4.83, 9.12)	7.20 (5.48, 9.98)	8.44 (6.45, 12.51)	$H = 13.649$	**0.003**
Neutrophil percentage (%)	69.90 (62.85, 80.60)	76.50 (68.40, 85.30)	80.20 (71.70, 87.60)	86.70 (79.93, 91.13)	$H = 45.212$	**<0.001**
Neutrophil count (10^9/L)	4.60 (3.40, 6.30)	5.10 (3.20, 7.00)	5.90 (4.10, 8.10)	7.55 (5.10, 11.23)	$H = 27.300$	**<0.001**
Lymphocyte percentage (%)	19.30 (11.03, 26.53)	14.80 (8.50, 20.30)	10.30 (5.90, 16.90)	5.75 (3.48, 12.15)	$H = 55.444$	**<0.001**
Lymphocyte count (10^9/L)	1.20 (0.70, 1.53)	0.90 (0.60, 1.40)	0.80 (0.50, 1.10)	0.55 (0.40, 0.90)	$H = 28.781$	**<0.001**
CRP (mg/L)	15.55 (5.46, 48.01)	25.40 (9.60, 62.00)	38.90 (9.70, 87.30)	71.99 (27.21, 100.87)	$H = 23.440$	**<0.001**
LDH (U/L)	270.00 (21.75, 301.00)	297.00 (242.00, 309.00)	301.00 (280.00, 343.00)	341.00 (287.25, 430.25)	$H = 37.108$	**<0.001**
IL-6 (pg/mL)	7.12 (2.39, 16.27)	11.55 (3.17, 20.35)	12.90 (3.95, 27.02)	16.85 (7.65, 30.82)	$H = 16.268$	**<0.001**
Ferritin (ng/mL)	316.50 (226.48, 514.48)	387.00 (241.10, 537.50)	486.40 (267.00, 573.50)	537.50 (367.95, 705.13)	$H = 18.150$	**<0.001**
NLR	3.6667 (2.4152, 7.5000)	4.8889 (3.2500, 10.0000)	7.7500 (4.4444, 15.5000)	14.0417 (6.4560, 24.5000)	$H = 53.106$	**<0.001**
PLR	184.7802 (117.3438, 291.2500)	202.5000 (146.1905, 350.0000)	283.3333 (173.3333, 410.0000)	294.1667 (204.5000, 512.5000)	$H = 27.752$	**<0.001**
SII	763.0833 (464.8472, 1605.5000)	1147.0000 (540.5714, 2141.6667)	1515.5556 (854.0000, 2908.8889)	2527.0000 (1205.5130, 6080.9583)	$H = 41.347$	**<0.001**
Coagulation indicators						
platelet count (10^9/L)	200.00 (149.75, 258.00)	199.00 (142.00, 261.00)	207.00 (148.00, 287.00)	179.00 (141.00, 257.75)	$H = 1.426$	0.699
D-dime r (μg/L)	180.00 (96.75, 425.50)	215.00 (133.00, 499.00)	264.00 (160.00, 486.00)	599.00 (257.75, 2818.75)	$H = 29.910$	**<0.001**
Liver-related indicators						
AST (U/L)	21.50 (16.00, 31.75)	22.00 (15.00, 33.00)	24.00 (17.00, 35.00)	27.00 (22.00, 44.00)	$H = 7.235$	0.065
ALT (U/L)	23.00 (16.00, 37.50)	27.00 (18.00, 43.00)	28.00 (18.00, 42.00)	34.00 (18.50, 50.75)	$H = 4.934$	0.177
Cardiac-related indicators						
BNP (pg/mL)	50.15 (17.28, 82.03)	69.00 (16.30, 100.31)	80.30 (35.30, 122.00)	67.25 (33.93, 146.40)	$H = 16.990$	**<0.001**
Troponin (μg/L)	0.0056 (0.0038, 0.0088)	0.0072 (0.0041, 0.0113)	0.0089 (0.0058, 0.0160)	0.0120 (0.0077, 0.0345)	$H = 36.531$	**<0.001**

The values in bold indicate that the *p*-values are less than 0.05, meaning there are significant differences among the four groups of data with statistical significance.

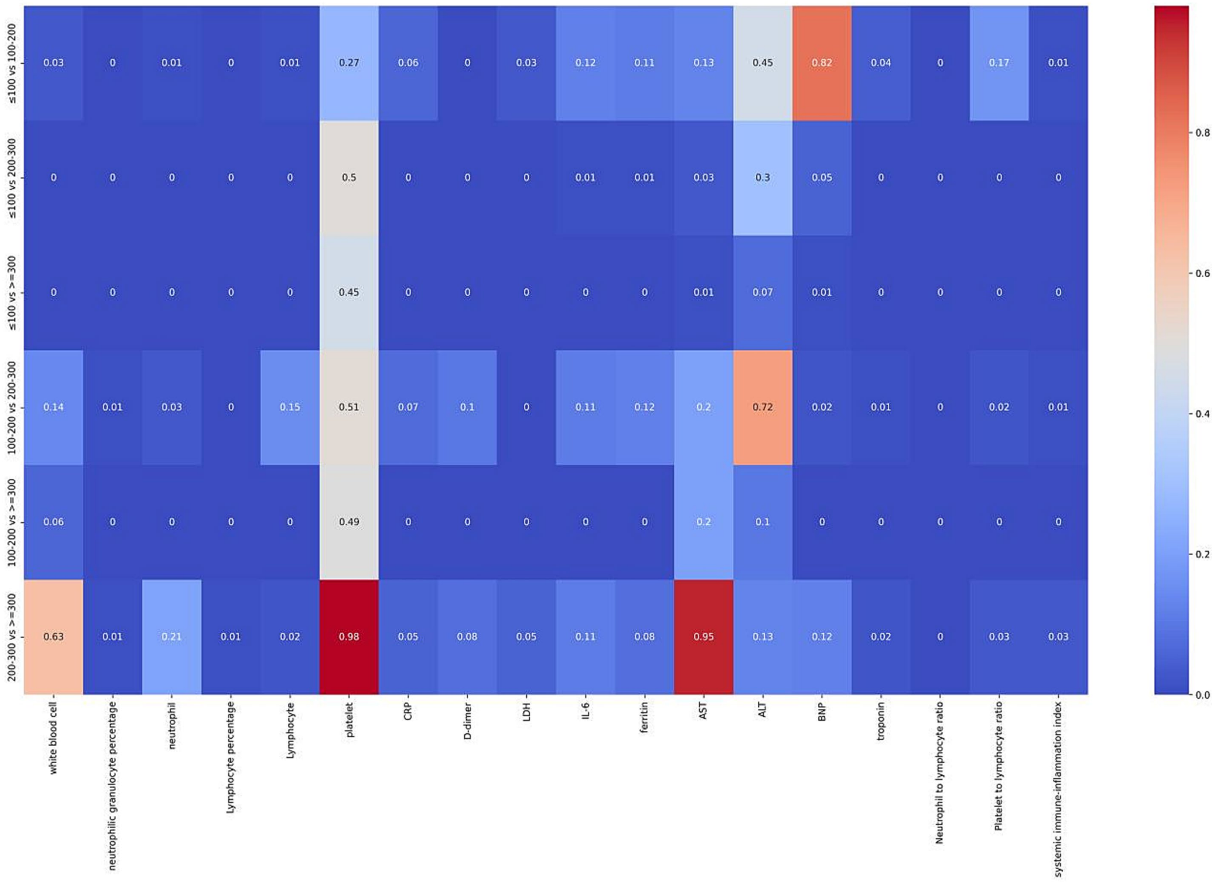


FIGURE 5
Heatmap of differences in laboratory indicators between pairs of PaO_2/FiO_2 groups. The x-axis represents laboratory indicators, and the y-axis represents pairwise comparisons between PaO_2/FiO_2 groups. Differences were analyzed using the Mann–Whitney *U* test. The bar in the figure indicates the *p*-value (0–1), with $p < 0.05$ indicating statistical significance and $p > 0.05$ indicating no statistical significance.

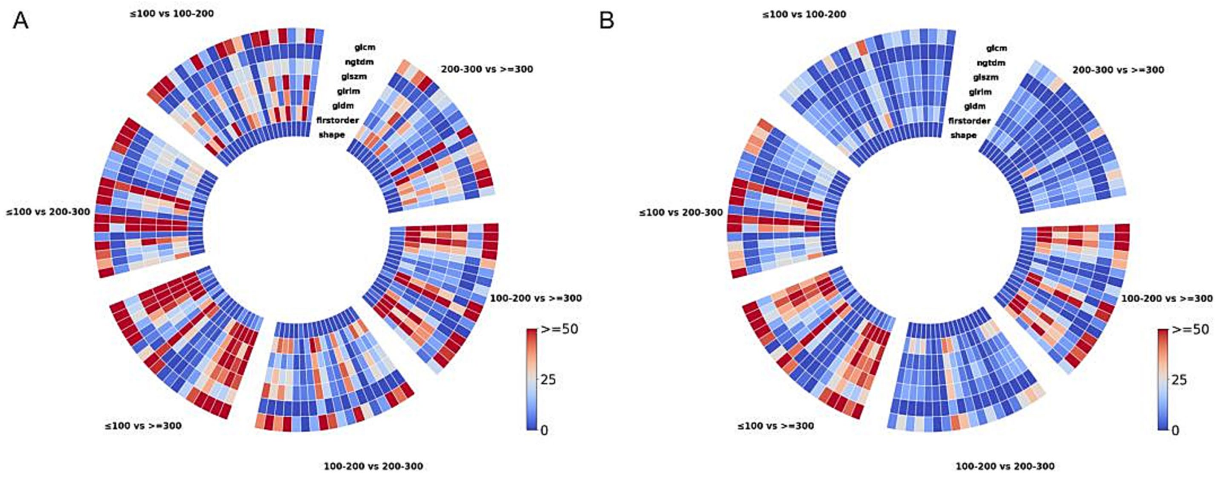


FIGURE 6
Texture feature map showing the correlation between laboratory indicators and radiomic features. In pairwise comparisons of different PaO_2/FiO_2 groups, different laboratory indicators are directly captured by the intensity of radiomic features **(A)**. Radiomic features that both capture laboratory indicators and show differences between different PaO_2/FiO_2 groups are illustrated in **(B)**.

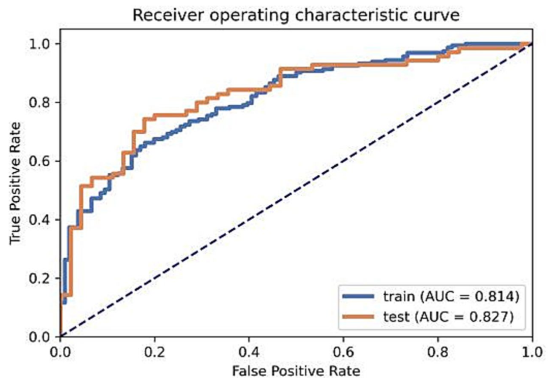


FIGURE 7
The ROC curves for the training and validation sets.

and adverse outcomes in COVID-19 patients (Fu et al., 2020; Liu J. et al., 2020). However, to date, there have been no studies directly predicting the degree of lung injury and oxygenation function through the analysis of changes in laboratory indicators. To address this, we explored whether multiple laboratory indicators directly correlate with the PaO_2/FiO_2. Results showed that while there are overall differences in laboratory indicators across four different PaO_2/FiO_2 groups, the differences are inconsistent when comparing pairwise groups. For example, white blood cell count, CRP, IL-6, and ferritin were statistically significant between the $PaO_2/FiO_2 \leq 100$ mmHg group and the PaO_2/FiO_2 200-300 mmHg group, but there were no statistical differences between the PaO_2/FiO_2 200-300 mmHg group and the $PaO_2/FiO_2 > 300$ mmHg group. These results suggest that laboratory indicators alone do not fully and accurately assess oxygenation function and the extent of lung damage in COVID-19 patients. Therefore, more appropriate assessment indicators are needed clinically.

We subsequently focused on early chest CT images to analyze their correlation with the PaO_2/FiO_2. We found significant differences in radiomic features among different PaO_2/FiO_2 groups, particularly in the patients with $PaO_2/FiO_2 \leq 100$ mmHg, whose radiomic features showed a very significant difference compared to the other three groups. Thus, the radiomic features derived from non-contrast chest CT images may provide a valuable tool for predicting the PaO_2/FiO_2. Furthermore, we conducted a correlation analysis between radiomic features and laboratory indicators, revealing significant correlations between them. Notably, even laboratory indicators that were not directly related to the PaO_2/FiO_2 showed a strong association with the radiomic features, indicating that these features may also serve as accurate reflections of the body's inflammatory response level.

This study is the first to extract radiomic features from early chest CT scans of over 380 COVID-19 patients, using the PaO_2/FiO_2 as the stratification criterion. We established an AI model based on early chest CT radiomic features, which achieve an accuracy of 0.8 in predicting stratification for the PaO_2/FiO_2 above and below 200 mmHg. Although AI-driven quantitative analysis of CT scans has shown promise in assessing clinical classifications, predicting disease progression, and evaluating sequelae in COVID-19 patients, the current research often relies on radiologists visually assessing and manually annotating CT images (Salahshour et al., 2021; Tanaka et al., 2023; Wasilewski et al., 2020). This heavy workload limits the ability to evaluate large samples, and reducing human error remains a significant challenge. Furthermore, studies utilizing AI technology for CT imaging primarily focus on identifying and analyzing specific features such as lesion volume, inflammatory area, or lesion density (Pang et al., 2021; Alilou et al., 2023; Chung et al., 2021). This narrow focus may lead to incomplete assessments and, similarly, suffers from issues related to high error margins and low accuracy.

This study departs from traditional visual assessment methods by disruptively applying computer programming languages to extract over 900 radiomic numerical features from CT images, including first-order, shape, texture, Gaussian Laplacian filters, and wavelet filters. Using machine learning for training and validation, we ultimately selected the feature parameter combinations most strongly correlated with the PaO_2/FiO_2 to construct a CT-AI model for lung assessment, achieving high accuracy and specificity in predicting oxygenation function. Clinically, patients with an PaO_2/FiO_2 below 200 mmHg generally require mechanical ventilation (Qadir et al., 2024). Santus P and Zhou W have confirmed that an $PaO_2/FiO_2 < 200$ mmHg at admission is independently associated with higher mortality, which can help clinicians identify high-risk patients early in their hospital stay (Santus et al., 2020; Zhou et al., 2021). Therefore, we selected PaO_2/FiO_2 200 mmHg as the threshold value in clinical practice, dividing patients into two groups: the mechanical ventilation group (including the $PaO_2/FiO_2 \leq 100$ mmHg group and the PaO_2/FiO_2 100–200 mmHg group) and the non-mechanical ventilation group (including the PaO_2/FiO_2 200–300 mmHg group and the $PaO_2/FiO_2 > 300$ mmHg group). The results indicate that this model can predict stratification tasks with an accuracy of 0.8 for determining whether the PaO_2/FiO_2 is above or below 200 mmHg. This capability can assist clinicians in automatically identifying high-risk patients through early admission CT scans, effectively guiding the monitoring of critically ill patients, the need for increased oxygen supplementation, and decisions regarding mechanical ventilation.

This study does have some limitations. First, it is a single-center study, lacking multi-center data to further validate these conclusions. Second, the study only explored the AI model's ability to predict the lowest PaO_2/FiO_2 during hospitalization, lacking comprehensive monitoring throughout the patient's disease course, the further model can be established for dynamic monitoring and the prediction of the long COVID-19 in the future. Third, we used only the PaO_2/FiO_2 as the primary parameter for assessing COVID-19 severity, without considering other complications that may arise during the disease course. Finally, given the high heterogeneity of COVID-19, future research will further explore their corresponding mechanism and the impact of genetic susceptibility on the PaO_2/FiO_2.

5 Conclusion

This study found that the early chest CT radiomic features of COVID-19 patients show a strong correlation with early laboratory indicators and the lowest PaO_2/FiO_2. Therefore, we established an AI model based on the early chest CT radiomic characteristics of COVID-19 patients, which can be used to predict the deterioration of oxygenation function in COVID-19 patients, providing a basis for selecting further clinical management and treatment measures.

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 the Ethics Committee of Shengjing Hospital Affiliated to China Medical University (Ethics Number: 2021PS621K). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

WK: Writing – original draft, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation. YL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. WL: Data curation, Software, Writing – original draft. KY: Investigation, Validation, Writing – original draft. LY: Investigation, Writing – original draft. GJ: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Liaoning University of Traditional Chinese Medicine's Student Innovation and Entrepreneurship Training Program Project (202310162004x); Applied Basic Research Project of Liaoning Province in 2022 (2022JH2/101500051).

Acknowledgments

Thanks to the authors of the included studies to provide primary datas and develop the model.

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/fmicb.2024.1495432/full#supplementary-material

References

Alilou, S., Zangiabadian, M., Pouramini, A., Jaberinezhad, M., Shobeiri, P., Ghozy, S., et al. (2023). Radiological findings as predictors of COVID-19 lung sequelae: a systematic review and Meta-analysis. *Acad. Radiol.* 30, 3076–3085. doi: 10.1016/j.acra.2023.06.002

Alzaabi, A. H., Ahmed, L. A., Rabooy, A. E., Zaabi, A. A., Alkaabi, M., AlMahmoud, F., et al. (2021). Longitudinal changes in IgG levels among COVID-19 recovered patients: a prospective cohort study. *PLoS One* 16:e0251159. doi: 10.1371/journal.pone.0251159

Arian, A., Mehrabi Nejad, M. M., Zoorpaikar, M., Hasanzadeh, N., Sotoudeh-Paima, S., Kolahi, S., et al. (2023). Accuracy of artificial intelligence CT quantification in predicting COVID-19 subjects' prognosis. *PLoS One* 18:e0294899. doi: 10.1371/journal.pone.0294899

Cai, W., Liu, T., Xue, X., Luo, G., Wang, X., Shen, Y., et al. (2020). CT quantification and machine-learning models for assessment of disease severity and prognosis of COVID-19 patients. *Acad. Radiol.* 27, 1665–1678. doi: 10.1016/j.acra.2020.09.004

Camporota, L., Cronin, J. N., Busana, M., Gattinoni, L., and Formenti, F. (2022). Pathophysiology of coronavirus-19 disease acute lung injury. *Curr. Opin. Crit. Care* 28, 9–16. doi: 10.1097/MCC.0000000000000911

Caso, F., Costa, L., Ruscitti, P., Navarini, L., Del Puente, A., Giacomelli, R., et al. (2020). Could Sars-coronavirus-2 trigger autoimmune and/or autoinflammatory mechanisms in genetically predisposed subjects? *Autoimmun. Rev.* 19:102524. doi: 10.1016/j.autrev.2020.102524

Chen, R., Sang, L., Jiang, M., Yang, Z., Jia, N., Fu, W., et al. (2020). Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J. Allergy Clin. Immunol.* 146, 89–100. doi: 10.1016/j.jaci.2020.05.003

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395, 507–513. doi: 10.1016/S0140-6736(20)30211-7

Chung, H., Ko, H., Kang, W. S., Kim, K. W., Lee, H., Park, C., et al. (2021). Prediction and feature importance analysis for severity of COVID-19 in South Korea using artificial intelligence: model development and validation. *J. Med. Internet Res.* 23:e27060. doi: 10.2196/27060

Del Valle, D. M., Kim-Schulze, S., Huang, H. H., Beckmann, N. D., Nirenberg, S., Wang, B., et al. (2020). An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* 26, 1636–1643. doi: 10.1038/s41591-020-1051-9

Fatima, N., Khokhar, S. A., and Farooq Ur Rehman, R. M. (2023). Correlation between oxygen saturation of patient and severity index of Covid 19 pneumonia on CT. *J. Pak. Med. Assoc.* 73, 60–63. doi: 10.47391/JPMA.5586

Fu, J., Kong, J., Wang, W., Wu, M., Yao, L., Wang, Z., et al. (2020). The clinical implication of dynamic neutrophil to lymphocyte ratio and D-dimer in COVID-19: a retrospective study in Suzhou China. *Thromb. Res.* 192, 3–8. doi: 10.1016/j.thromres.2020.05.006

Gallais, F., Velay, A., Nazon, C., Wendling, M. J., Partisani, M., Sibilia, J., et al. (2021). Intrafamilial exposure to SARS-CoV-2 associated with cellular immune response without seroconversion, France. *Emerg. Infect. Dis.* 27, 113–121. doi: 10.3201/eid2701.203611

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/S0140-6736(20)30183-5

Liu, J., Li, S., Liu, J., Liang, B., Wang, X., Wang, H., et al. (2020). Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine* 55:102763. doi: 10.1016/j.ebiom.2020.102763

Liu, F., Zhang, Q., Huang, C., Shi, C., Wang, L., Shi, N., et al. (2020). CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* 10, 5613–5622. doi: 10.7150/thno.45985

Matsubara, S., Sudo, K., Kushimoto, K., Yoshii, R., Inoue, K., Kinoshita, M., et al. (2024). Prediction of acute lung injury assessed by chest computed tomography, oxygen saturation/fraction of inspired oxygen ratio, and serum lactate dehydrogenase in patients with COVID-19. *J. Infect. Chemother.* 30, 406–416. doi: 10.1016/j.jiac.2023.11.013

National Health Commission (2023). NHC of the PRC and National Administration of Traditional Chinese Medicine of the PRC. 10th version of the National Health Commission of China's guidelines for diagnosis and treatment of novel coronavirus. Beijing: National Health Commission.

Pang, B., Li, H., Liu, Q., Wu, P., Xia, T., Zhang, X., et al. (2021). CT quantification of COVID-19 pneumonia at admission can predict progression to critical illness: a retrospective multicenter cohort study. *Front. Med.* 8:689568. doi: 10.3389/fmed.2021.689568

Poston, J. T., Patel, B. K., and Davis, A. M. (2020). Management of Critically ill Adults with COVID-19. *JAMA* 323, 1839–1841. doi: 10.1001/jama.2020.4914

Pu, J., Leader, J. K., Bandos, A., Ke, S., Wang, J., Shi, J., et al. (2021). Automated quantification of COVID-19 severity and progression using chest CT images. *Eur. Radiol.* 31, 436–446. doi: 10.1007/s00330-020-07156-2

Qadir, N., Sahetya, S., Munshi, L., Summers, C., Abrams, D., Beitler, J., et al. (2024). An update on Management of Adult Patients with acute respiratory distress syndrome: an official American Thoracic Society clinical practice guideline. *Am. J. Respir. Crit. Care Med.* 209, 24–36. doi: 10.1164/rccm.202311-2011ST

Qin, R., He, L., Yang, Z., Jia, N., Chen, R., Xie, J., et al. (2023). Identification of parameters representative of immune dysfunction in patients with severe and fatal COVID-19 infection: a systematic review and Meta-analysis. *Clin. Rev. Allergy Immunol.* 64, 33–65. doi: 10.1007/s12016-021-08908-8

Ranieri, V. M., Rubenfeld, G. D., Thompson, B. T., Ferguson, N. D., Caldwell, E., Fan, E., et al. (2012). Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 307, 2526–2533. doi: 10.1001/jama.2012.5669

Salahshour, F., Mehrabinejad, M. M., Nassiri Toosi, M., Gity, M., Ghanaati, H., Shakiba, M., et al. (2021). Clinical and chest CT features as a predictive tool for COVID-19 clinical progress: introducing a novel semi-quantitative scoring system. *Eur. Radiol.* 31, 5178–5188. doi: 10.1007/s00330-020-07623-w

Santus, P., Radovanovic, D., Saderi, L., Marino, P., Cogliati, C., De Filippis, G., et al. (2020). Severity of respiratory failure at admission and in-hospital mortality in patients with COVID-19: a prospective observational multicentre study. *BMJ Open* 10:e043651. doi: 10.1136/bmjopen-2020-043651

Shaikh, F., Andersen, M. B., Sohail, M. R., Mulero, F., Awan, O., Dupont-Roettger, D., et al. (2021). Current landscape of imaging and the potential role for artificial intelligence in the management of COVID-19. *Curr. Probl. Diagn. Radiol.* 50, 430–435. doi: 10.1067/j.cpradiol.2020.06.009

Sudre, C. H., Murray, B., Varsavsky, T., Graham, M. S., Penfold, R. S., Bowyer, R. C., et al. (2021). Attributes and predictors of long COVID. *Nat. Med.* 27, 626–631. doi: 10.1038/s41591-021-01292-y

Tanaka, H., Maetani, T., Chubachi, S., Tanabe, N., Shiraishi, Y., Asakura, T., et al. (2023). Clinical utilization of artificial intelligence-based COVID-19 pneumonia quantification using chest computed tomography - a multicenter retrospective cohort study in Japan. *Respir. Res.* 24:241. doi: 10.1186/s12931-023-02530-2

Tobin, M. J., Laghi, F., and Jubran, A. (2020). Why COVID-19 silent hypoxemia is baffling to physicians. *Am. J. Respir. Crit. Care Med.* 202, 356–360. doi: 10.1164/rccm.202006-2157CP

Torres-Castro, R., Vasconcello-Castillo, L., Alsina-Restoy, X., Solis-Navarro, L., Burgos, F., Puppo, H., et al. (2021). Respiratory function in patients post-infection by COVID-19: a systematic review and meta-analysis. *Pulmonology* 27, 328–337. doi: 10.1016/j.pulmoe.2020.10.013

Wang, L., Liu, T., Yue, H., Zhang, J., Sheng, Q., Wu, L., et al. (2023). Clinical characteristics and high risk factors of patients with omicron variant strain infection in Hebei, China. *Front. Cell. Infect. Microbiol.* 13:1294904. doi: 10.3389/fcimb.2023.1294904

Wasilewski, P. G., Mruk, B., Mazur, S., Półtorak-Szymczak, G., Sklinda, K., and Walecki, J. (2020). COVID-19 severity scoring systems in radiological imaging - a review. *Pol. J. Radiol.* 85, e361–e368. doi: 10.5114/pjr.2020.98009

Xu, L., Raitoharju, J., Iosifidis, A., and Gabbouj, M. (2022). Saliency-based multilabel linear discriminant analysis. *IEEE Trans. Cybern.* 52, 10200–10213. doi: 10.1109/TCYB.2021.3069338

Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., et al. (2020). Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* 8, 420–422. doi: 10.1016/S2213-2600(20)30076-X

Zhang, L., Chen, H., Zhang, L., and Chen, X. (2021). Analysis of the risk factors for progression from mild to severe cases of COVID-19. *Prev. Med.* 4, 751–764.

Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., et al. (2020). Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 181, 1423–1433.e11. doi: 10.1016/j.cell.2020.04.045

Zhao, Q., Meng, M., Kumar, R., Wu, Y., Huang, J., Deng, Y., et al. (2020). Lymphopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a systemic review and meta-analysis. *Int. J. Infect. Dis.* 96, 131–135. doi: 10.1016/j.ijid.2020.04.086

Zhou, W., Liu, Y., Xu, B., Wang, S., Li, S., Liu, H., et al. (2021). Early identification of patients with severe COVID-19 at increased risk of in-hospital death: a multicenter case-control study in Wuhan. *J. Thorac. Dis.* 13, 1380–1395. doi: 10.21037/jtd-20-2568

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 November 2024
DOI 10.3389/fmicb.2024.1495709

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Xiaoran Li,
Capital Medical University, China
Huanming Xu,
Northern Theater Command General
Hospital, China
ningning Gao,
Liaoning Cancer Hospital and Institute, China

*CORRESPONDENCE
Jiuhui Su
✉ sjhmedicine@126.com

RECEIVED 13 September 2024
ACCEPTED 16 October 2024
PUBLISHED 06 November 2024

CITATION
Xia Y, Kang Q, Gao Y and Su J (2024) A transformer-based deep learning model for identifying the occurrence of acute hematogenous osteomyelitis and predicting blood culture results.
Front. Microbiol. 15:1495709.
doi: 10.3389/fmicb.2024.1495709

A transformer-based deep learning model for identifying the occurrence of acute hematogenous osteomyelitis and predicting blood culture results

Yingtu Xia[1], Qiang Kang[2], Yi Gao[3] and Jiuhui Su[4]*

[1]Department of Orthopedics, The Second Hospital of Dalian Medical University, Dalian, Liaoning, China, [2]Department of Radiology, Xing'an League People's Hospital of Inner Mongolia, Ulanhot, Inner Mongolia, China, [3]Department of Laboratory Medicine, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [4]Department of Orthopaedics, Haicheng Bonesetting Hospital, Haicheng, Liaoning, China

Background: Acute hematogenous osteomyelitis is the most common form of osteomyelitis in children. In recent years, the incidence of osteomyelitis has been steadily increasing. For pediatric patients, clearly describing their symptoms can be quite challenging, which often necessitates the use of complex diagnostic methods, such as radiology. For those who have been diagnosed, the ability to culture the pathogenic bacteria significantly affects their treatment plan.

Method: A total of 634 patients under the age of 18 were included, and the correlation between laboratory indicators and osteomyelitis, as well as several diagnoses often confused with osteomyelitis, was analyzed. Based on this, a Transformer-based deep learning model was developed to identify osteomyelitis patients. Subsequently, the correlation between laboratory indicators and the length of hospital stay for osteomyelitis patients was examined. Finally, the correlation between the successful cultivation of pathogenic bacteria and laboratory indicators in osteomyelitis patients was analyzed, and a deep learning model was established for prediction.

Result: The laboratory indicators of patients are correlated with the presence of acute hematogenous osteomyelitis, and the deep learning model developed based on this correlation can effectively identify patients with acute hematogenous osteomyelitis. The laboratory indicators of patients with acute hematogenous osteomyelitis can partially reflect their length of hospital stay. Although most laboratory indicators lack a direct correlation with the ability to culture pathogenic bacteria in patients with acute hematogenous osteomyelitis, our model can still predict whether the bacteria can be successfully cultured.

Conclusion: Laboratory indicators, as easily accessible medical information, can identify osteomyelitis in pediatric patients. They can also predict whether pathogenic bacteria can be successfully cultured, regardless of whether the patient has received antibiotics beforehand. This not only simplifies the diagnostic process for pediatricians but also provides a basis for deciding whether to use empirical antibiotic therapy or discontinue treatment for blood cultures.

KEYWORDS

deep learning, osteomyelitis, blood culture, anti-infection treatment, pathogenic microorganism

1 Introduction

Osteomyelitis is one of the common bone and muscle infections in children. Acute hematogenous osteomyelitis (AHO) is the most common form of this disease in pediatrics. In recent years, the incidence of osteomyelitis has been increasing annually (Walter et al., 2021). The incidence is generally higher in males compared to females, and lower limb infections are more prevalent than upper limb infections (Kremers et al., 2015; Disch et al., 2023). Typically, healthy bones have strong resistance to pathogen invasion. The occurrence of osteomyelitis mainly occurs through three mechanisms: direct inoculation, extension from adjacent lesions, and hematogenous dissemination. Additionally, in conditions of bone ischemia, trauma, or foreign bodies, pathogens are more likely to adhere to exposed bone locations, leading to bone infection. Most cases of osteomyelitis can be cured; however, a small number of affected children may experience discrepancies in limb length between the affected and unaffected sides (Liu et al., 2024).

The presentation of AHO varies, ranging from localized infections at a single epiphyseal site to multifocal infections accompanied by septic shock (Funk and Copley, 2017). Fever and pain are the most common manifestations of bone infections. Common signs of osteomyelitis include fever, pain, swelling, erythema, localized warmth, and varying degrees of functional impairment. The onset of symptoms can differ depending on the type of pathogen involved (Calvo et al., 2016). When the lower limb bones are affected, children often have difficulty bearing weight or may exhibit noticeable limping, whereas involvement of the pelvis may lead to a waddling gait. Overall, the functional impairment and the location of the infection are highly correlated (Dich et al., 1975). In fact, similar symptoms can be observed in various pediatric orthopedic conditions. For example, osteosarcoma (OSC) and Ewing's sarcoma (EWS) are the most common primary malignant bone tumors in children and young adults, and they also present with significant pain and swelling (Wang et al., 2022). Fractures also present with localized pain, swelling, functional impairment, deformities, and abnormal movements. Considering that children may have difficulty responding accurately to physical examinations, confirming a diagnosis of osteomyelitis through simple procedures is more challenging.

Acute hematogenous osteomyelitis is often caused by pathogen infections, so identifying the type of pathogen early in the disease is crucial for treatment (Jahan et al., 2024). This not only directly guides the physician's treatment but also provides psychological comfort to the family. Although *Staphylococcus aureus* is the most common pathogen in osteomyelitis (McNeil, 2020), pathogen culture is the gold standard for pathogen diagnosis, with blood culture being the most common method (Woods et al., 2021). To obtain accurate culture results, samples need to be taken before the use of antibiotics. However, in practice, it is difficult to ensure that patients have not self-medicated with antibiotics before admission. Bone biopsy or aspiration is also limited to the early stages of the disease. Thus, to achieve accurate culture results, discontinuing antibiotics is often necessary (Manz et al., 2018). For critically ill patients, discontinuing antibiotics to obtain accurate bacterial culture results is clearly impractical. Despite the stringent sampling requirements, not all children with acute osteomyelitis can successfully culture the pathogen (Section et al., 2015). Discontinuing antibiotics is often difficult for patients due to the potential risks of disease progression. Therefore, discontinuing medication to diagnose the pathogen type is a significant clinical challenge.

Artificial intelligence has been widely applied in the diagnosis and treatment of osteomyelitis. AI methods not only efficiently handle repetitive tasks and improve diagnostic efficiency but also explore complex relationships between medical information, mapping features to manifestations, and establishing quantitative relationships between medical information and clinical outcomes. A 2022 study classified

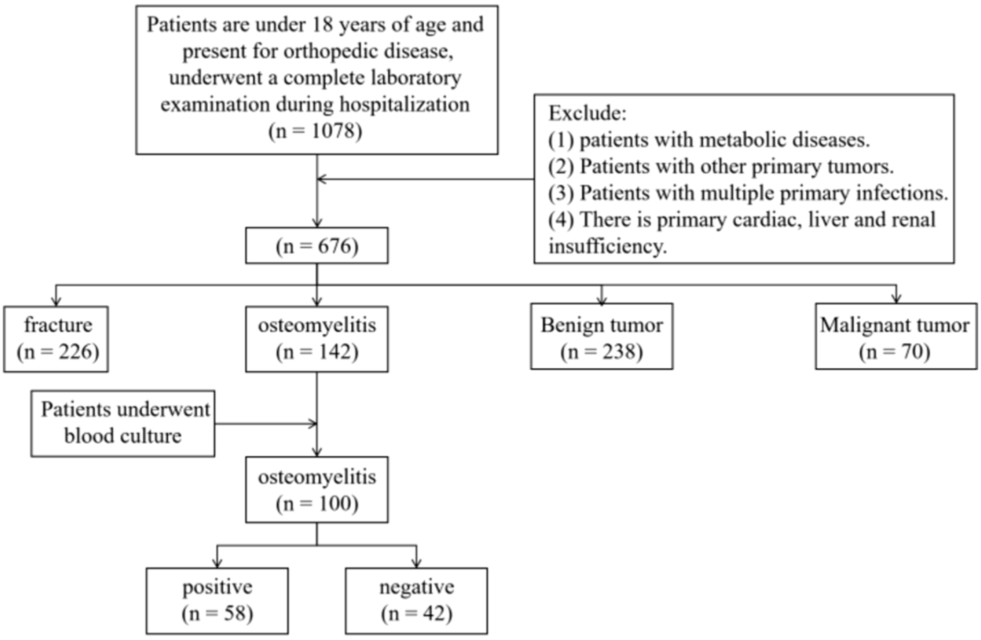


FIGURE 1
Inclusion and exclusion flowchart of patients in this study.

acute osteomyelitis, chronic osteomyelitis, and Ewing's sarcoma using patient X-ray images (Consalvo et al., 2022). Another 2022 study proposed a machine learning model based on clinical features and biomarkers to classify diabetic foot, necrotizing fasciitis, and osteomyelitis, trained and validated on a dataset of 1,581 samples (Kim et al., 2022). A 2024 study included 145 patients with diagnosed spinal infections undergoing metagenomic next-generation sequencing (mNGS) to differentiate pathogen types in iatrogenic vertebral osteomyelitis (IVO) and native vertebral osteomyelitis (NVO) (Gao et al., 2024). These studies demonstrate the potential of AI in the diagnosis and treatment of osteomyelitis.

This study explored the correlation between laboratory parameters and diseases commonly confused with acute hematogenous osteomyelitis. Based on this, we developed an intelligent diagnostic system based on clinical and laboratory features, which can classify patients into categories of acute hematogenous osteomyelitis, benign bone tumors, malignant bone tumors, and fractures. At the same time, we investigated the correlation between the ability to culture pathogenic bacteria and the number of hospital days, and established a deep learning model to predict whether pathogenic bacteria can be cultured. This model supports clinicians in deciding whether to discontinue antibiotics for blood culture purposes.

2 Materials and methods

2.1 Patient

This study is a retrospective analysis that includes patients under 18 years of age who were hospitalized at our institution from January 1, 2016, to June 1, 2024. Baseline characteristics, including age and gender, were collected, as shown in Figure 1.

Inclusion criteria were as follows:

1. Age under 18 years at the time of admission.
2. Diagnosis of one of the following conditions: acute hematogenous osteomyelitis, benign bone tumor, malignant bone tumor, or fracture. AHO was diagnosed based on both laboratory indicators and clinical presentation. Benign and malignant bone tumors were diagnosed through pathology,

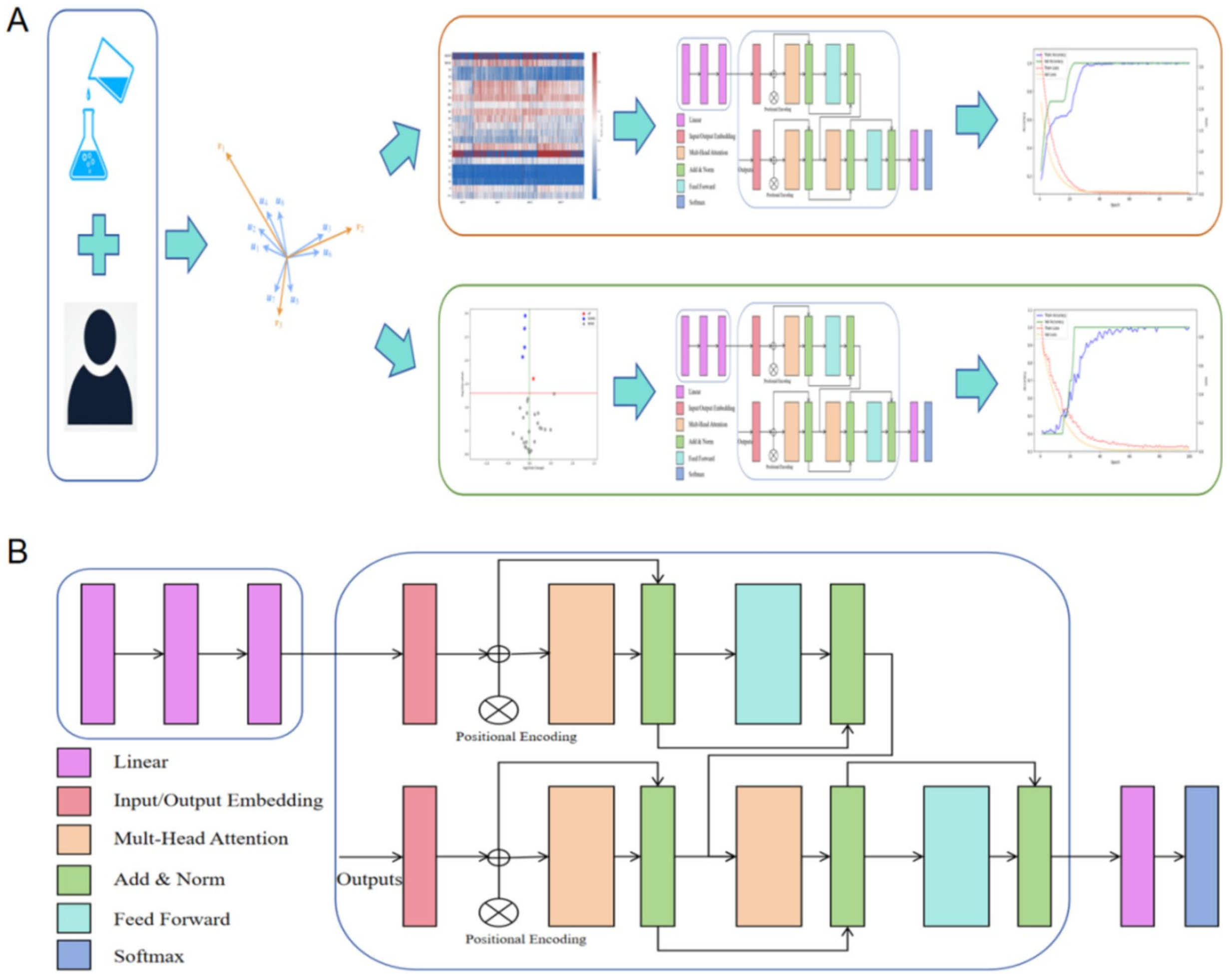


FIGURE 2
Research workflow of this study. (A) We constructed feature vectors using patients' clinical characteristics and laboratory parameters, and utilized the deep learning architecture proposed in this study for classification tasks. (B) The structural diagram of the model in this study. Input was encoded through a standard ANN (with three linear layers and RELU activation function) and decoded through a transformer.

TABLE 1 Baseline information of patients.

	AHO/100	Benign bone tumor/238	Malignant bone tumor/70	Fracture/226
Sex	M67/F33	M146/F92	M31/F39	M153/F73
Gender	9.00 (5.00, 11.25)	11.50 (9.00, 14.00)	13.00 (10.25, 15.00)	9.00 (6.00, 11.00)
Length of hospital stay	19.50 (14.75, 27.00)	6.00 (5.00, 8.00)	15.00 (11.25, 19.00)	4.00 (3.00, 7.00)
EO#	0.15 (0.08, 0.22)	0.11 (0.10, 0.20)	0.10 (0.10, 0.20)	0.07 (0.02, 0.12)
EO%	2.40 (1.58, 3.73)	2.00 (1.30, 3.30)	1.70 (0.90, 2.90)	0.90 (0.20, 1.70)
HCT	33.95 (31.58, 36.95)	39.65 (37.60, 41.80)	34.45142	37.70 (35.25, 39.98)
HGB	112.00 (103.75, 121.00)	133.00 (124.00, 140.75)	115	126.00 (118.00, 133.00)
MCH	27.50 (26.48, 28.73)	28.40 (27.90, 29.50)	29.11285	28.40 (27.40, 29.00)
MCHC	331.18	335.00 (332.00, 340.00)	335.00 (329.00, 342.75)	335.00 (328.00, 339.00)
MCV	82.50 (80.00, 86.00)	85.00 (83.72, 87.83)	87.38571	85.00 (82.00, 87.00)
MPV	7.508	9.60 (8.50, 10.30)	9.26142	8.20 (7.50, 8.50)
PCT	0.2599	0.26394	0.278	0.25 (0.22, 0.27)
PLT	337.00 (285.75, 394.75)	280.50 (240.25, 323.75)	298.00 (260.25, 349.50)	298.00 (262.25, 342.75)
WBC	6.29 (4.98, 7.38)	6.89 (5.70, 8.00)	7.39 (5.93, 8.68)	8.41 (7.39, 11.82)
RBC	4.20 (3.80, 4.50)	4.40 (4.40, 4.71)	4.32 (4.00, 4.40)	4.40 (4.20, 4.80)
RH	1.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 0.00)	1.00 (1.00, 1.00)
A/G	1.30 (1.10, 1.60)	1.70 (1.60, 1.90)	1.586714286	1.68 (1.60, 1.90)
ALT	13.50 (10.00, 25.25)	13.00 (9.00, 18.00)	18.00 (12.00, 30.00)	13.00 (11.00, 15.00)
AST	21.00 (17.00, 25.00)	19.00 (15.00, 23.00)	19.00 (15.00, 26.00)	22.00 (21.00, 26.00)
GGT	17.00 (14.00, 28.00)	14.00 (11.00, 18.00)	21.00 (14.00, 30.25)	14.00 (12.00, 15.00)
TP	69.889	68.42521	64.57142	68.90 (68.22, 73.20)
Urea	4.31 (3.52, 4.77)	4.42 (3.81, 5.36)	4.00 (3.39, 4.59)	4.31 (3.83, 4.82)

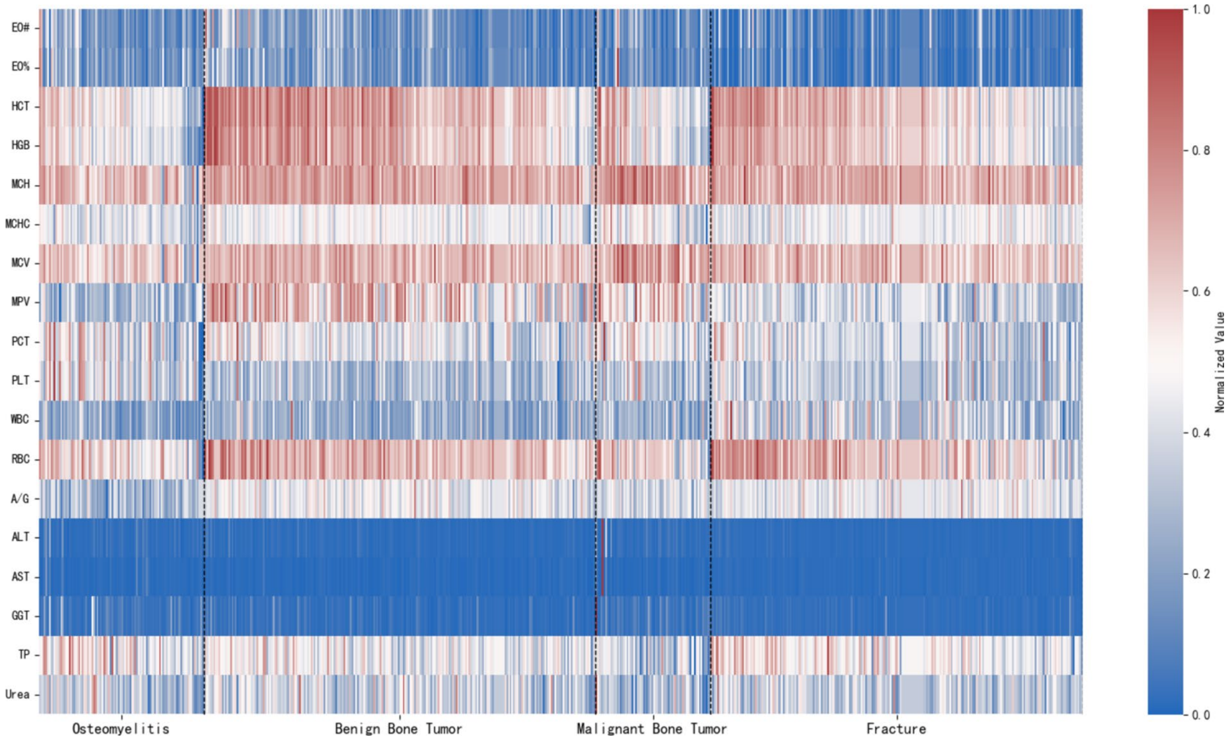


FIGURE 3
Expression of laboratory parameters in different disease types.

TABLE 2 Association between categorical variables and disease classification using cramér's *V*.

Indicator	Cramér's *V*	χ^2	*P*-value
Gender	0.14607	13.528	*0.00362
RH	0.41983	111.745	*0

*Has statistical significance.

while fractures were diagnosed through imaging or clear clinical manifestations (i.e., pain, deformity, and restricted movement).

3 For patients with AHO, at least one blood-based bacterial culture must have been performed during hospitalization.

Exclusion criteria were as follows:

1 Presence of metabolic diseases.
2 Presence of other primary tumors.
3 Presence of multiple primary infections.
4 Presence of primary heart, kidney, or liver dysfunction.

2.2 Laboratory parameters and blood culture

Collect laboratory data for the included population, The parameters related to blood cell counts include white blood cell count (WBC), red blood cell count (RBC), and platelets (PLT), all of which are obtained through the electrical impedance method (Coulter principle). Hemoglobin (HGB) is obtained using the sodium dodecyl sulfate hemoglobin colorimetric method. The white blood cell classification parameters include absolute eosinophil count (EO#, Eosinophil Count) and percentage (EO%, Eosinophil Percentage), absolute basophil count (BA#, Basophil Count) and percentage (BA%, Basophil Percentage), absolute lymphocyte count (LY#, Lymphocyte Count) and percentage (LY%, Lymphocyte Percentage), absolute monocyte count (MO#, Monocyte Count) and percentage (MO%, Monocyte Percentage), and absolute neutrophil count (NE#, Neutrophil Count) and percentage (NE%, Neutrophil Percentage). EO# and EO%, BA# and BA%, LY# and LY%, MO# and MO%, NE# and NE% are all obtained through VCS counting [V represents dual motor direct current characteristics (Coulter principle); C represents radio frequency conduction characteristics; S represents laser scattering]. Biochemical indicators include the albumin/globulin ratio (A/G), alanine aminotransferase (ALT), aspartate aminotransferase (AST), where AST is obtained through the reduced coenzyme method (NADH method). Albumin (AlbG) is obtained using the bromocresol green (BCG) method, and direct bilirubin (BilD) and total bilirubin (BilT) are obtained through the vanadate oxidation method. Gamma-glutamyl transferase (GGT) is obtained using the L-γ-glutamyl-3-carboxy-4-nitroaniline substrate method, and total protein (TP) is obtained using the biuret method. RH represents the Rh blood type (Rhesus Blood Type), obtained through the gel microcolumn method. Hematocrit (HCT), platelet count (PLT), mean platelet volume (MPV), plateletcrit (PCT), platelet distribution width (PDW), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW), and unconjugated bilirubin (UNBIL) are all calculated values.

TABLE 3 Analysis of variations in continuous variables across different disease groups.

Indicator	*H*-value	*P*-value	Indicator	*H*-value	*P*-value
Age	88.93294	*0	PCT	21.60439	*0
EO#	80.67734	*0	PLT	37.97623	*0
EO%	108.21281	*0	WBC	109.35881	*0
HCT	136.72223	*0	RBC	66.77762	*0
HGB	133.80915	*0	A/G	133.28753	*0
MCH	47.72300	*0	ALT	24.30355	*0
MCHC	14.29948	*0.00252	AST	43.32638	*0
MCV	49.52752	*0	GGT	68.77069	*0
MPV	229.19484	*0	TP	50.19360	*0
Urea	16.94605	*0.00072			

*Has statistical significance.

2.3 Blood culture

After the patient visits, blood samples should be collected as soon as possible. Draw 10 mL of blood from each arm of the patient for both aerobic and anaerobic blood cultures. Subsequently, send the blood culture bottles to the laboratory for testing. The detection involves monitoring whether bacteria consume nutrients in the culture bottles or produce new metabolites, triggering an alarm. Generally, if no alarm is triggered by the instrument within 5 days, the result is considered negative. If an alarm is triggered by the culture bottle, it is necessary to perform Gram staining microscopy and plate inoculation with the mixed solution from the culture bottle. For positive alarms in aerobic blood culture bottles, blood agar and MacConkey agar plates are commonly used. For positive alarms in anaerobic blood culture bottles, ensure anaerobic procedures and perform incubation in an anaerobic or microaerophilic environment.

2.4 Deep learning model

We developed a deep learning model structure to process laboratory parameters. The model consists of an ANN and a transformer. The transformer is a sequence-to-sequence model based on the attention mechanism, with the core idea of using self-attention to capture contextual relationships at different positions in the input sequence (Vaswani et al., 2017). We constructed feature vectors from clinical and laboratory data. The input is encoded through an ANN (including three linear layers with RELU activation functions) and then decoded by the transformer to classify the patients. For the two tasks in this study—differentiating AHO and identifying infection bacteria through blood cultures—we used the same model architecture, only changing the final fully connected layer to suit different tasks. The research workflow of this study is shown in Figure 2. The deep learning model in this study was trained on a workstation equipped with an NVIDIA RTX 4090 GPU (24GB VRAM) and 64GB of system memory. The training was based on PyTorch 2.4.1 and utilized CUDA 11.8 for acceleration.

TABLE 4 Baseline characteristics of the training, validation, and test sets for multicategory deep learning of laboratory parameters in AHO, benign bone tumors, malignant bone tumors, and fractures.

	Train/380				Validation/126				Test/128			
	AHO/70	Benign bone tumor/142	Malignant bone tumor/37	Fracture/131	AHO/18	Benign bone tumor/40	Malignant bone tumor/16	Fracture/52	AHO/12	Benign bone tumor/56	Malignant bone tumor/17	Fracture/43
Sex	M47/F23	M88/F54	M17/F20	M85/F46	M12/F6	M28/F12	M5/F11	M32/F20	M8/F4	M30/F26	M9/F8	M35/F8
Gender	9.00 (5.00, 11.00)	11.00 (8.00, 13.00)	12.64865	9.00 (6.00, 11.00)	8.50000	12.05000	14.00 (11.00, 14.25)	8.94118	7.66667	12.00 (9.50, 14.00)	11.82353	8.72093
EO#	0.15 (0.08, 0.22)	0.12 (0.10, 0.21)	0.10 (0.10, 0.20)	0.07 (0.02, 0.11)	0.11 (0.08, 0.14)	0.10 (0.10, 0.20)	0.10 (0.10, 0.23)	0.06 (0.01, 0.13)	0.22250	0.14 (0.10, 0.20)	0.10 (0.02, 0.19)	0.10 (0.03, 0.11)
EO%	2.60 (1.60, 3.80)	2.05 (1.20, 3.48)	2.00 (0.90, 3.20)	0.90 (0.20, 1.70)	2.05556	1.80 (1.50, 2.77)	2.00 (1.08, 3.10)	0.80 (0.15, 1.70)	3.65833	2.00 (1.60, 3.30)	1.41765	1.00 (0.30, 1.75)
HCT	34.40 (32.10, 37.20)	39.43873	35.23514	37.70 (35.45, 40.05)	34.06667	40.05 (37.70, 43.52)	32.57500	37.70 (34.80, 39.90)	31.28333	39.09455	31.70 (30.70, 35.80)	37.16977
HGB	113.00 (106.00, 122.00)	131.99296	117.94595	126.00 (119.00, 133.50)	113.44444	135.20000	109.62500	126.00 (115.50, 132.00)	103.75000	130.38182	113.64706	124.74419
MCH	27.30 (26.50, 28.60)	28.50 (27.90, 29.60)	29.07297	28.40 (27.40, 29.30)	27.80556	28.40 (27.93, 28.73)	28.98750	27.91961	26.74167	28.39818	29.31765	28.61163
MCHC	331.08696	335.00 (332.00, 339.75)	335.27027	335.00 (329.00, 339.00)	332.50000	336.75000	339.43750	333.58824	330.75000	334.92727	330.94118	335.69767
MCV	82.00 (80.00, 85.00)	85.51479	86.95946	85.00 (82.00, 87.00)	83.61111	85.00 (83.88, 86.27)	86.42500	83.70588	80.83333	85.18545	89.21765	85.30233
MPV	7.48116	9.60 (8.60, 10.30)	9.13514	8.20 (7.60, 8.50)	7.62222	9.54500	9.58750	8.01569	7.47500	9.31455	9.22941	8.50 (7.50, 8.60)
PCT	0.26536	0.26711	0.28 (0.25, 0.30)	0.25 (0.21, 0.27)	0.22778	0.25525	0.28125	0.26 (0.21, 0.28)	0.27750	0.26309	0.27294	0.26 (0.23, 0.28)
PLT	345.00 (292.00, 401.00)	279.00 (242.25, 331.25)	298.00 (267.00, 339.00)	298.00 (251.50, 337.50)	285.00 (254.25, 325.50)	274.00 (232.75, 298.00)	312.00000	316.54902	375.25000	296.00 (256.00, 324.50)	304.41176	303.00 (278.00, 342.00)
WBC	6.29 (5.04, 7.75)	6.90 (5.83, 8.54)	7.11541	8.40 (7.39, 11.75)	6.09000	6.48675	7.23875	9.97098	6.92917	7.03382	8.65471	8.27 (7.39, 9.96)
RBC	4.20 (3.90, 4.60)	4.42 (4.40, 4.71)	4.39 (4.00, 4.44)	4.40 (4.20, 4.80)	4.07778	4.40 (4.40, 4.65)	4.20 (4.04, 4.40)	4.40 (4.20, 4.80)	3.88333	4.48 (4.40, 4.80)	4.14294	4.40 (4.15, 4.75)
RH	1.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 0.00)	1.00 (1.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 0.00)	1.00 (1.00, 1.00)	0.50 (0.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 0.00)	1.00 (0.00, 1.00)
A/G	1.30 (1.10, 1.60)	1.75690	1.57541	1.68 (1.60, 1.80)	1.45 (1.20, 1.66)	1.74300	1.57813	1.68 (1.50, 1.90)	1.38833	1.70 (1.60, 1.86)	1.61941	1.68 (1.60, 1.90)
ALT	14.00 (10.00, 25.00)	13.00 (9.00, 18.00)	21.00 (12.00, 30.00)	13.00 (11.50, 15.00)	19.50 (13.00, 37.00)	12.50 (9.75, 18.25)	18.00 (9.75, 34.50)	13.00 (10.00, 15.00)	13.25000	13.00 (9.00, 16.00)	14.00 (10.00, 24.00)	13.00 (11.50, 14.00)

(Continued)

TABLE 4 (Continued)

	Train/380				Validation/126				Test/128			
	AHO/70	Benign bone tumor/142	Malignant bone tumor/37	Fracture/131	AHO/18	Benign bone tumor/40	Malignant bone tumor/16	Fracture/52	AHO/12	Benign bone tumor/56	Malignant bone tumor/17	Fracture/43
AST	21.00 (17.00, 25.00)	20.00 (16.00, 24.00)	19.00 (17.00, 27.00)	22.00 (21.00, 26.00)	21.50 (19.25, 30.75)	19.00 (14.75, 22.25)	18.00 (14.75, 33.75)	21.00 (18.00, 26.50)	20.58333	19.00 (15.50, 21.00)	20.00 (14.00, 23.00)	22.00 (21.00, 26.00)
GGT	16.00 (14.00, 24.00)	13.00 (10.00, 18.00)	21.00 (14.00, 31.00)	14.00 (11.00, 15.00)	30.00 (15.00, 38.50)	14.00 (11.00, 20.25)	16.00 (12.75, 26.50)	14.00 (13.00, 17.00)	16.50 (13.25, 46.75)	14.00 (12.00, 18.00)	26.35294	13.00 (11.00, 14.00)
TP	69.50725	68.25634	64.69189	68.90 (67.60, 72.20)	70.67778	69.38500	65.40625	71.28824	70.50833	68.40000	63.52353	68.90 (67.55, 71.30)
Urea	4.24899	4.49 (3.88, 5.28)	3.61784	4.31 (3.79, 4.80)	4.21278	4.29 (3.76, 5.62)	4.06 (3.53, 4.68)	4.31 (4.01, 5.03)	4.90750	4.55418	4.41529	4.22070

2.5 Statistical analysis

Statistical analysis was performed using IBM SPSS software version 26.0. First, the Kolmogorov–Smirnov test was conducted to assess the normality of all data. For normally distributed data, the mean ± standard deviation was used for description, while for non-normally distributed data, the median (interquartile range) was used. Differences in continuous variables among multiple groups were analyzed using the Kruskal–Wallis *H* test, while categorical variables were analyzed using the chi-square test, with Cramér's *V* used to quantify the strength of association. In the analysis between two groups, after testing for normality, the Mann–Whitney *U* test was applied to non-normally distributed data, and the *t*-test was used for normally distributed data. Differences in categorical variables were analyzed using the chi-square test, and Cramér's *V* was used to quantify the association between categorical variables. Correlations between continuous variables were assessed using Pearson's test for normally distributed data and Spearman's test for non-normally distributed data. A *p*-value of <0.05 was considered statistically significant. Statistical plots were generated using Python's matplotlib.

3 Results

3.1 Baseline

A total of 634 patients were included in this study, with variables having missing values greater than 20% being removed. Table 1 shows the clinical information and laboratory parameters of the included patients.

3.2 Multicategory correlation analysis of laboratory parameters for AHO, benign bone tumors, malignant bone tumors, and fractures

The expression of laboratory parameters in patients with different disease types is shown in Figure 3.

Correlation test results for disease types and clinical characteristics, as well as laboratory parameters, are presented in Tables 2, 3. We found that, among clinical characteristics, both gender and age were related to orthopedic disease types. In laboratory parameters, all parameters included in this study were related to orthopedic disease types.

3.3 Multicategory deep learning of laboratory parameters for AHO, benign bone tumors, malignant bone tumors, and fractures

By removing columns with more than 20% missing values and imputing the missing values, a total of 21 common variables across the four diseases were used in modeling. The training, validation, and test sets were split in a 6:2:2 ratio, with baseline characteristics shown in Table 4.

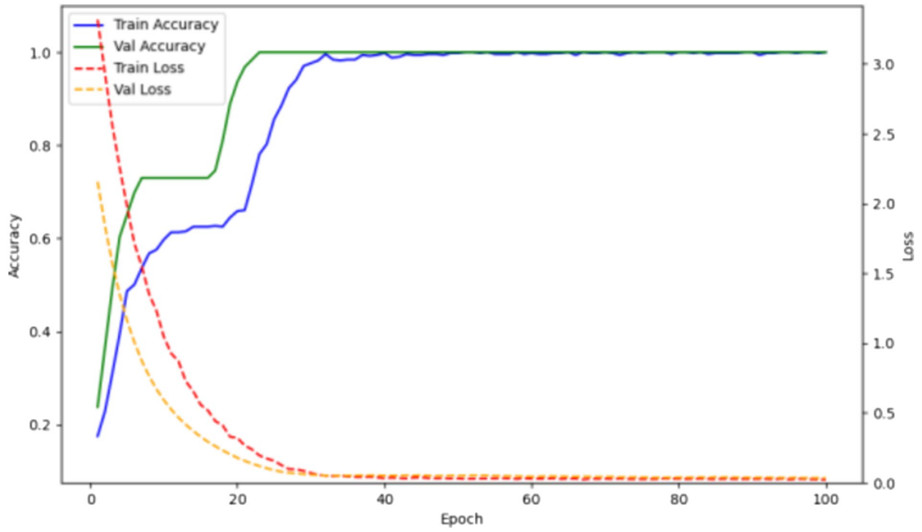


FIGURE 4
Accuracy and loss curves of the multicategory deep learning model for laboratory parameters in AHO, benign bone tumors, malignant bone tumors, and fractures on the training and validation sets.

TABLE 5 Correlation between continuous variables and hospitalization duration.

Indicator	Correlation	*P*-value	Indicator	Correlation	*P*-value
Age	0.00963	0.92427	LY#	−0.02446	0.80911
EO#	0.04269	0.67326	LY%	0.05748	0.56999
EO%	0.09702	0.33692	MO#	0.08611	0.39429
HCT	−0.33353	*0.00070	MO%	0.07880	0.43583
HGB	−0.35001	*0.00036	NE#	−0.07093	0.48315
MCH	−0.25294	*0.01112	NE%	−0.09425	0.35096
MCHC	−0.15533	0.12281	RDW	0.37916	*0.00010
MCV	−0.26286	*0.00824	A/G	−0.03404	0.73673
MPV	0.03388	0.73787	ALT	0.02556	0.80071
PCT	−0.05340	0.59771	AST	−0.15158	0.13222
PLT	0.08313	0.41091	AlbG	−0.06810	0.50082
WBC	−0.05722	0.57176	BilD	0.01904	0.85087
RBC	−0.22104	*0.02710	BilT	−0.05883	0.56096
PDW	0.04390	0.66454	GGT	−0.07641	0.44987
BA#	0.08405	0.40572	TP	−0.08939	0.37648
BA%	0.06003	0.55300	UNBIL	−0.12966	0.19854

*Has statistical significance.

TABLE 6 Analysis of variations in hospitalization duration across different categorical variables.

Indicator	*U*-value	*P*-value
Bacterial culture	1636.0	*0.00362
Gender	1272.0	0.22296
RH	943.0	0.07069

*Has statistical significance.

The training results are shown in Figure 4. Over 32 epochs, the model achieved optimal performance on the validation set and attained an accuracy of 1.0 on the test set.

3.4 Correlation between hospitalization duration for AHO and laboratory parameters

As shown in Tables 5, 6, although AHO is more common in males, there is no correlation between the patient's gender or age and their hospitalization duration. Among the laboratory parameters, RDW is positively correlated with the length of hospital stay, while HCT, HGB, MCH, MCV, and RBC are negatively correlated with the length of hospital stay, with a *p*-value of <0.05. This indicates a strong link between the patient's hospitalization duration and their red blood cell physiological state. Although inflammation markers can assess disease

TABLE 7 Baseline information for AHO bacterial culture.

	Negative/42	Positive/58
Gender	M26/F16	M41/F17
Age	8.00 (4.00, 11.00)	10.00 (6.00, 12.00)
Length of hospital stay	16.00 (12.00, 23.00)	21.00 (16.00, 28.00)
EO#	0.14 (0.08, 0.27)	0.15 (0.09, 0.21)
EO%	2.40 (1.55, 4.00)	2.45 (1.60, 3.55)
HCT	35.25476	32.65172
HGB	119.00 (110.00, 125.00)	107.91379
MCH	27.65714	27.50 (26.10, 28.75)
MCHC	332.40476	330.29310
MCV	83.21429	82.00 (78.50, 85.00)
MPV	7.52857	7.49310
PCT	0.25905	0.26052
PLT	333.50 (288.25, 376.00)	352.81034
WBC	5.99 (4.90, 7.69)	6.36 (5.03, 7.30)
RBC	4.21429	4.10 (3.80, 4.45)
PDW	16.20952	16.20 (15.93, 16.40)
BA#	0.03310	0.03 (0.02, 0.04)
BA%	0.60 (0.40, 0.70)	0.50 (0.40, 0.70)
LY#	2.35 (1.80, 3.05)	2.20 (1.80, 2.90)
LY%	41.83571	37.92931
MO#	0.40 (0.33, 0.60)	0.40 (0.40, 0.60)
MO%	7.60000	7.35 (5.62, 8.10)
NE#	2.70 (2.02, 3.75)	3.10 (2.02, 4.20)
NE%	46.96429	51.12759
RDW	14.44524	14.60 (14.03, 16.10)
RH	1.00 (0.00, 1.00)	1.00 (0.00, 1.00)
A/G	1.30 (1.20, 1.40)	1.22759
ALT	16.00 (10.00, 20.00)	16.00 (11.25, 30.75)
AST	20.00 (18.25, 25.75)	20.00 (16.00, 23.00)
AlbG	39.55 (38.50, 41.30)	38.40 (36.05, 39.27)
BilD	1.60 (1.30, 2.27)	1.70 (1.30, 2.85)
BilT	4.75 (3.17, 6.35)	4.80 (3.82, 6.80)
GGT	18.00 (13.25, 23.00)	19.50 (16.25, 30.00)
TP	70.06905	69.91552
UNBIL	3.00 (2.02, 4.28)	3.00 (2.42, 4.07)

TABLE 8 Association between categorical variables and bacterial culture negative/positive groups using cramér's *V*.

Indicator	Cramér's *V*	χ^2	*P*-value
Gender	0.07067	0.49937	0.47978
RH	0.05376	0.28907	0.59082

TABLE 9 Analysis of variations in continuous variables between bacterial culture negative and positive groups.

Indicator	*U*-statistic	*P*-value
Age	1,416	0.16587
EO#	1,187	0.83119
EO%	1137.5	0.57622
HGB	751	*0.00111
MCH	1,073	0.31273
MCV	1058.5	0.26504
PLT	1,257	0.78801
WBC	1238.5	0.88891
RBC	982.5	0.10002
PDW	1199.5	0.89968
BA#	1,251	0.81698
BA%	1165.5	0.71384
LY%	1,012	0.15122
MO#	1193.5	0.86423
MO%	1,124	0.51365
NE%	1433.5	0.13321
RDW	1,431	0.13761
A/G	896.5	*0.02347
ALT	1381.5	0.25372
AST	1,071	0.30436
AlbG	762.5	*0.00146
BilD	1444.5	0.11271
BilT	1,362	0.31502
GGT	1516.5	*0.03702
UNBIL	1284.5	0.64422

Indicator	*t*-statistic	*P*-value
HCT	3.15776	*0.00211
MCHC	0.97213	0.33338
MPV	0.21465	0.83049
PCT	−0.10334	0.91791
LY%	1.39233	0.16697
NE%	−1.47695	0.14289
TP	0.11779	0.90647

*Has statistical significance.

severity, they are not directly related to the patient's hospitalization duration. Liver function also shows no correlation with the hospitalization duration for AHO. The correlation between bacterial culture results and clinical information or laboratory parameters in AHO patients reveals statistical differences in HCT, HGB, RDW, A/G, and AlbG. Additionally, there is a significant difference in the length of hospital stay between the negative and positive culture sample groups. As shown in Table 7, patients with positive cultures have a longer hospital stay.

3.5 Correlation analysis of bacterial culture results in AHO

Baseline information for AHO patients is shown in Table 7. Additionally, among the patients with positive cultures, the identified strains included 51 *Staphylococcus aureus*, 1 *Acinetobacter baumannii*,

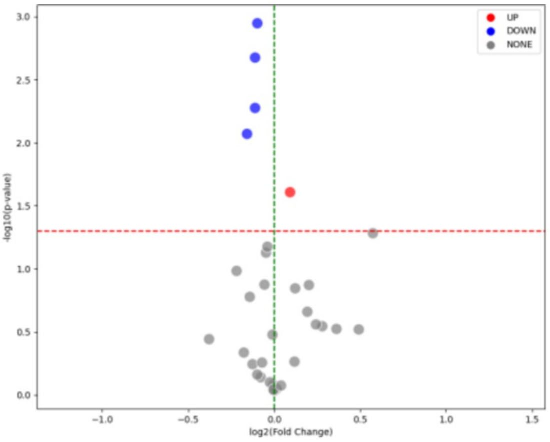


FIGURE 5
Upregulation and downregulation of parameters in the groups with or without detection of pathogenic bacteria in AHO patients: parameters were adjusted as follows: RDW was upregulated, while A/G, HGB, HCT, and AlbG were downregulated.

1 *Staphylococcus epidermidis*, 1 *Bacillus subtilis*, 1 *Achromobacter xylosoxidans*, 1 Klebsiella, 1 Malassezia furfur, and 1 *Pseudomonas aeruginosa*.

As shown in Tables 8, 9, there are no significant differences in gender and RH between the negative and positive culture sample groups. Only HGB, A/G, AlbG, GGT, and HCT showed significant differences between the negative and positive culture sample groups. According to Table 7, the positive culture group had lower levels of HGB, A/G, AlbG, and HCT, and higher levels of GGT. The upregulation and downregulation situations in the two groups are illustrated in Figure 5.

3.6 Deep learning model for detecting pathogenic bacteria in AHO

By removing columns with more than 20% missing values and filling in the missing values, 34 common variables from both bacterial culture negative and positive AHO patients were used for modeling. The baseline features are shown in Table 10.

The training/testing ratio was 6:2:2. The training results are shown in Figure 6. Over 58 epochs, the model achieved optimal performance on the validation set and an accuracy of 1.0 on the test set.

4 Discussion

Acute hematogenous osteomyelitis is a challenging diagnosis in pediatric emergency departments. The condition can develop gradually over a few days but typically manifests within 2 weeks. Patients may present with localized symptoms such as redness, swelling, and fever at the infection site. They might experience dull pain, with or without movement, and sometimes systemic symptoms such as fever or chills. In subacute cases, some patients may exhibit generalized discomfort, mild pain over several weeks accompanied by slight fever or other systemic symptoms (Schmitt, 2017). The variety of symptoms and the difficulty children have in clearly describing their condition make the diagnosis and treatment of osteomyelitis quite challenging (Stephan et al., 2022). In this study, we propose a diagnostic model consisting of two deep learning models that can accurately diagnose AHO in children and predict whether the pathogenic bacteria can be identified through blood cultures. This study not only provides clinicians with a straightforward method for confirming AHO but also offers support for decisions regarding the necessity of stopping antibiotics for bacterial culture.

There are several methods for diagnosing AHO, with blood cultures and X-rays currently being strongly recommended. However, both methods have their drawbacks. Blood cultures require a lengthy period to yield results and cannot always ensure accuracy (Doern et al., 2019). X-rays, while simple, quick, and safe, have low sensitivity (Zaki and Morrison, 2024). Additionally, MRI often requires sedation for pediatric patients, and collecting samples from the affected area for bacterial culture through invasive methods faces the same issues as blood culture (Dartnell et al., 2012; Dong et al., 2019). In addition to these two types of examination methods, recent research by Paliwal et al. (2021) has revealed the diagnostic capability of ultrasound for acute hematogenous osteomyelitis (AHO). By assessing the accumulation of deep soft tissue fluid around the bones in AHO cases, rapid diagnosis can be achieved. This method is also expected to advance the diagnosis of AHO (Paliwal et al., 2021).

Additionally, Stephan et al. (2022) identified elevated C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) as the most sensitive laboratory markers in pediatric emergency settings. In the study by Manz et al. (2020), elevated CRP was also found to be associated with poor long-term outcomes in AHO. Based on previous studies, Stephan et al. (2024) also demonstrated in subsequent experiments that elevated CRP and ESR are closely related to poor long-term prognosis in AHO. In this study, we propose a method for diagnosing AHO through laboratory parameters. By employing a multiclass model structure, it can effectively diagnose several common orthopedic conditions, significantly enhancing diagnostic efficiency. Considering that the model in this study needs to differentiate between osteomyelitis and other orthopedic diseases, and that C-reactive protein and erythrocyte sedimentation rate are not essential laboratory parameters for orthopedic diseases, they were not included in this study.

Different types of pathogenic bacteria may lead to variations in hospital length of stay. A 2023 study indicated that there is a difference in hospital stay duration for osteomyelitis caused by methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-sensitive *Staphylococcus aureus* (MSSA) (Wen et al., 2023). Additionally, children with acute osteomyelitis who are Black, Hispanic, or of other races and ethnicities have longer hospital stays compared to White children (Campbell et al., 2023). In our study, we found that the infection level and liver function levels of patients were not associated with the length of hospital stay. Considering that in China, the length of hospital stay is highly correlated with the patient's recovery of physical health and is less influenced by the wishes of the patients and their families, we believe that the status of the red blood cells is an important factor affecting the recovery process of AHO. This study found a positive correlation between RDW and length of hospital stay, while HCT, HGB, MCH, MCV, and RBC showed a negative correlation with length of hospital stay. This may be due to the important role of red blood cells

TABLE 10 Baseline features of the deep learning model for pathogen detection in training, validation, and test sets.

	Train		Validation		Test	
	Negative	Positive	Negative	Positive	Negative	Positive
Sex	M15/F13	M 25/F7	M 3/F3	M 8/F6	M 8/F1	M 8/F3
Gender	6.51852	10.00 (6.75, 11.25)	11.40000	7.85714	9.75000	11.00 (8.50, 13.00)
EO#	0.15 (0.08, 0.27)	0.14 (0.09, 0.20)	0.07200	0.13 (0.09, 0.17)	0.26375	0.18636
EO%	2.60 (1.15, 4.00)	2.55 (1.68, 3.08)	2.10 (1.10, 2.10)	1.60 (1.10, 4.28)	4.12500	3.23636
HCT	35.19259	32.66875	34.18000	34.32857	36.16250	29.70909
HGB	116.88889	108.46875	106.80000	112.92857	122.12500	104.00 (85.50, 108.50)
MCH	27.67778	27.65 (26.60, 29.02)	27.14000	27.28571	28.12500	24.42727
MCHC	332.33333	331.90625	328.20000	328.64286	337.62500	327.45455
MCV	83.33333	82.00 (80.75, 85.25)	82.60000	83.00000	83.37500	74.63636
MPV	7.57407	7.39688	7.62000	7.34286	7.33750	7.85455
PCT	0.24 (0.23, 0.29)	0.26094	0.22600	0.24929	0.27375	0.27727
PLT	331.00 (289.50, 365.50)	336.00 (289.75, 407.75)	301.40000	342.35714	379.37500	359.90909
WBC	6.78630	6.35 (4.83, 6.96)	4.78000	7.03786	6.29625	6.27636
RBC	4.22593	3.99688	3.94000	4.13571	4.35000	4.04545
PDW	16.26667	16.24063	16.10000	15.98571	16.07500	16.34545
BA#	0.03556	0.03 (0.02, 0.04)	0.02800	0.04071	0.03 (0.02, 0.03)	0.03091
BA%	0.60 (0.40, 0.70)	0.50 (0.40, 0.72)	0.64000	0.59286	0.55000	0.49091
LY#	2.70 (1.90, 3.30)	2.24375	1.56000	2.56429	2.11250	2.30 (1.80, 2.60)
LY%	45.11481	37.85000	32.88000	38.67857	34.76250	39.34545
MO#	0.40 (0.35, 0.60)	0.48125	0.46000	0.40 (0.40, 0.47)	0.45000	0.43636
MO%	7.31481	7.75 (6.57, 8.30)	9.00000	6.65 (5.38, 7.95)	7.12500	6.87273
NE#	2.70 (2.00, 3.40)	3.05 (2.18, 3.88)	2.86000	3.76429	3.43750	3.18182
NE%	43.90741	51.00313	56.04000	50.32857	53.46250	50.05455
RDW	14.57037	14.60 (14.10, 15.72)	15.28000	14.45000	13.65000	17.81818
RH	1.00 (0.00, 1.00)	0.50 (0.00, 1.00)	1.00 (1.00, 1.00)	1.00 (0.00, 1.00)	1.00 (0.75, 1.00)	1.00 (0.50, 1.00)
A/G	1.30 (1.20, 1.45)	1.18125	1.48000	1.31429	1.27500	1.22727
ALT	15.00 (9.00, 20.00)	16.00 (11.00, 28.25)	23.80000	17.00 (12.25, 40.25)	14.50000	16.00 (13.50, 23.50)
AST	23.00 (19.50, 26.00)	19.00 (16.00, 22.25)	19.40000	20.50 (19.00, 26.75)	18.12500	20.00 (17.00, 21.50)
AlbG	39.30 (38.45, 41.20)	36.88438	41.20000	37.76429	39.95000	38.60 (37.35, 40.55)
BilD	1.70 (1.30, 2.45)	1.70 (1.30, 2.70)	1.40 (1.30, 2.30)	1.60 (1.30, 2.55)	1.40000	1.70 (1.65, 2.40)
BilT	5.50000	4.80 (3.10, 6.42)	3.70 (3.40, 5.40)	5.19286	3.98750	4.80 (4.60, 6.95)
GGT	15.00 (12.50, 20.00)	21.00 (16.75, 30.00)	26.40000	29.00 (17.50, 50.25)	21.00000	19.18182
TP	69.84444	70.17500	69.44000	67.52143	71.50000	71.26364
UNBIL	3.52222	3.00 (1.95, 4.17)	2.60 (2.30, 3.10)	3.00 (2.60, 3.33)	2.57500	3.20 (2.75, 4.05)

in oxygen transport and bone tissue repair. If there is an insufficient number of red blood cells (as seen in anemia) or abnormal red blood cell function, the bone tissue may not receive enough oxygen, which can delay repair.

The Pediatric Infectious Diseases Society and the Infectious Diseases Society of America recommend performing blood cultures before the use of antibiotics (Woods et al., 2021). Despite the presence of clear symptoms, the detection rate of blood cultures remains unsatisfactory, and this rate rarely exceeds 60% in various cohorts (Russell et al., 2015; McNeil et al., 2019). Although sampling from adjacent infected sites can somewhat improve the detection rate, it does not provide an overwhelming advantage over blood cultures (Athey et al., 2019). In practical medical work, patients often have already been on oral antibiotics for some time by the time they arrive at the hospital. Once effective antibiotics are started, the detection rate of blood cultures usually drops rapidly within a few hours of exposure. Therefore, deciding whether to stop antibiotics for culture testing is a challenging decision for physicians. Previous studies have analyzed different clinical features as independent predictors of positive blood cultures (Burns et al., 2023). In this study, our model accurately predicts whether pathogens can be identified through blood cultures. This provides guidance for clinicians on whether to

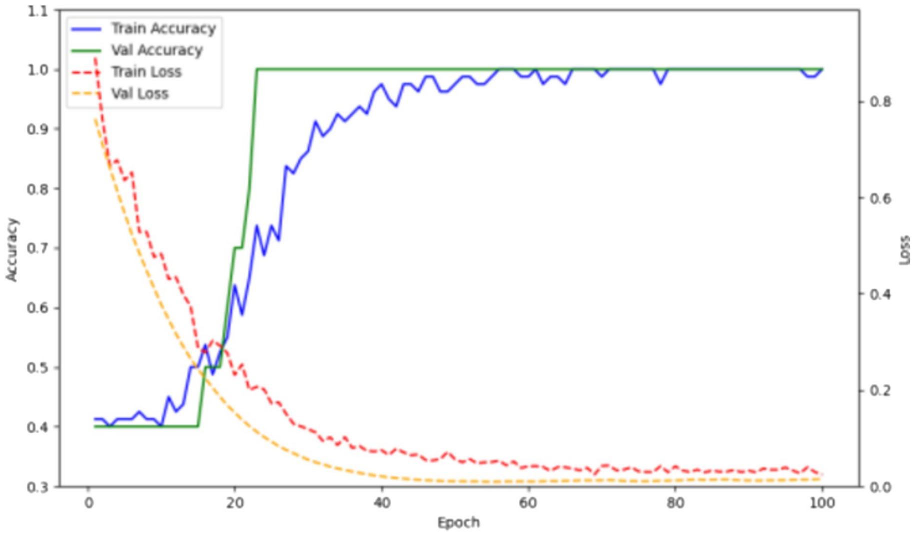


FIGURE 6
Accuracy and loss curves of the deep learning model for pathogen detection on training and validation sets.

proceed with empirical antibiotic treatment. This study reveals that, in addition to affecting the recovery process of AHO patients, red blood cell status also influences the results of bacterial culture. This may be related to how red blood cell status affects the overall function of the immune system, thereby impacting the proliferation and distribution of bacteria within the body. Abnormal red blood cell function may weaken the immune response to infections, allowing pathogens to become localized in tissues and making them more difficult to detect in blood cultures.

Despite this, the study has some limitations. First, it is a single-center study, including both laboratory parameters and blood culture results. This suggests that the study may have potential issues with generalizability. Second, the study only established a classification method between acute hematogenous osteomyelitis and benign bone tumors, malignant bone tumors, and fractures, without exploring classifications for different subtypes and disease progressions of osteomyelitis, which is also due to a lack of data. Third, while the study predicted whether pathogens could be successfully cultured from patients, it did not further classify the cultured pathogens. Fourth, this study only included laboratory parameters from AHO patients, while the diagnosis of AHO can also rely on imaging information. Due to a lack of data, this study did not include any medical imaging information. Therefore, the next step for this research is to collect data from multiple centers, expand the cohort, and include more data modalities.

5 Conclusion

Early laboratory parameters can accurately diagnose whether pediatric patients have acute hematogenous osteomyelitis. Laboratory parameters can describe the severity of acute hematogenous osteomyelitis and are somewhat correlated with the patient's length of hospital stay. Early laboratory parameters can predict whether a patient's blood sample will successfully culture pathogens, thereby guiding clinical decision-making, indirectly improving clinical outcomes, and shortening the hospital stay.

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 Ethics Committee of The Second Hospital of Dalian Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

YX: Formal analysis, Investigation, Methodology, Writing – original draft. QK: Data curation, Investigation, Validation, Writing – original draft. YG: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – original draft. JS: Funding acquisition, Project administration, 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.

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

Athey, A. G., Mignemi, M. E., Gheen, W. T., Lindsay, E. A., Jo, C. H., and Copley, L. A. (2019). Validation and modification of a severity of illness score for children with acute Hematogenous osteomyelitis. *J. Pediatr. Orthop.* 39, 90–97. doi: 10.1097/bpo.0000000000000879

Burns, J. D., Upasani, V. V., Bastrom, T. P., Baldwin, K. D., Schoenecker, J. G., Shore, B. J., et al. (2023). Age and CRP associated with improved tissue pathogen identification in children with blood culture negative osteomyelitis: results from the CORTICES multicenter database. *J. Pediatr. Orthop.* 43, e603–e607. doi: 10.1097/bpo.0000000000002448

Calvo, C., Núñez, E., Camacho, M., Clemente, D., Fernández-Cooke, E., Alcobendas, R., et al. (2016). Epidemiology and management of acute, uncomplicated septic arthritis and osteomyelitis: Spanish multicenter study. *Pediatr. Infect. Dis. J.* 35, 1288–1293. doi: 10.1097/inf.0000000000001309

Campbell, J. I., Shanahan, K. H., Bartick, M., Ali, M., Goldmann, D., Shaikh, N., et al. (2023). Racial and ethnic differences in length of stay for US children hospitalized for acute osteomyelitis. *J. Pediatr.* 259:113424. doi: 10.1016/j.jpeds.2023.113424

Consalvo, S., Hinterwimmer, F., Neumann, J., Steinborn, M., Salzmann, M., Seidl, F., et al. (2022). Two-phase deep learning algorithm for detection and differentiation of Ewing sarcoma and acute osteomyelitis in paediatric radiographs. *Anticancer Res.* 42, 4371–4380. doi: 10.21873/anticanres.15937

Dartnell, J., Ramachandran, M., and Katchburian, M. (2012). Haematogenous acute and subacute paediatric osteomyelitis: a systematic review of the literature. *J. Bone Joint Surg. Br.* 94, 584–595. doi: 10.1302/0301-620x.94b5.28523

Dich, V. Q., Nelson, J. D., and Haltalin, K. C. (1975). Osteomyelitis in infants and children. A review of 163 cases. *Am. J. Dis. Child.* 129, 1273–1278. doi: 10.1001/archpedi.1975.02120480007004

Disch, K., Hill, D. A., Snow, H., and Dehority, W. (2023). Clinical outcomes of pediatric osteomyelitis. *BMC Pediatr.* 23:54. doi: 10.1186/s12887-023-03863-z

Doern, G. V., Carroll, K. C., Diekema, D. J., Garey, K. W., Rupp, M. E., Weinstein, M. P., et al. (2019). Practical guidance for clinical microbiology laboratories: a comprehensive update on the problem of blood culture contamination and a discussion of methods for addressing the problem. *Clin. Microbiol. Rev.* 33:e00009-19. doi: 10.1128/cmr.00009-19

Dong, S. Z., Zhu, M., and Bulas, D. (2019). Techniques for minimizing sedation in pediatric MRI. *J. Magn. Reson. Imaging* 50, 1047–1054. doi: 10.1002/jmri.26703

Funk, S. S., and Copley, L. A. (2017). Acute Hematogenous osteomyelitis in children: pathogenesis, diagnosis, and treatment. *Orthop. Clin. North Am.* 48, 199–208. doi: 10.1016/j.ocl.2016.12.007

Gao, Q., Liu, Q., Zhang, G., Lu, Y., Li, Y., Tang, M., et al. (2024). Identification of pathogen composition in a Chinese population with iatrogenic and native vertebral osteomyelitis by using mNGS. *Ann. Med.* 56:2337738. doi: 10.1080/07853890.2024.2337738

Jahan, T. A., Lapin, N. A., O'Connell, M. T., Jo, C., Ma, Y., Tareen, N. G., et al. (2024). Accelerated severity of illness score enhances prediction of complicated acute hematogenous osteomyelitis in children. *Pediatr. Infect. Dis. J.* doi: 10.1097/inf.0000000000004535

Kim, J., Yoo, G., Lee, T., Kim, J. H., Seo, D. M., and Kim, J. (2022). Classification model for diabetic foot, necrotizing fasciitis, and osteomyelitis. *Biology* 11:1310. doi: 10.3390/biology11091310

Kremers, H. M., Nwojo, M. E., Ransom, J. E., Wood-Wentz, C. M., Melton, L. J., and Huddleston, P. M. (2015). Trends in the epidemiology of osteomyelitis: a population-based study, 1969 to 2009. *J. Bone Joint Surg. Am.* 97, 837–845. doi: 10.2106/jbjs.N.01350

Liu, Y., Zhao, K., Liu, Y., Sun, Y. H., Li, M. X., Yu, M., et al. (2024). Bone and joint infection complicated with sepsis in neonates and infants under three months of age. *J. Pediatr.* 100, 156–162. doi: 10.1016/j.jped.2023.09.003

Manz, N., Krieg, A. H., Buettcher, M., Ritz, N., and Heininger, U. (2020). Long-term outcomes of acute osteoarticular infections in children. *Front. Pediatr.* 8:587740. doi: 10.3389/fped.2020.587740

Manz, N., Krieg, A. H., Heininger, U., and Ritz, N. (2018). Evaluation of the current use of imaging modalities and pathogen detection in children with acute osteomyelitis and septic arthritis. *Eur. J. Pediatr.* 177, 1071–1080. doi: 10.1007/s00431-018-3157-3

McNeil, J. C. (2020). Acute Hematogenous osteomyelitis in children: clinical presentation and management. *Infect Drug Resist* 13, 4459–4473. doi: 10.2147/idr.S257517

McNeil, J. C., Vallejo, J. G., Kok, E. Y., Sommer, L. M., Hultén, K. G., and Kaplan, S. L. (2019). Clinical and microbiologic variables predictive of orthopedic complications following *Staphylococcus aureus* acute hematogenous osteoarticular infections in children. *Clin. Infect. Dis.* 69, 1955–1961. doi: 10.1093/cid/ciz109

Paliwal, A. K., Sahdev, R., Deshwal, A., and Ram, B. (2021). Role of ultrasound in the diagnosis of paediatric acute osteomyelitis. *J Ultrason* 21, e34–e40. doi: 10.15557/JoU.2021.0005

Russell, C. D., Ramaesh, R., Kalima, P., Murray, A., and Gaston, M. S. (2015). Microbiological characteristics of acute osteoarticular infections in children. *J. Med. Microbiol.* 64, 446–453. doi: 10.1099/jmm.0.000026

Schmitt, S. K. (2017). Osteomyelitis. *Infect. Dis. Clin. N. Am.* 31, 325–338. doi: 10.1016/j.idc.2017.01.010

Section, J., Gibbons, S. D., Barton, T., Greenberg, D. E., Jo, C. H., and Copley, L. A. (2015). Microbiological culture methods for pediatric musculoskeletal infection: a guideline for optimal use. *J. Bone Joint Surg. Am.* 97, 441–449. doi: 10.2106/jbjs.N.00477

Stephan, A. M., Faino, A., Caglar, D., and Klein, E. J. (2022). Clinical presentation of acute osteomyelitis in the pediatric emergency department. *Pediatr. Emerg. Care* 38, e209–e213. doi: 10.1097/pec.0000000000002217

Stephan, A. M., Platt, S., Levine, D. A., Qiu, Y., Buchhalter, L., Lyons, T. W., et al. (2024). A novel risk score to guide the evaluation of acute hematogenous osteomyelitis in children. *Pediatrics* 153:3153. doi: 10.1542/peds.2023-063153

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al. (2017). Attention is all you need. Advances in neural information processing systems. 30.

Walter, N., Bärtl, S., Alt, V., and Rupp, M. (2021). The epidemiology of osteomyelitis in children. *Children* 8:1000. doi: 10.3390/children8111000

Wang, J., Zhanghuang, C., Tan, X., Mi, T., Liu, J., Jin, L., et al. (2022). A nomogram for predicting Cancer-specific survival of osteosarcoma and Ewing's sarcoma in children: a SEER database analysis. *Front. Public Health* 10:837506. doi: 10.3389/fpubh.2022.837506

Wen, Y., Wang, C., Jia, H., Liu, T., Yu, J., and Zhang, M. (2023). Comparison of diagnosis and treatment of MSSA and MRSA osteomyelitis in children: a case-control study of 64 patients. *J. Orthop. Surg. Res.* 18:197. doi: 10.1186/s13018-023-03670-3

Woods, C. R., Bradley, J. S., Chatterjee, A., Copley, L. A., Robinson, J., Kronman, M. P., et al. (2021). Clinical practice guideline by the pediatric infectious diseases society and the infectious diseases society of America: 2021 guideline on diagnosis and management of acute hematogenous osteomyelitis in pediatrics. *J Pediatric Infect Dis Soc* 10, 801–844. doi: 10.1093/jpids/piab027

Zaki, I., and Morrison, W. B. (2024). Osteomyelitis and septic arthritis of the foot and ankle: imaging update. *Clin. Podiatr. Med. Surg.* 41, 745–758. doi: 10.1016/j.cpm.2024.04.007

frontiers | Frontiers in Microbiology

TYPE Systematic Review
PUBLISHED 15 November 2024
DOI 10.3389/fmicb.2024.1510139

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Yuzhuo Gu,
Beihang University, China
Hanyu Tian,
The Chinese University of Hong Kong,
Shenzhen, China

*CORRESPONDENCE
Yuting Wang
✉ 897259567@qq.com
Ying Liu
✉ sjliuying@sj-hospital.org
Jia-He Wang
✉ wangjh1@sj-hospital.org

[†]These authors have contributed equally to this work

RECEIVED 12 October 2024
ACCEPTED 06 November 2024
PUBLISHED 15 November 2024

CITATION
Tian T, Zhang X, Zhang F, Huang X, Li M, Quan Z, Wang W, Lei J, Wang Y, Liu Y and Wang J-H (2024) Harnessing AI for advancing pathogenic microbiology: a bibliometric and topic modeling approach.
Front. Microbiol. 15:1510139.
doi: 10.3389/fmicb.2024.1510139

Harnessing AI for advancing pathogenic microbiology: a bibliometric and topic modeling approach

Tian Tian[1†], Xuan Zhang[1†], Fei Zhang[1], Xinghe Huang[2], Minglin Li[1], Ziwei Quan[1], Wenyue Wang[3], Jiawei Lei[1], Yuting Wang[4]*, Ying Liu[5]* and Jia-He Wang[1]*

[1]Department of Family Medicine, Shengjing Hospital of China Medical University, Shenyang, China, [2]College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou, China, [3]Department of General Practice, The First Hospital of China Medical University, Shenyang, China, [4]Department of Cardiology, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [5]Department of Nephrology, Shengjing Hospital of China Medical University, Shenyang, China

Introduction: The integration of artificial intelligence (AI) in pathogenic microbiology has accelerated research and innovation. This study aims to explore the evolution and trends of AI applications in this domain, providing insights into how AI is transforming research and practice in pathogenic microbiology.

Methods: We employed bibliometric analysis and topic modeling to examine 27,420 publications from the Web of Science Core Collection, covering the period from 2010 to 2024. These methods enabled us to identify key trends, research areas, and the geographical distribution of research efforts.

Results: Since 2016, there has been an exponential increase in AI-related publications, with significant contributions from China and the USA. Our analysis identified eight major AI application areas: pathogen detection, antibiotic resistance prediction, transmission modeling, genomic analysis, therapeutic optimization, ecological profiling, vaccine development, and data management systems. Notably, we found significant lexical overlaps between these areas, especially between drug resistance and vaccine development, suggesting an interconnected research landscape.

Discussion: AI is increasingly moving from laboratory research to clinical applications, enhancing hospital operations and public health strategies. It plays a vital role in optimizing pathogen detection, improving diagnostic speed, treatment efficacy, and disease control, particularly through advancements in rapid antibiotic susceptibility testing and COVID-19 vaccine development. This study highlights the current status, progress, and challenges of AI in pathogenic microbiology, guiding future research directions, resource allocation, and policy-making.

KEYWORDS

pathogenic microorganisms, artificial intelligence (AI), machine learning (ML), deep learning (DL), bibliometrics, topic modeling, antimicrobial resistance (AMR)

1 Introduction

Pathogenic microorganisms, including viruses, bacteria, fungi, and parasites, cause infections and diseases in hosts. Since the 1960s, the widespread use of antibiotics has driven the evolution of these microorganisms through natural selection, gene recombination, and horizontal gene transfer (HGT), leading to antibiotic resistance (AMR).

AMR results in millions of deaths annually worldwide, posing a severe threat to public health (Saha and Sarkar, 2021; Uddin et al., 2021). Traditional culture-based methods fail to address the increasing genetic diversity and resistance of pathogens. In the era of big data, research on pathogenic microorganisms heavily relies on high-throughput sequencing, metagenomics, proteomics, and targeted techniques (Lewis et al., 2021; Wani et al., 2022). Effectively organizing, analyzing, and interpreting the vast amounts of biomedical data generated has emerged as a new challenge.

AI, a field that simulates and extends human intelligence through computational devices, provides powerful tools to address these challenges. Machine learning (ML) improves computer performance through pattern recognition and analysis, enabling precise microbial classification, biomarker identification, small molecule compound library screening, and novel anti-infective drug discovery.

Deep learning (DL), comprising multilayer neural networks, boosts data generation capabilities for pathogenic microorganisms through neural networks, generative models, and variational autoencoders (Huo and Wang, 2024; Wong et al., 2023). Computer vision (CV) rapidly detects pathogens in microscope or fluorescence sensor images (Zhao et al., 2024; Matias et al., 2021). Natural language processing (NLP) automatically identifies information from scientific literature on pathogen research (Jimeno-Yepes and Verspoor, 2023) and analyzes bacteriophage genomes to predict their life cycles (Tynecki et al., 2020).

The application of AI in pathogenic microbiology has been widely explored, with many scholars evaluating its use in related research. Literature reviews date back to 2014. Specifically, Nourani et al. studied ML, homology prediction, and structural prediction in predicting pathogen-host protein interactions (PHI) between 2009 and 2014, crucial for understanding infection mechanisms (Nourani et al., 2015). Rondon-Villarreal et al. reviewed ML in antimicrobial peptide design, a potential new class of antimicrobial drugs to combat AMR (Rondon-Villarreal et al., 2014). Qu et al. comprehensively reviewed ML in microbiology, covering microbial classification from high-throughput sequencing data, environmental and host phenotype prediction, and microbial-disease association analysis (Qu et al., 2019). Agany et al. explored data mining and ML in understanding vector-host-pathogen relationships from 2012 to 2020, highlighting advances in DL and association rule analysis (Agany et al., 2020). Peiffer-Smadja et al. studied ML in clinical microbiology, identifying 97 ML systems aimed at assisting clinical microbiologists with bacterial, parasitic, viral, and fungal infection analysis and antimicrobial sensitivity assessment up to 2020 (Peiffer-Smadja et al., 2020). Pillai et al. summarized various AI models (e.g., logistic regression, random forests, support vector machines, neural networks, ensemble methods) in predicting zoonotic disease outbreaks and identifying risk factors (Pillai et al., 2022). He et al. introduced AI's role in infectious disease drug delivery, including drug development, resistance prediction, dose optimization, and drug combination selection (He et al., 2021). Hu et al. discussed ML's broad applications in protozoan pathogen and infectious disease research, covering detection, diagnosis, monitoring, host–parasite interactions, drug discovery, and vaccine development (Hu et al., 2022). Kaur et al. reviewed AI techniques in predicting and monitoring vector-borne diseases and their pathogens, noting significant progress in disease prediction, vector identification, and outbreak monitoring through ML and DL (Kaur et al., 2022).

Despite several studies exploring the application of AI in specific areas of pathogenic microbiology, a systematic analysis of the overall development trends and knowledge structure of the field is lacking. Previous literature reviews have primarily focused on AI's performance in specific application scenarios. These studies provide important references for understanding AI's value in specific applications but fail to offer a comprehensive grasp of the overall development landscape of AI in pathogenic microbiology research.

In this context, bibliometrics and topic modeling offer powerful methods to explore and understand scientific research in this domain. Bibliometrics, a statistical method widely used to analyze publication trends and relationships in the medical field, includes evaluative and relational bibliometrics. The latter reveals hidden relationships and research status by analyzing metadata from authors, papers, and journals (Ninkov et al., 2022). Topic modeling, a natural language processing technique, identifies latent semantic patterns in document collections, helping researchers discover cross-disciplinary themes and research trends. Latent Dirichlet Allocation (LDA) is the most widely utilized technique for this purpose (Vayansky and Kumar, 2020). This study aims to conduct a large-scale quantitative analysis of AI applications in the field of pathogenic microbiology through bibliometrics and topic modeling methods. Compared to existing studies, this paper has the following innovations and contributions: (1) By utilizing 27,420 publications spanning 2010 to 2024, the study conducts a comprehensive quantitative analysis of the field for the first time, covering a wide scope; (2) The integration of bibliometrics and topic modeling techniques not only reveals research hotspots and trends but also deeply explores the potential knowledge structure; and (3) A systematic review of AI's advancements in eight major application areas within pathogenic microbiology provides a scientific basis for future research directions and resource allocation.

2 Methods

2.1 Data collection

To ensure the scientific rigor and authority of the literature review, we retrieved data from the Web of Science Core Collection (WoSCC), the oldest and most widely used research publication and citation database globally (Birkle et al., 2020). The citation index includes various versions of WoSCC, such as the Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Current Chemical Reactions (CCR-EXPANDED), and Index Chemicus (IC).

The search terms were derived from key phrases mentioned in previous review articles on AI applications in pathogenic microbiology (Table 1). The final search string was: TS = ("Pathogen-host protein–protein interactions" OR Host OR Pathogen OR "Drug Resistance" OR "Antimicrobial Peptides" OR Viruses OR Bacteria OR Fungi OR "Vector-Host-Pathogen Relationships" OR Vector OR Parasites OR "Infectious Diseases" OR "Pathogenic Microbes").

AND TS = ("Deep Learning" OR "Association Rule Mining" OR "Artificial Neural Network" OR "Support Vector Machine" OR "K-nearest Neighbors" OR "Decision Trees" OR "Regression Trees" OR "Classification Trees" OR "Gradient Boosting" OR "Adaptive Boosting" OR "eXtreme Gradient Boosting" OR "Long Short Term Memory network" OR "Generative Adversarial Network" OR "Auto-Encoder" OR "Convolutional Neural Networks" OR "Ensemble Classifiers" OR "Support Vector Machine"). This search yielded 151,593 results.

TABLE 1 Sources of search terms for this study.

Author	Paper title	Keywords related to AI	Keywords related to pathogenic microorganisms	Year
Nourani	Computational approaches for prediction of pathogen-host protein–protein interactions	Homology-based prediction Structure-based prediction	Pathogen-host protein–protein interactions Host Pathogen	2015
Rondon-Villarrea	Machine Learning in the Rational Design of Antimicrobial Peptides	Machine learning	Drug Resistance Antimicrobial Peptides	2014
Qu	Application of Machine Learning in Microbiology	Supervised Learning Unsupervised Learning Support Vector Machine Naïve Bayes Random Forest K Nearest Neighbor	Viruses Bacteria Fungi	2019
Agany	Assessment of vector-host-pathogen relationships using data mining and machine learning	Data Mining Deep Learning Association Rule Mining	Vector-Host-Pathogen Relationships Vector	2020
Peiffer-Smadja	Machine learning in the clinical microbiology laboratory: has the time come for routine practice?	Artificial Neural Network Support Vector Machine Logistic Regression K-nearest Neighbors Decision\Regression\Classification Trees Gradient Boosting Adaptive Boosting	Parasites	2020
Pillai	Artificial Intelligence Models for Zoonotic Pathogens: A Survey	eXtreme Gradient Boosting Long Short Term Memory network Generative Adversarial Network Auto-Encoder		2020
He	Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases		Infectious Diseases	2021
Hu	Machine Learning and Its Applications for Protozoal Pathogens and Protozoal Infectious Diseases	Convolutional Neural Networks		2022
Kaur	Artificial Intelligence Techniques for Predictive Modeling of Vector-Borne Diseases and its Pathogens: A Systematic Review	Ensemble Classifiers Support Vector Machine	Pathogenic Microbes	2022

2.1.1 Screening criteria

Inclusion and exclusion criteria were established to filter the results. The inclusion criteria focused on research articles related to AI and pathogenic microorganisms. Exclusion criteria were set for outdated articles, non-English literature, irrelevant disciplines, conference abstracts, and duplicate documents. The screening was performed independently by two authors to ensure accuracy and consistency. The specific steps were:

1. **Year Limitation:** Restricting the publication years to 2010–2024 eliminated 4,858 articles. This timeframe was selected because the rapid advancements in AI and pathogenic microbiology, particularly the widespread application of metagenomics and high-throughput sequencing, began around 2010 (Sheetal Ambardar et al., 2016; Park et al., 2016).
2. **Language Restriction:** Non-English articles (163) were removed to ensure the inclusion of high-impact research and important results published in major journals.
3. **Disciplinary Focus:** Articles from non-medical fields (78,574), such as engineering or computer science, were excluded. These fields often focus more on technical developments and algorithm optimization, which could introduce noise into the bibliometric analysis.
4. **Document Type:** Only “Article” and “Review Article” categories were included, eliminating 40,457 documents from other types like proceeding papers, book chapters, letters, and news items.
5. **Duplicate Removal:** Using Endnote software, 121 duplicate documents were automatically removed to ensure data uniqueness and completeness.

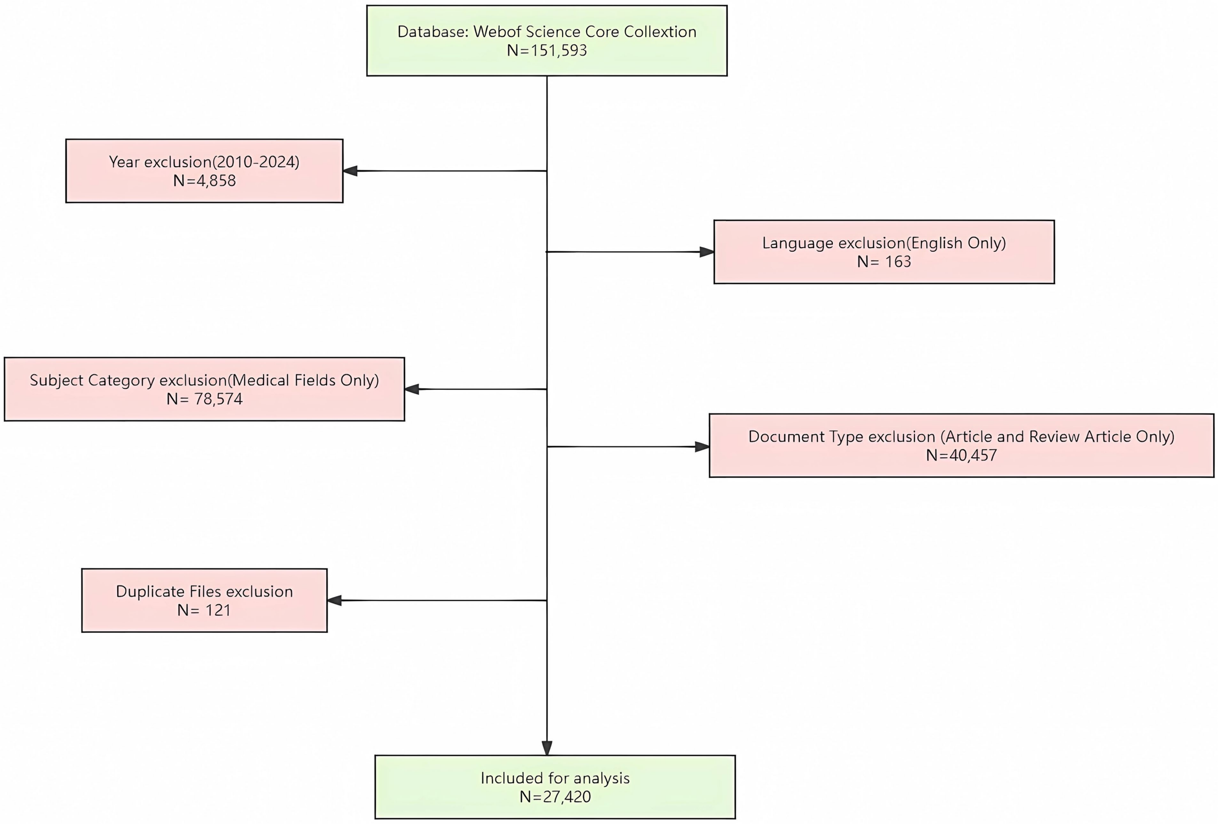


FIGURE 1
Flow chart of literature search and selection.

The final dataset comprised 27,420 articles for analysis (Figure 1).

2.2 Bibliometric analysis

The data was exported in plain text format, including full records and cited references. The clean dataset was imported into three software programs for visualization analysis: R-Bibliometrix 4.3.1, CiteSpace 6.1.R6 (64-bit) Advanced, and VOSviewer 1.6.19.

2.2.1 Bibliometrix

This R-based bibliometric analysis package, launched in 2017, offers robust data processing and multi-dimensional chart generation capabilities (Aria and Cuccurullo, 2017; Arruda et al., 2022). We used Bibliometrix 4.3.1 to analyze annual publications, author H-indices, productivity over time, high-impact journals, and highly cited papers.

2.2.2 CiteSpace

Developed by Professor Chaomei Chen in 2004, CiteSpace is a Java-based scientific literature analysis software capable of document co-citation, collaboration network, and burst term analysis (Chen, 2018). We used CiteSpace 6.1.R6 (64-bit) Advanced to generate co-occurrence maps of institutions.

2.2.3 VOSviewer

Launched in 2010, VOSviewer is a software tool for creating and exploring maps based on network data, offering network, overlay, and density visualizations (Arruda et al., 2022). We used VOSviewer for author and country co-occurrence collaboration analysis.

2.3 Topic modeling analysis

Compared to traditional bibliometric keyword clustering methods, topic modeling offers more precise and detailed research classifications, uncovering the underlying structures and dynamic trends within research fields. In our study, we employed Latent Dirichlet Allocation (LDA) for topic modeling using the Python Gensim library. LDA is a generative model that leverages unsupervised machine learning to analyze large volumes of unstructured data, eliminating the need to divide data into training and test sets. It assumes that documents comprise multiple topics, each represented by a probability distribution over words (Chauhan and Shah, 2021; Yang et al., 2024).

Initially, we conducted text preprocessing, including the removal of stopwords and punctuation, as well as stemming, to ensure data consistency and cleanliness. Subsequently, we set the parameters for the topic modeling. Optimization of topics was performed through perplexity and coherence evaluation. Finally, we generated the topic-word distribution and a topic-term relationship network graph.

The specific hyperparameter choices for the LDA model were as follows: alpha = "symmetric" (symmetric prior), and eta = None (default prior). These parameters control the prior beliefs regarding the document-topic and topic-word distributions. The chunksize was set to 2000, meaning that the corpus of 27,420 documents was divided into approximately 14 chunks for processing, thereby avoiding the necessity of loading all documents into memory simultaneously. The passes parameter was set to 1 since the model's performance was satisfactory with a single pass through the corpus. The model was trained using the LdaModel class provided by Gensim, which implements an online LDA algorithm that enables streaming and incremental training of the corpus, thereby effectively handling large-scale text data. The save and load methods were used for model persistence, ensuring the reproducibility of the experimental results.

3 Results

3.1 Publication trends analysis

Polynomial regression analysis (Figure 2) from 2010 to 2023 reveals a significant upward trend in the number of publications, with exponential growth evident since 2016. The number of publications is projected to reach approximately 4,500 by 2024. This growth is primarily attributed to the increase in interdisciplinary collaboration, the impact of global health challenges such as the COVID-19 pandemic, and the rapid advancements in AI and computational technologies, particularly breakthroughs in deep learning algorithms, convolutional neural networks, and disease prediction models since 2016 (Jelodar et al., 2019).

3.2 Authors

Figure 3A presents the metrics for the top four contributing authors. Wang Wei leads with 69 publications and an H-index of 25, signifying that at least 25 publications have been cited at least 25 times (Zhang et al., 2022). Wang Jing was notably prolific in 2018, publishing 11 papers and achieving a Total Citations per Year (TCpY) score of 98.67 (Figure 3B). VOSviewer analysis (Figures 3C,D) of 70 authors, each with a minimum of 5 publications and 1,000 citations, reveals three primary collaborative groups centered around Li Hao, Zhang Wei, and Wang Lie. Notably, Zhang Wei's collaborative network is the largest, comprising 13 members. The collaboration between Li Hao and Chen Wei is the most frequent, with 28 co-authored papers (Bihari et al., 2023).

3.3 Institutions

Figure 4 illustrates that there are 917 collaborative interactions among 776 institutions. Although the overall network density is low (0.003), certain institutions display frequent and intensive collaborations. This phenomenon can be attributed to the high specialization in pathogenic microbiology and AI technologies, which leads collaborations to be concentrated among a select few capable institutions. As shown in Table 2, the Chinese Academy of Sciences leads with 486 publications, while the Universitair Medisch Centrum Utrecht demonstrates significant research impact with a betweenness centrality of 0.49.

3.4 Countries

Figure 5A indicates that China and the United States have been leading in pathogenic microbiology research. Notably, China's publication volume significantly decreased in 2021, likely due to the impact of the COVID-19 pandemic. However, since 2022, China's publication rate has grown exponentially, surpassing other countries. VOSviewer analysis (filtering for countries with at least *100 publications) revealed an international collaboration network comprising 32 countries. The thickness of the connecting lines indicates collaboration strength, with China and the U.S. exhibiting the tightest cooperation (link strength = 974). This suggests 974 instances of collaboration between researchers from these two countries, reflecting their central role and significant advantages in knowledge and resource sharing, which are crucial for advancing pathogenic microbiology research and addressing global health challenges (Figure 5B).

3.5 Journals

In analyzing the evolution and trends of AI in pathogenic microbiology research, we identified the top ten journals in this field, including their H-index, impact factor, and JCR indicators (Table 3). These metrics reflect the research activity and the journals' influence within the academic community. "Computers in Biology and Medicine" has the highest number of publications in this field, while "Clinical Infectious Diseases," the only Q1 journal among the top ten by publication volume, is the most cited, demonstrating its authority. Notably, four of the top ten journals are Q2, indicating that research outcomes are increasingly being published in higher-quality journals.

3.6 Topic modeling

To determine the optimal number of topics, we undertook the following steps: First, we trained LDA models with varying numbers of topics (2 to 15) and calculated their perplexity scores on the test set. Lower perplexity indicates a better model fit (Figure 6A). Second, we computed the topic coherence score for each model, which measures the semantic consistency of words within a topic; higher values indicate more coherent topic structures (Figure 6B). Finally, we plotted perplexity and coherence scores on a scatter plot (Figure 6C). The top-right region of the plot shows data points for 8, 9, 10, and 12 topics, which performed well in balancing perplexity and coherence. Further manual analysis revealed that although 9, 10, and 12 topics offered higher model performance, they led to overly fine and dispersed classifications, which are impractical for real-world applications. An 8-topic model provided an efficient and practical classification structure, laying a solid foundation for further interpretation. Consequently, we selected the 8-topic model. The resulting themes include AI in pathogen detection, drug resistance, transmission and control, genomics, treatment optimization, ecology, vaccine development, and data analysis and management (Table 4). A word cloud was then utilized to visually represent the intrinsic connections between different research themes, with each cluster marked in different colors. The clusters were interconnected through shared keywords (red nodes), and the size of each node reflected keyword frequency, while the thickness of connecting lines indicated the distribution strength of words within specific topics (Figure 7).

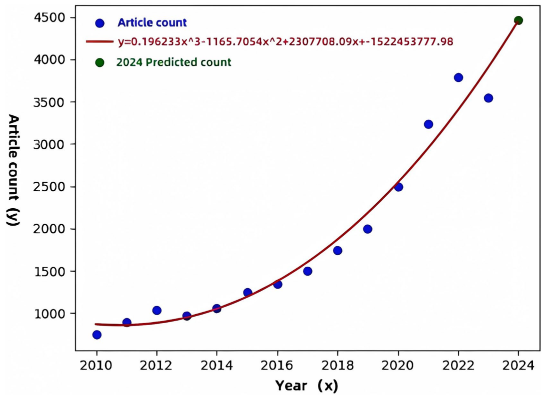


FIGURE 2
Article quantity trend analysis.

4 Discussion

4.1 Eight major topics

This section summarizes the eight topics identified through topic modeling and discusses the research advancements within each domain.

4.1.1 Application of artificial intelligence in pathogen detection

Traditional diagnostic methods, such as microbial culture and isolation, are often time-consuming and prone to false-negative results (Daim et al., 2006). The application of multi-modal data fusion techniques in pathogen detection has gained significant attention in recent years. These techniques integrate image data with genetic data to provide more comprehensive and accurate diagnostic outcomes. For instance, Khan et al. (2019) developed models using automated image capture technology and convolutional neural networks (CNN), successfully classifying and identifying Gram-stained blood cultures. Their model achieved a classification accuracy of 94.9% for both Gram-positive cocci and Gram-negative bacteria. By integrating microbial genome sequencing data with the capabilities of CNNs, researchers can further subtype pathogens based on pattern recognition. This multi-modal fusion approach significantly enhances sensitivity and specificity in diagnostics by simultaneously analyzing the visual and genetic characteristics of pathogens, thus making pathogen diagnosis more efficient and precise.

Additionally, machine learning models can rapidly analyze complex data patterns, thereby improving diagnostic speed and accuracy (Smith et al., 2018). For instance, models used for DNA sequencing can quickly process the genomes of bacteria and viruses (Ali et al., 2023).

4.1.2 Application of artificial intelligence in antimicrobial resistance research

Artificial intelligence has been effectively utilized in the analysis and prediction of microbial drug resistance, marking a significant advancement in antimicrobial resistance research. The increasing prevalence of resistant bacteria underscores the critical importance of analyzing genomic and sequence data. Traditional antibiotic susceptibility testing (AST) methodologies require a minimum of 4 days, which is excessively time-consuming for urgent clinical scenarios where swift decision-making is crucial, especially in the face of rapidly spreading infections (Mardis, 2008). This predicament underscores the urgent need for innovative diagnostic techniques that can adapt to the rapid evolution of antibiotic resistance.

The integration of techniques like MALDI-TOF MS with sophisticated data analysis algorithms has been shown to expedite the identification of resistant strains (Garcia et al., 2024). A retrospective clinical case study involving 63 patients revealed that adopting such methodologies would have altered the clinical management of nine patients, benefiting eight of them (89%). Consequently, machine learning based on MALDI-TOF mass spectrometry emerges as an essential new tool for therapy optimization and antibiotic stewardship (Theodosiou and Read, 2023). Deep learning algorithms, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, can perform rapid and accurate antibiotic susceptibility testing by classifying bacteria into active or non-active strains (Weis et al., 2022).

AI technologies, encompassing machine learning and natural language processing, enable the processing of vast quantities of genomic data, which leads to the identification of resistance-associated genetic mutations and evolutionary patterns (Yu et al., 2018). This capability not only enhances our understanding of how bacteria develop drug resistance but also provides invaluable insights for novel drug development (Zhou and Troyanskaya, 2015). Furthermore, machine learning models have the potential to predict mutational trends and resistance to specific drugs, thereby aiding clinicians in selecting the most effective treatment regimens (Gupta et al., 2021).

4.1.3 Application of artificial intelligence in pathogen transmission and control

Artificial intelligence demonstrates significant potential in monitoring and controlling pathogen transmission. By employing machine learning to recognize transmission patterns, it provides vital decision support for public health authorities, enabling the implementation of more effective outbreak control strategies (Májek et al., 2021). AI can analyze historical epidemic data to predict future disease transmission (Ren et al., 2023). Additionally, AI technologies are utilized for real-time monitoring of infection trends, allowing rapid responses to outbreaks. AI-driven warning systems enhance the predictive capacity for future outbreaks, improving resource allocation and management strategies (Vahedi et al., 2021). These technologies provide scientific evidence for disease control and prevention, bolstering the resilience of public health systems.

4.1.4 Application of artificial intelligence in pathogen genomics

Deep learning, as a crucial AI technology, offers new perspectives and tools for analyzing diversity and evolution in pathogen genomics research. Xu et al. employed deep learning algorithms to efficiently identify various antimicrobial peptides from metagenomic data, significantly advancing the development of next-generation antimicrobials (Wang et al., 2024). AI's ability to analyze large volumes of genomic sequence data allows it to identify and compare genetic characteristics of diverse microorganisms, revealing their evolutionary

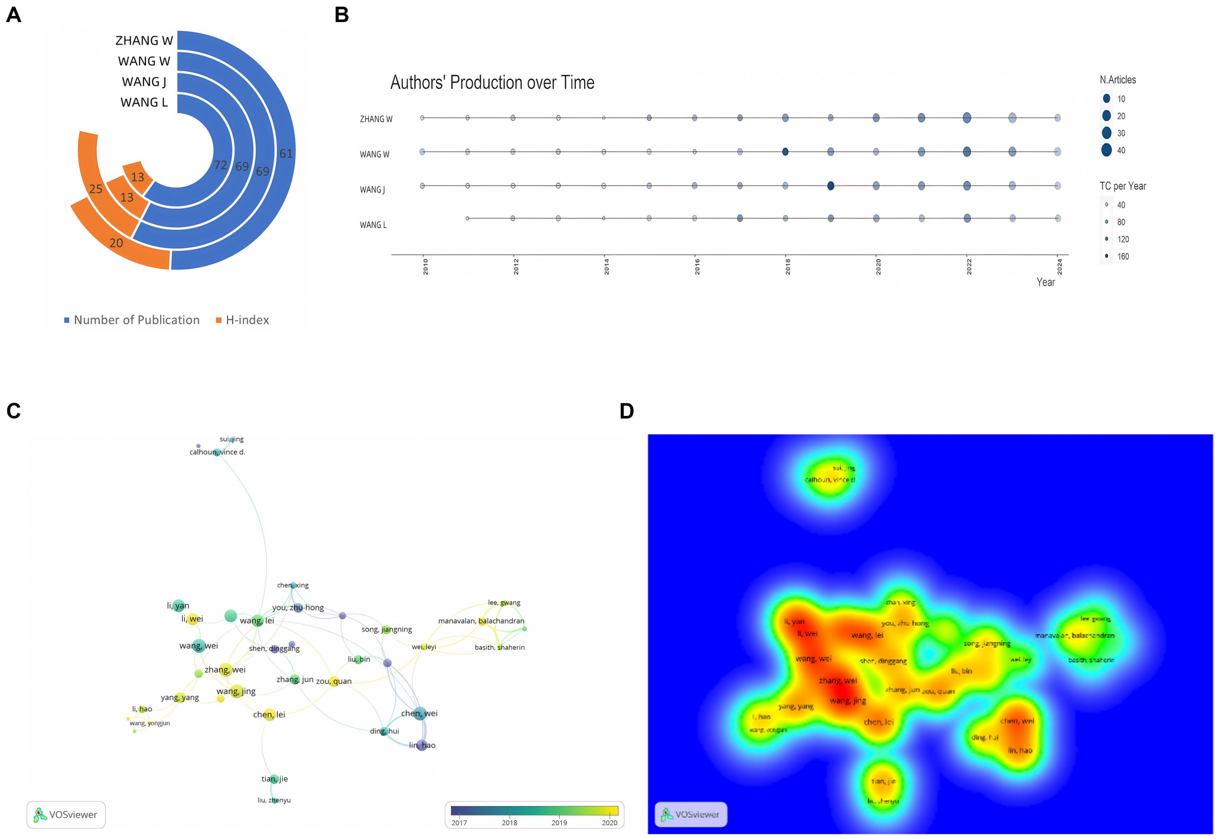


FIGURE 3
(A) Total number of publications and H-index of top 4 authors; **(B)** Authors' production over time; **(C)** Network visualization of author co-authorship analysis; **(D)** Density visualization of author co-authorship analysis.

relationships and functional traits (Xu et al., 2020). This robust potential for genome annotation and functional prediction provides vital support for microbial ecology and functional research (Angly et al., 2006). Furthermore, constructed databases allow for in-depth exploration of complex interactions between microorganisms and environmental contexts (Sun et al., 2023). This provides valuable insights for research in microbial genomics, ecological analysis, and disease prevention.

4.1.5 Application of artificial intelligence in optimization of treatment strategies

Artificial intelligence is playing an increasingly important role in optimizing treatment strategies for pathogens. By analyzing clinical data, AI can predict the efficacy of various treatment regimens and adjust strategies in real-time according to individual patient condition changes. Li Jinquan utilized AI to identify differences in high-dimensional features of antimicrobial candidate proteins, discovering the best-in-class lytic enzyme LLysSA9, effective in treating bovine mastitis and combating *Staphylococcus aureus* infections (Hirose et al., 2024). This personalized medicine approach not only enhances treatment outcomes but also reduces unnecessary treatments and potential side effects. AI-powered decision support systems integrate medical literature, patient data, and clinical trial results to provide scientific foundations for optimizing treatment plans (Wong et al., 2023).

Moreover, AI excels in drug repurposing and new drug development, using models to simulate the effects of different drugs on pathogens, thereby advancing personalized treatment (Zhang et al., 2024). Liu G et al. highlighted challenges in discovering new antibiotics against *Acinetobacter baumannii* through traditional screening methods, while James J. Collins and colleagues utilized machine learning to screen approximately 7,500 molecules, swiftly identifying those inhibiting *A. baumannii* growth *in vitro* (Melo et al., 2021). Khaledi et al. predicted antimicrobial susceptibility based on genomic and transcriptomic markers, enhancing diagnostic performance by identifying resistance characteristics early in disease progression (Liu et al., 2023).

4.1.6 Application of artificial intelligence in ecology studies of pathogens

The application of artificial intelligence in ecological studies of pathogens opens up new avenues for understanding the ecological roles of microorganisms in various environments (Khaledi et al., 2020). Neural network technology, in particular, demonstrates remarkable performance in this domain. For instance, the vedoNet neural network algorithm, developed by Ananthakrishnan et al., integrates microbiome and clinical data and achieves superior classification capability for clinical remission in inflammatory bowel disease (IBD). Detailed research indicates that early trajectories of microbiome changes can serve as markers for treatment response

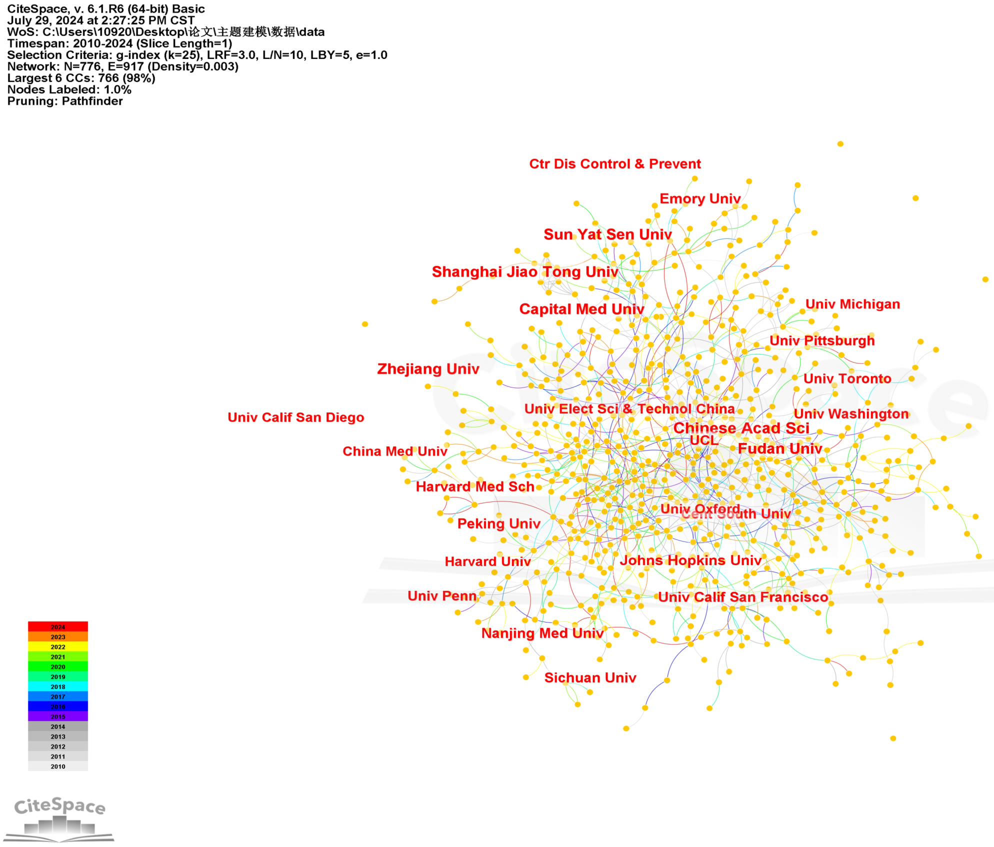


FIGURE 4
Institution co-occurrence map (node labels: by centrality).

TABLE 2 The top five institutions by number of publications and intermediate centrality.

Number of publications	Institution
486	Chinese Acad Sci
397	Fudan Univ
383	Capital Med Univ
371	Zhejiang Univ
369	Shanghai Jiao Tong Univ

BC (betweenness centrality)	Institution
0.49	Univ Med Ctr Utrecht
0.35	German Canc Res Ctr
0.28	Second Mil Med Univ
0.28	Univ Amsterdam
0.26	Chinese Acad Sci

(Lopatkin and Collins, 2020). Additionally, machine learning and data mining techniques are extensively applied to model and predict microbial community behavior under various environmental conditions, thus helping to reduce disease incidence associated with environmental changes (Ananthakrishnan et al., 2017). This interdisciplinary research not only enhances the understanding of microbial ecology but also provides a scientific foundation for formulating effective environmental management strategies.

4.1.7 Application of artificial intelligence in vaccine development

Traditional vaccine development has largely relied on laborious experimental methods that, while effective, are often time-consuming and have limited success rates (Ai et al., 2020). Recently, data mining and big data analytics have paved new pathways for vaccine development, with artificial intelligence (AI) revolutionizing the field as a tool for antigen selection and immunogen design (Brisse et al., 2020). By utilizing advanced algorithms, AI extracts crucial data from extensive genomic datasets, protein structure information, and immune system interactions, quickly identifying potential vaccine candidate antigens (Aswathy and Sumathi, 2024). For example, AI-driven neural network prediction models trained on a large dataset of over 24,000 peptides can accurately recognize key epitopes detected by the immune system. Prioritizing these epitopes and recommending experimental validation allows AI to significantly shorten the discovery time while minimizing resource investment (Olawade et al., 2024).

By integrating AI algorithms with experimental validation and clinical trials, the vaccine development process is substantially accelerated. This data-driven approach enhances vaccine development efficiency and demonstrates significant potential during global health

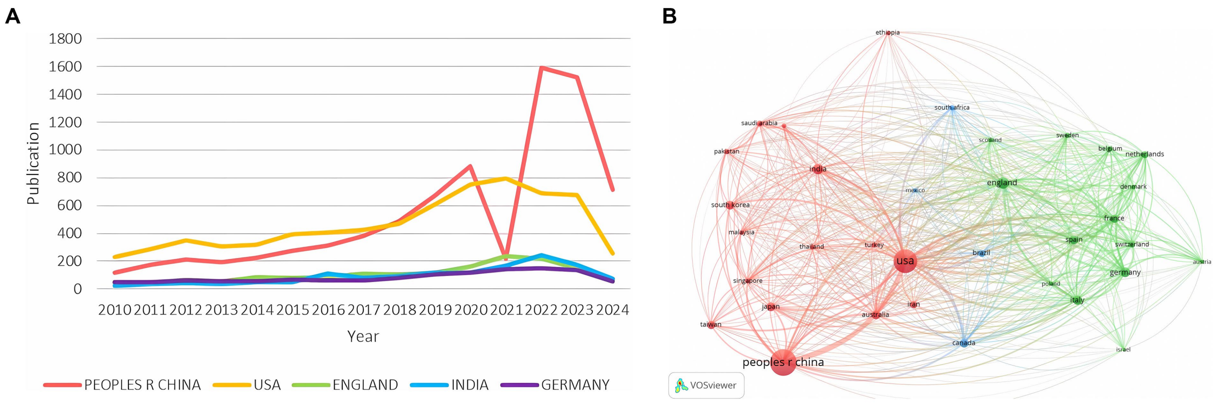


FIGURE 5
(A) Top 5 countries' production over time; **(B)** Country network visualization.

TABLE 3 Top 10 journals.

Source	Document	Citation	IF	JCR
Computers in Biology and Medicine	455	14,681	7	Q2
BMC Bioinformatics	371	10,308	2.9	Q3
Clinical Infectious Diseases	333	16,107	8.2	Q1
Computer Methods and Programs in Biomedicine	318	10,240	4.9	Q2
Diagnostics	318	2,447	3	Q3
IEEE Computational Intelligence Magazine	302	2,815	10.3	Q2
Frontiers in Oncology	272	2,546	3.5	Q3
IEEE Journal of Biomedical and Health Informatics	228	8,545	6.7	Q2
Bioinformatics	223	10,347	4.4	Q4
Medicine	212	2079	1.3	Q4

crises (Ward et al., 2021). During the COVID-19 pandemic, AI played a crucial role in quickly identifying novel antigens through detailed data mining, providing essential support for the rapid development of vaccines (Brisse et al., 2020). In mRNA-based COVID-19 vaccines, AI not only optimized vaccine sequences but also effectively screened delivery vectors, improving overall research and development efficiency (Federico et al., 2023).

4.1.8 Application of artificial intelligence in data analysis and management of pathogens

With the explosion of data volume, the application of AI in image data processing technology for pathogen detection becomes increasingly critical (Zhang et al., 2023). Traditional detection methods, such as nucleic acid and immunological assays, are often time-consuming and complex (Haymond and McCudden, 2021). Through the incorporation of machine learning, particularly deep convolutional neural network (CNN)-based image processing algorithms, AI can rapidly process and analyze microscopic image data, automatically identifying pathogens, thus significantly reducing diagnostic time. For instance, Rahman et al. utilized the DenseNet CNN model to classify 89 fungal genera from microscopic images, achieving a prediction accuracy of 65.35% (Whiley and Taylor, 2016), marking a notable enhancement in detection efficiency. Tao Chenglong integrated the HMI system with Buffer Net, developing a CNN-based AI-assisted system for rapid and automatic bacterial identification (Rahman et al., 2023). Additionally, Devan et al. employed a transfer learning method based on CNN, requiring minimal preprocessing to detect HCMV nucleocapsids in TEM images (Tao et al., 2022). In tuberculosis detection, Kuok et al. attained an 86% detection rate using a region-refined Faster R-CNN algorithm to automatically detect acid-fast bacilli on slides, significantly outperforming the traditional support vector machine (SVM) method, which had a detection rate of 70.93% (Shaga Devan et al., 2021). Chung et al. combined MALDI-TOF MS (matrix-assisted laser desorption ionization-time of flight mass spectrometry) with CNN technology for the rapid identification of hemolytic streptococci, quickly pinpointing infection sources, effectively preventing epidemic spread, and providing robust technical support for public health management (Kuok et al., 2019).

4.2 Interconnections among topics

As illustrated in Figure 7, there is significant lexical overlap among the various research topics, reflecting a strong

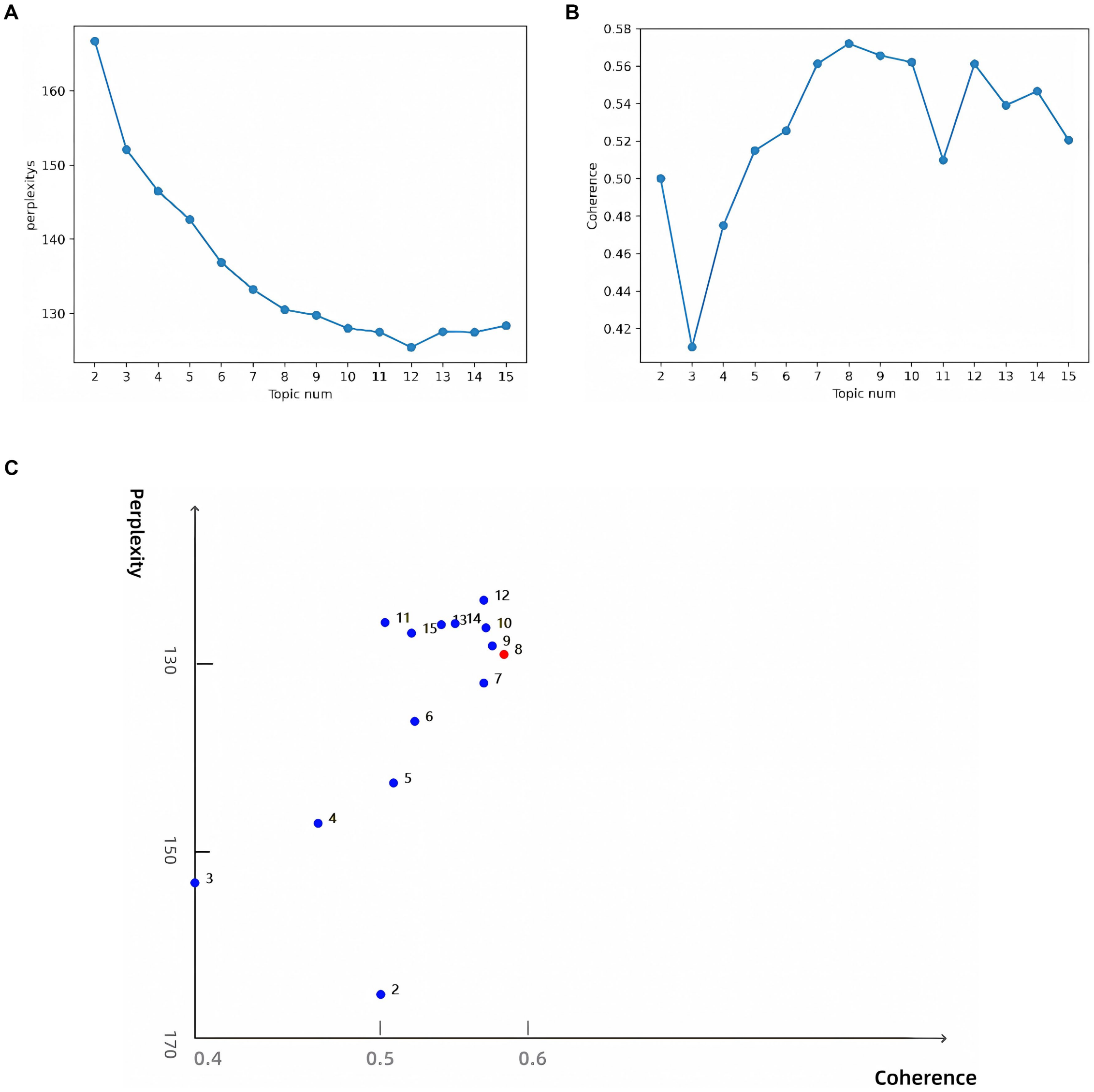


FIGURE 6
(A) Perplexity for topics 2–15; **(B)** Coherence for topics 2–15; **(C)** Topic model optimal parameter selection diagram.

interconnection and a trend towards interdisciplinary integration in the field. In the topic modeling analysis, drug resistance (Topic 1) and vaccine development (Topic 6) exhibited the highest weights (0.226 and 0.195, respectively). The growing global challenge of bacterial drug resistance and the threat from emerging infectious diseases in recent years have heightened the need for large-scale immunization efforts. AI contributes to the rapid development of vaccines by accelerating antigen identification and predicting immune responses.

Genomics research (Topic 3) and drug resistance research (Topic 1) are closely linked through shared genetic analysis methods. Genomics plays a critical role in drug resistance research; AI can swiftly analyze genomic sequencing data to identify and classify antibiotic resistance genes (Qu et al., 2019), and this genomic data can be integrated into machine learning models to predict antibiotic sensitivity and resistance phenotypes (Chung et al., 2019).

The word cloud also reveals a synergy between transmission control (Topic 2) and ecological research (Topic 5), particularly in environmental monitoring. For example, combining AI algorithms to develop predictive models can forecast high- and low-risk areas for pathogen outbreaks under future climate conditions. This approach is especially effective when linking climatology research (analyzing factors such as temperature and precipitation) with ecological studies (focusing on pathogen vectors or hosts), thereby significantly enhancing the predictive accuracy and interpretability of these models, enabling precise control and prevention (Melo et al., 2021; Farooq et al., 2022).

Data analysis and management (Topic 7) appears to be a crucial link across all research topics. Data analysis and management is not merely an independent theme but rather a key element throughout the pathogen research process. AI algorithms heavily depend on the quality of pathogen data and metadata to enhance research accuracy and reliability. From pathogen

TABLE 4 Topic-word distribution (manually screened).

Theme	Intensity	Distribution
Topic0	0.043646	Recognition, detection, microorganisms, diagnosis, sensitivity, specificity, algorithms, models, deep learning, rapid, automation, pathogens
Topic1	0.225899	Resistance, drugs, genes, sequences, mutations, evolution, prediction, analysis, antibiotic resistance, bacteria, experimental data, antibiotics
Topic2	0.082465	Transmission, infection, control, epidemic, prediction, monitoring, public health, transmission routes, risk assessment, early warning systems, spread
Topic3	0.089147	Genome, sequencing, genes, analysis, microorganisms, diversity, evolution, comparative, database, functional prediction, gene expression
Topic4	0.174416	Treatment, optimization, plans, personalized, efficacy, prediction, assessment, therapeutic outcomes, decision support, patient data, medical protocols
Topic5	0.107509	Ecology, microorganisms, environment, interactions, communities, ecosystems, analysis, monitoring, modeling, biodiversity
Topic6	0.195139	Vaccine, development, antigens, immune response, prediction, experimental data, simulation, efficacy, protection rate, bioinformatics, clinical trials
Topic7	0.08178	Data, analysis, management, databases, informatics, storage, big data, data mining, computation, statistics, automated processing

detection to predicting antibiotic resistance and optimizing treatments, substantial amounts of genomic sequencing data, electronic health records, and other clinical data are collected, processed, and analyzed, forming the training datasets for machine learning models.

4.3 Practical applications

The application of artificial intelligence in pathogen research is gradually transitioning from laboratory research to clinical practice, with official approval in certain regions. For example, the U.S. Food and Drug Administration (FDA) has approved Clever Culture Systems' APAS Compact system for the automated assessment of plates in clinical microbiology laboratories, demonstrating high sensitivity and specificity in detecting urine cultures (Peiffer-Smadja et al., 2020).

Many hospitals have already implemented AI for pathogen detection. Taiwan's Tri-Service General Hospital, along with four secondary hospitals, has successfully deployed a solution powered by an AI clinical decision support system (AI-CDSS) to expedite the detection of carbapenem-resistant *Klebsiella pneumoniae* (KP). This system integrates MALDI-TOF MS technology with machine learning algorithms, accelerating the prediction of bacterial resistance—particularly to carbapenems and colistin—by 1 day compared to traditional antibiotic susceptibility tests (AST). It provides healthcare professionals with resistance probability scores through a web interface, enabling rapid and informed treatment decisions (Ali et al., 2024). Massachusetts General Hospital employs AI to assess the risk of *Clostridium difficile* infections. In a multicenter study involving at least nine hospitals, Dascena's machine learning algorithms have been used for early sepsis detection and stratification, antimicrobial prescription recommendations, and resistant microorganism colonization predictions, demonstrating the potential to reduce hospital mortality rates, shorten hospital stays, and decrease 30-day readmission rates (Jian et al., 2024; Shimabukuro et al., 2017; Burdick et al., 2020).

AI and machine learning (ML) technologies are also extensively applied in addressing healthcare-associated infections (HAIs). AI systems are capable of predicting surgical site infections (SSIs), hospital-acquired pneumonia (HCAP), and hospital-acquired urinary tract infections (HA-UTI) (McCoy and Das, 2017). For instance, a machine learning model monitoring SSI in colon surgeries has reduced manual workload by 83.9% (Radaelli et al., 2024). A new AI-based training and monitoring system (AITMS) has improved personal protective equipment (PPE) wearing and doffing skills, successfully reducing pathogen infection rates from 1.31 to 0.58% in a Japanese hospital (Cho et al., 2024). The University of Iowa Hospitals and Clinics utilized machine learning to decrease surgical site infection rates by 74%, while Philips' "Connected Care" system reduced detection time for nosocomial infections by 87% (Huang et al., 2023).

Artificial intelligence has also played a practical role in global public health. Systems like HealthMap utilize natural language processing to analyze online news and professional resources, providing global alert information for outbreaks such as the Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Agrebi and Larbi, 2020; Ali et al., 2023). The U.S. CDC employs machine learning models to predict influenza trends (Hossain and Househ, 2016). During the COVID-19 pandemic, AI technologies were implemented in genomic classification, lineage mapping, and optimization of testing strategies. The ZOE COVID Study collected symptom data via a smartphone app, offering invaluable insights for public health (Reich et al., 2019). Singapore airport implemented thermal imaging for temperature monitoring of potential infections, combining physiological parameters with advanced analytical methods to classify high-risk influenza patients (Menni et al., 2020).

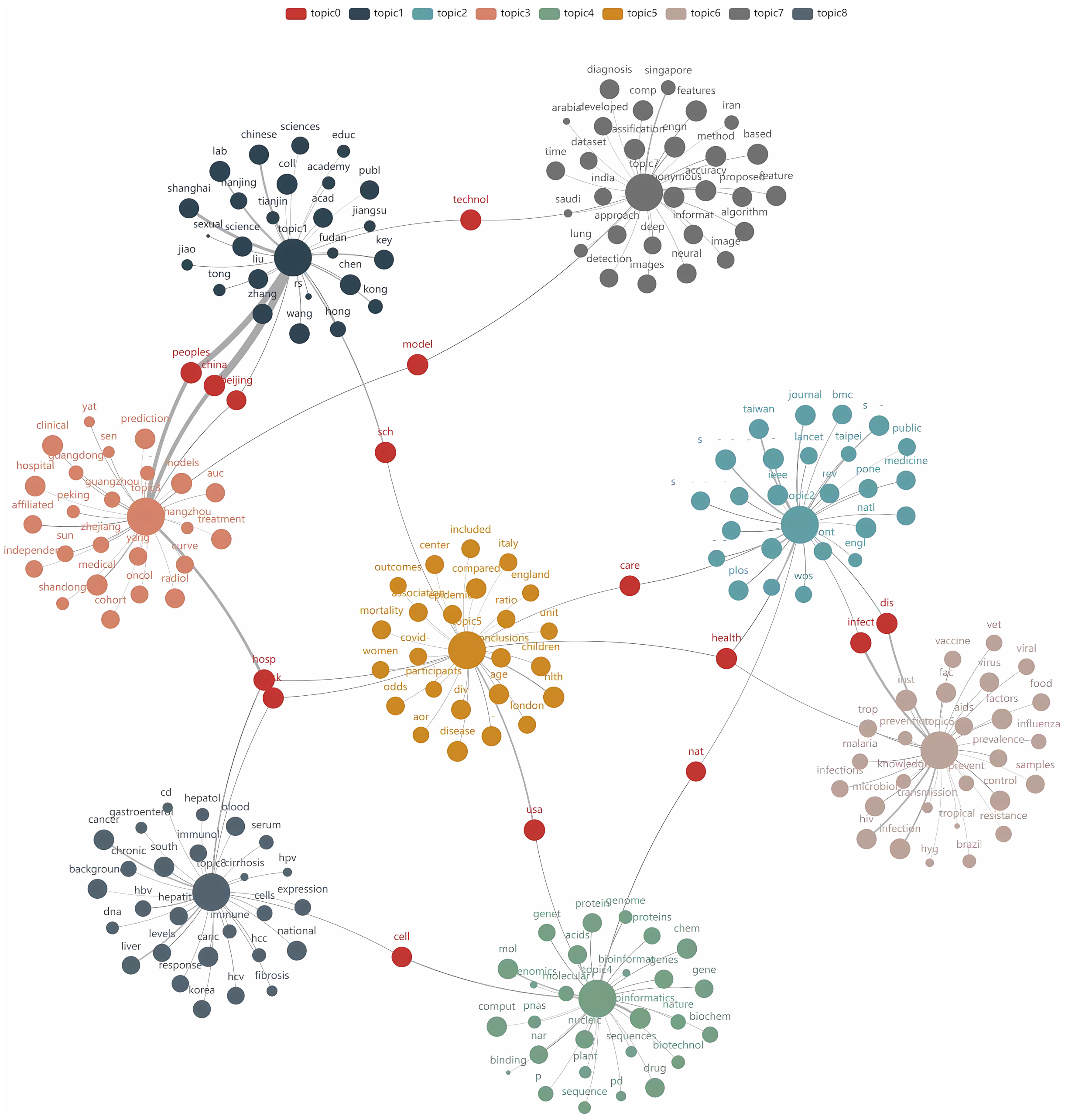


FIGURE 7
Topic-word relationship diagram.

4.4 Opportunities and challenges

The real-world application of artificial intelligence (AI) in pathogen research is still in its infancy; however, it reveals immense potential for development while facing numerous challenges and obstacles. The following are four key directions for enhancing AI application in this field:

Advanced Machine Learning Algorithms: With the increase in computational power and data accumulation, more sophisticated and accurate deep learning models can be applied to pathogen research to improve the accuracy of disease prediction and enhance the capability to handle multidimensional data.

Richer Sample Data: By collecting additional sample data from diverse clinical settings worldwide, AI systems can improve their generalization ability, thereby increasing their robustness across varied medical environments.

User-Friendly Interface Design: Developing intuitive and easy-to-use interfaces, along with providing adequate training for healthcare professionals, can significantly promote the widespread application of AI technologies in clinical practice.

Application of Extreme Value Theory: Integrating extreme value theory with robust statistical methods in epidemiology and public health can aid in the early detection of anomalies in transmission dynamics. This is particularly beneficial for the early warning of rare infectious events, such as emerging infectious diseases, providing strong support for public health interventions.

However, several challenges must be overcome to advance AI applications:

High Costs: The development, deployment, and maintenance of AI models are capital-intensive. Solutions include utilizing open-source AI tools and models and creating government subsidy policies to lower the barrier to technology access.

Training and Talent Shortage: Healthcare professionals require appropriate training to effectively use AI tools. This issue can be addressed by implementing targeted AI training programs and cultivating more medical professionals with expertise in AI.

Data Quality and Accessibility: High-quality data is crucial for training AI models. Challenges can be tackled by establishing standardized data-sharing mechanisms, improving data collection and annotation methods, and enhancing data security and privacy protections.

Ethical and Legal Issues: The use of AI in medical decision-making involves ethical and legal responsibilities. This necessitates the development of ethical guidelines and legal regulations for AI applications, clearly defining accountability and establishing effective oversight mechanisms to ensure lawful and compliant use of AI systems.

Model Explainability: The "black box" nature of AI models affects their applicability and acceptance in clinical practice. Therefore, developing more interpretable AI models can help clinicians understand their decision processes, thereby increasing trust and encouraging their use (Sun et al., 2015; Hassija et al., 2024).

5 Conclusion

In this study, we conducted a comprehensive analysis of the application of artificial intelligence (AI) in pathogenic microbiology research using bibliometrics and topic modeling. We examined 27,420 relevant publications from 2010 to 2024, uncovering an exponential growth trend in publications since 2016, primarily focused on eight key areas: pathogen detection, antibiotic resistance prediction, transmission and control, genomic analysis, therapeutic optimization, ecological studies, vaccine development, and data management systems.

The results from topic modeling indicate that the application of AI in pathogen research has become diverse and specialized. For instance, in pathogen detection, AI has significantly improved diagnostic speed and accuracy through the integration of multimodal data fusion technologies. In the realm of antibiotic resistance prediction, machine learning and deep learning models have expedited the identification and analysis of resistance genes. In vaccine development, AI has facilitated rapid progress in antigen recognition and immunogen design, thus playing a critical supportive role in the development of COVID-19 vaccines.

Despite AI's substantial potential in pathogenic microbiology research, its practical implementation remains in the early stages and faces numerous challenges. Key factors limiting effective AI application include the acquisition and sharing of high-quality data, AI system interpretability, ethical and legal responsibilities, and the high cost of development. To foster further advancements in this field, we recommend strengthening interdisciplinary collaboration to enrich AI model training data, enhancing the user-friendliness of AI tools to promote their adoption and application in clinical practice, and supporting policies to reduce the economic barriers to AI utilization. Addressing these issues collaboratively will enable a fuller realization of AI technologies in tackling challenges in the field of pathogenic microbiology, ultimately contributing to the resilience of health management and public health systems and providing unprecedented opportunities to address global public health challenges.

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

TT: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Writing – original draft. XZ: Conceptualization, Formal analysis, Investigation, Project administration, Writing – original draft. FZ: Methodology, Software, Validation, Visualization, Writing – original draft. XH: Conceptualization, Data curation, Methodology, Project administration, Software, Validation, Writing – original draft. ML: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing. ZQ: Writing – review & editing. WW: Investigation, Writing – review & editing. JL: Writing – review & editing. YW: Funding acquisition, Resources, Writing – review & editing. YL: Funding acquisition, Resources, Writing – review & editing. J-HW: Formal analysis, Funding acquisition, Resources, 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.

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

Agany, D. D. M., Pietri, J. E., and Gnimpieba, E. Z. (2020). Assessment of vector-host-pathogen relationships using data mining and machine learning. *Comput. Struct. Biotechnol. J.* 18, 1704–1721. doi: 10.1016/j.csbj.2020.06.031

Agrebi, S., and Larbi, A. (2020). "Use of artificial intelligence in infectious diseases" in Artificial intelligence in precision health. ed. D. Barh (Amsterdam, Netherlands: Elsevier), 415–438.

Ai, D., Wang, Y., Li, X., and Pan, H. (2020). Colorectal cancer prediction based on weighted gene co-expression network analysis and variational autoencoder. *Biomol. Ther.* 10:1207. doi: 10.3390/biom10091207

Ali, T., Ahmed, S., and Aslam, M. (2023). Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. *Antibiotics* 12:523. doi: 10.3390/antibiotics12030523

Ali, F., Rehman, A., and Hameed, A. (2024). "Climate change impact on plant pathogen emergence: artificial intelligence (AI) approach" in Plant quarantine challenges under climate change anxiety. eds. K. A. Abd-Elsalam and S. M. Abdel-Momen (Cham: Springer Nature Switzerland), 281–303.

Ananthakrishnan, A. N., Luo, C., Yajnik, V., Khalili, H., Garber, J. J., Stevens, B. W., et al. (2017). Gut microbiome function predicts response to anti-integrin biologic therapy in inflammatory bowel diseases. *Cell Host Microbe* 21, 603–610.e3. doi: 10.1016/j.chom.2017.04.010

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., et al. (2006). The marine viromes of four oceanic regions. *PLoS Biol.* 4:e368. doi: 10.1371/journal.pbio.0040368

Aria, M., and Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *J. Informet.* 11, 959–975. doi: 10.1016/j.joi.2017.08.007

Arruda, H., Silva, E. R., and Lessa, M. (2022). VOSviewer and Bibliometrix. *J. Med. Libr. Assoc.* 110, 392–395. doi: 10.5195/jmla.2022.1434

Aswathy, R., and Sumathi, S. (2024). The evolving landscape of cervical cancer: breakthroughs in screening and therapy through integrating biotechnology and artificial intelligence. *Mol. Biotechnol.* 1–17. doi: 10.1007/s12033-024-01124-7

Bihari, A., Tripathi, S., and Deepak, A. (2023). A review on h-index and its alternative indices. *J. Inf. Sci.* 49, 624–665. doi: 10.1177/01655515211014478

Birkle, C., Pendlebury, D. A., Schnell, J., and Adams, J. (2020). Web of science as a data source for research on scientific and scholarly activity. *Quant. Sci. Stud.* 1, 363–376. doi: 10.1162/qss_a_00018

Brisse, M., Vrba, S. M., Kirk, N., Liang, Y., and Ly, H. (2020). Emerging concepts and technologies in vaccine development. *Front. Immunol.* 11:583077. doi: 10.3389/fimmu.2020.583077

Burdick, H., Pino, E., Gabel-Comeau, D., McCoy, A., Gu, C., Roberts, J., et al. (2020). Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. *BMJ Health Care Inform.* 27:e100109. doi: 10.1136/bmjhci-2019-100109

Chauhan, U., and Shah, A. (2021). Topic modeling using latent Dirichlet allocation: a survey. *ACM Comput. Surv.* 54, 1–35. doi: 10.1145/3462478

Chen, C. (2018). Visualizing and exploring scientific literature with Citespace: an introduction. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval [New York, NY, United States: Association for Computing Machinery (ACM Press)], (pp. 369–370).

Cho, S. Y., Kim, Z., Chung, D. R., Cho, B. H., Chung, M. J., Kim, J. H., et al. (2024). Development of machine learning models for the surveillance of colon surgical site infections. *J. Hosp. Infect.* 146, 224–231. doi: 10.1016/j.jhin.2023.03.025

Chung, C. R., Wang, H. Y., Lien, F., Tseng, Y. J., Chen, C. H., Lee, T. Y., et al. (2019). Incorporating statistical test and machine intelligence into strain typing of *Staphylococcus haemolyticus* based on matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Front. Microbiol.* 10:2120. doi: 10.3389/fmicb.2019.02120

Daim, T. U., Rueda, G., Martin, H., and Gerdsri, P. (2006). Forecasting emerging technologies: use of bibliometrics and patent analysis. *Technol. Forecast. Soc. Chang.* 73, 981–1012. doi: 10.1016/j.techfore.2006.04.004

Farooq, Z., Rocklöv, J., Wallin, J., Abiri, N., Sewe, M. O., Sjödin, H., et al. (2022). Artificial intelligence to predict West Nile virus outbreaks with eco-climatic drivers. *Lancet Reg. Health Eur.* 17:100370. doi: 10.1016/j.lanepe.2022.100370

Federico, L., Malone, B., Tennøe, S., Chaban, V., Osen, J. R., Gainullin, M., et al. (2023). Experimental validation of immunogenic SARS-CoV-2 T cell epitopes identified by artificial intelligence. *Front. Immunol.* 14:1265044. doi: 10.3389/fimmu.2023.1265044

Garcia, E., Diep, J. K., Sharma, R., and Rao, G. G. (2024). Model-based learn and confirm: designing effective treatment regimens against multidrug resistant gram-negative pathogens. *Int. J. Antimicrob. Agents* 63:107100. doi: 10.1016/j.ijantimicag.2024.107100

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., and Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers.* 25, 1315–1360. doi: 10.1007/s11030-021-10217-3

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., et al. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cogn. Comput.* 16, 45–74. doi: 10.1007/s12559-023-10179-8

Haymond, S., and McCudden, C. (2021). Rise of the machines: artificial intelligence and the clinical laboratory. *J. Appl. Lab. Med.* 6, 1640–1654. doi: 10.1093/jalm/jfab075

He, S., Leanse, L. G., and Feng, Y. (2021). Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases. *Adv. Drug Deliv. Rev.* 178:113922. doi: 10.1016/j.addr.2021.113922

Hirose, Y., Zielinski, D. C., Poudel, S., Rychel, K., Baker, J. L., Toya, Y., et al. (2024). A genome-scale metabolic model of a globally disseminated hyperinvasive M1 strain of *Streptococcus pyogenes*. *mSystems* 9, e00736–e00724. doi: 10.1128/msystems.00736-24

Hossain, N., and Househ, M. (2016). "Using HealthMap to analyse Middle East respiratory syndrome (MERS) data" in Unifying the applications and foundations of biomedical and health informatics. eds. J. Mantas, A. Hasman, P. Gallos, A. Kolokathi, and M. S. Househ. (Amsterdam, Netherlands: IOS Press), 150–157.

Hu, R. S., El Hesham, A., and Zou, Q. (2022). Machine learning and its applications for protozoal pathogens and protozoal infectious diseases. *Front. Cell. Infect. Microbiol.* 12:882995. doi: 10.3389/fcimb.2022.882995

Huang, T., Ma, Y., Li, S., Ran, J., Xu, Y., Asakawa, T., et al. (2023). Effectiveness of an artificial intelligence-based training and monitoring system in prevention of nosocomial infections: a pilot study of hospital-based data. *Drug Discov. Ther.* 17, 351–356. doi: 10.5582/ddt.2023.01068

Huo, D., and Wang, X. (2024). A new era in healthcare: the integration of artificial intelligence and microbial. *Med. Nov. Technol. Devices* 23:100319. doi: 10.1016/j.medntd.2024.100319

Jelodar, H., Wang, Y., and Yuan, C. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimed. Tools Appl.* 78, 15169–15211. doi: 10.1007/s11042-018-6894-4

Jian, M. J., Lin, T. H., and Chung, H. Y. (2024). Artificial intelligence-clinical decision support system in infectious disease control: combatting multidrug-resistant *Klebsiella pneumoniae* with machine learning. *Infect. Drug Resist.* 17, 2899–2912. doi: 10.2147/IDR.S470821

Jimeno-Yepes, A. J., and Verspoor, K. (2023). Classifying literature mentions of biological pathogens as experimentally studied using natural language processing. *J. Biomed. Semantics* 14:1. doi: 10.1186/s13326-023-00282-y

Kaur, I., Sandhu, A. K., and Kumar, Y. (2022). Artificial intelligence techniques for predictive modeling of vector-borne diseases and its pathogens: a systematic review. *Arch. Comput. Methods Eng.* 29, 3741–3771. doi: 10.1007/s11831-022-09724-9

Khaledi, A., Weimann, A., Schniederjans, M., Asgari, E., Kuo, T. H., Oliver, A., et al. (2020). Predicting antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular diagnostics. *EMBO Mol. Med.* 12:e10264. doi: 10.15252/emmm.201910264

Khan, Z. A., Siddiqui, M. F., and Park, S. (2019). Current and emerging methods of antibiotic susceptibility testing. *Diagnostics* 9:49. doi: 10.3390/diagnostics9020049

Kuok, C. P., Horng, M. H., Liao, Y. M., Chow, N. H., and Sun, Y. N. (2019). An effective and accurate identification system of *Mycobacterium tuberculosis* using convolutional neural networks. *Microsc. Res. Tech.* 82, 709–719. doi: 10.1002/jemt.23217

Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., and Ettema, T. J. G. (2021). Innovations to culturing the uncultured microbial majority. *Nat. Rev. Microbiol.* 19, 225–240. doi: 10.1038/s41579-020-00458-8

Liu, G., Catacutan, D. B., Rathod, K., Swanson, K., Jin, W., Mohammed, J. C., et al. (2023). Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nat. Chem. Biol.* 19, 1342–1350. doi: 10.1038/s41589-023-01349-8

Lopatkin, A. J., and Collins, J. J. (2020). Predictive biology: modelling, understanding and harnessing microbial complexity. *Nat. Rev. Microbiol.* 18, 507–520. doi: 10.1038/s41579-020-0372-5

Májek, P., Lüftinger, L., Beisken, S., Rattei, T., and Materna, A. (2021). Genome-wide mutation scoring for machine-learning-based antimicrobial resistance prediction. *Int. J. Mol. Sci.* 22:13049. doi: 10.3390/ijms222313049

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402. doi: 10.1146/annurev.genom.9.081307.164359

Matias, A. V., Amorim, J. G. A., Macarini, L. A. B., Cerentini, A., Onofre, A. S. C., Onofre, F. B. D. M., et al. (2021). What is the state of the art of computer vision-assisted cytology? A systematic literature review. *Comput. Med. Imaging Graph.* 91:101934. doi: 10.1016/j.compmedimag.2021.101934

McCoy, A., and Das, R. (2017). Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual.* 6:e000158. doi: 10.1136/bmjoq-2017-000158

Melo, M. C. R., Maasch, J. R. M. A., and de la Fuente-Nunez, C. (2021). Accelerating antibiotic discovery through artificial intelligence. *Commun. Biol.* 4:1050. doi: 10.1038/s42003-021-02586-0

Menni, C., Valdes, A. M., Freidin, M. B., Sudre, C. H., Nguyen, L. H., Drew, D. A., et al. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* 26, 1037–1040. doi: 10.1038/s41591-020-0916-2

Ninkov, A., Frank, J. R., and Maggio, L. A. (2022). Bibliometrics: methods for studying academic publishing. *Perspect. Med. Educ.* 11, 173–176. doi: 10.1007/s40037-021-00695-4

Nourani, E., Khunjush, F., and Durmuş, S. (2015). Computational approaches for prediction of pathogen-host protein-protein interactions. *Front. Microbiol.* 6:94. doi: 10.3389/fmicb.2015.00094

Olawade, D. B., Teke, J., Fapohunda, O., Weerasinghe, K., Usman, S. O., Ige, A. O., et al. (2024). Leveraging artificial intelligence in vaccine development: a narrative review. *J. Microbiol. Methods* 224:106998. doi: 10.1016/j.mimet.2024.106998

Park, S. J., Saito-Adachi, M., Komiyama, Y., and Nakai, K. (2016). Advances, practice, and clinical perspectives in high-throughput sequencing. *Oral Dis.* 22, 353–364. doi: 10.1111/odi.12403

Peiffer-Smadja, N., Dellière, S., Rodriguez, C., Birgand, G., Lescure, F. X., Fourati, S., et al. (2020). Machine learning in the clinical microbiology laboratory: has the time come for routine practice? *Clin. Microbiol. Infect.* 26, 1300–1309. doi: 10.1016/j.cmi.2020.02.006

Pillai, N., Ramkumar, M., and Nanduri, B. (2022). Artificial intelligence models for zoonotic pathogens: a survey. *Microorganisms* 10:1911. doi: 10.3390/microorganisms10101911

Qu, K., Guo, F., Liu, X., Lin, Y., and Zou, Q. (2019). Application of machine learning in microbiology. *Front. Microbiol.* 10:827. doi: 10.3389/fmicb.2019.00827

Radaelli, D., Di Maria, S., and Jakovski, Z. (2024). Advancing patient safety: the future of artificial intelligence in mitigating healthcare-associated infections: a systematic review. *Healthcare* 12:1996. doi: 10.3390/healthcare12191996

Rahman, M. A., Clinch, M., Reynolds, J., Dangott, B., Meza Villegas, D. M., Nassar, A., et al. (2023). Classification of fungal genera from microscopic images using artificial intelligence. *J. Pathol. Inform.* 14:100314. doi: 10.1016/j.jpi.2023.100314

Reich, N. G., McGowan, C. J., Yamana, T. K., Tushar, A., Ray, E. L., Osthus, D., et al. (2019). Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the US. *PLoS Comput. Biol.* 15:e1007486. doi: 10.1371/journal.pcbi.1007486

Ren, J., Liu, M., Liu, Y., and Liu, J. (2023). TransCode: uncovering COVID-19 transmission patterns via deep learning. *Infect. Dis. Poverty* 12:14. doi: 10.1186/s40249-023-01052-9

Rondon-Villarreal, P., Sierra, D., and Torres, R. (2014). Machine learning in the rational design of antimicrobial peptides. *Curr. Comput. Aided Drug Des.* 10, 183–190. doi: 10.2174/1573409910666140624124807

Saha, M., and Sarkar, A. (2021). Review on multiple facets of drug resistance: a rising challenge in the 21st century. *J. Xenobiot.* 11, 197–214. doi: 10.3390/jox11040013

Shaga Devan, K., Walther, P., von Einem, J., Ropinski, T., Kestler, H., and Read, C. (2021). Improved automatic detection of herpesvirus secondary envelopment stages in electron microscopy by augmenting training data with synthetic labelled images generated by a generative adversarial network. *Cell. Microbiol.* 23:e13280. doi: 10.1111/cmi.13280

Sheetal Ambardar, S. A., Rikita Gupta, R. G., Deepika Trakroo, D. T., Lal, R., and Vakhlu, J. (2016). High throughput sequencing: an overview of sequencing chemistry. *J. Appl. Microbiol.* 56, 394–404. doi: 10.1007/s12088-016-0606-4

Shimabukuro, D. W., Barton, C. W., Feldman, M. D., Mataraso, S. J., and das, R. (2017). Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir. Res.* 4:e000234. doi: 10.1136/bmjresp-2017-000234

Smith, K. P., Kang, A. D., and Kirby, J. E. (2018). Automated interpretation of blood culture gram stains by use of a deep convolutional neural network. *J. Clin. Microbiol.* 56, e01521–e01517. doi: 10.1128/jcm.01521-17

Sun, G., Matsui, T., and Hakozaki, Y. (2015). An infectious disease/fever screening radar system which stratifies higher-risk patients within ten seconds using a neural network and the fuzzy grouping method. *J. Infect.* 70, 230–236. doi: 10.1016/j.jinf.2014.12.007

Sun, T., Niu, X., He, Q., Chen, F., and Qi, R. Q. (2023). Artificial intelligence in microbiomes analysis: a review of applications in dermatology. *Front. Microbiol.* 14:1112010. doi: 10.3389/fmicb.2023.1112010

Tao, C., Du, J., Tang, Y., Wang, J., Dong, K., Yang, M., et al. (2022). A deep-learning-based system for rapid genus identification of pathogens under hyperspectral microscopic images. *Cells* 11:2237. doi: 10.3390/cells11142237

Theodosiou, A. A., and Read, R. C. (2023). Artificial intelligence, machine learning and deep learning: potential resources for the infection clinician. *J. Infect.* 87, 287–294. doi: 10.1016/j.jinf.2023.07.006

Tynecki, P., Guziński, A., Kazimierczak, J., Jadczuk, M., Dastych, J., and Onisko, A. (2020). PhageAI-bacteriophage life cycle recognition with machine learning and natural language processing. *bioRxiv*:2020.07.11.198606. doi: 10.1101/2020.07.11.198606

Uddin, T. M., Chakraborty, A. J., Khusro, A., Zidan, B. M. R. M., Mitra, S., Emran, T. B., et al. (2021). Antibiotic resistance in microbes: history, mechanisms, therapeutic strategies and future prospects. *J. Infect. Public Health* 14, 1750–1766. doi: 10.1016/j.jiph.2021.10.020

Vahedi, B., Karimzadeh, M., and Zoraghein, H. (2021). Spatiotemporal prediction of COVID-19 cases using inter-and intra-county proxies of human interactions. *Nat. Commun.* 12:6440. doi: 10.1038/s41467-021-26742-6

Vayansky, I., and Kumar, S. A. P. (2020). A review of topic modeling methods. *Inf. Syst.* 94:101582. doi: 10.1016/j.is.2020.101582

Wang, D., Lang, J. C., and Chen, Y. H. (2024). Assessment of using Google trends for real-time monitoring of infectious disease outbreaks: a measles case study. *Sci. Rep.* 14:9470. doi: 10.1038/s41598-024-60120-8

Wani, A. K., Roy, P., Kumar, V., and Mir, T. G. (2022). Metagenomics and artificial intelligence in the context of human health. *Infect. Genet. Evol.* 100:105267. doi: 10.1016/j.meegid.2022.105267

Ward, D., Higgins, M., Phelan, J. E., Hibberd, M. L., Campino, S., and Clark, T. G. (2021). An integrated in silico immuno-genetic analytical platform provides insights into COVID-19 serological and vaccine targets. *Genome Med.* 13, 4–12. doi: 10.1186/s13073-020-00822-6

Weis, C., Cuénod, A., Rieck, B., Dubuis, O., Graf, S., Lang, C., et al. (2022). Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat. Med.* 28, 164–174. doi: 10.1038/s41591-021-01619-9

Whiley, H., and Taylor, M. (2016). Legionella detection by culture and qPCR: comparing apples and oranges. *Crit. Rev. Microbiol.* 42, 65–74. doi: 10.3109/1040841X.2014.885930

Wong, F., de la Fuente-Nunez, C., and Collins, J. J. (2023). Leveraging artificial intelligence in the fight against infectious diseases. *Science* 381, 164–170. doi: 10.1126/science.adh1114

Xu, X., Gao, Z., and Yang, F. (2020). Antidiabetic effects of Gegen Qinlian decoction via the gut microbiota are attributable to its key ingredient berberine. *Genomics Proteomics Bioinformatics* 18, 721–736. doi: 10.1016/j.gpb.2019.09.007

Yang, Y., Ngai, E. W. T., and Wang, L. (2024). Resistance to artificial intelligence in healthcare: literature review, conceptual framework, and research agenda. *Inf. Manag.* 61:103961. doi: 10.1016/j.im.2024.103961

Yu, H., Jing, W., and Iriya, R. (2018). Phenotypic antimicrobial susceptibility testing with deep learning video microscopy. *Anal. Chem.* 90, 6314–6322. doi: 10.1021/acs.analchem.8b01128

Zhang, T., Chen, J., and Lu, Y. (2022). Identification of technology frontiers of artificial intelligence-assisted pathology based on patent citation network. *PLoS One* 17:e0273355. doi: 10.1371/journal.pone.0273355

Zhang, Y., Li, R., Zou, G., Guo, Y., Wu, R., Zhou, Y., et al. (2024). Discovery of antimicrobial lysins from the "dark matter" of uncharacterized phages using artificial intelligence. *Adv. Sci.* 11:e2404049. doi: 10.1002/advs.202404049

Zhang, H., Zhang, L., Lin, A., Xu, C., Li, Z., Liu, K., et al. (2023). Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* 621, 396–403. doi: 10.1038/s41586-023-06127-z

Zhao, J., Han, M., Ma, A., Jiang, F., Chen, R., Dong, Y., et al. (2024). A machine vision-assisted Argonaute-mediated fluorescence biosensor for the detection of viable Salmonella in food without convoluted DNA extraction and amplification procedures. *J. Hazard. Mater.* 466:133648. doi: 10.1016/j.jhazmat.2024.133648

Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* 12, 931–934. doi: 10.1038/nmeth.3547

Frontiers in Microbiology

TYPE Original Research
PUBLISHED 09 December 2024
DOI 10.3389/fmicb.2024.1519871

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Mengting Liu,
Sun Yat-sen University, China
Xia Hao,
Shandong Agricultural University, China
Tong Li,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Xiaodan Sui
✉ xiaodan.sui@qq.com

RECEIVED 30 October 2024
ACCEPTED 21 November 2024
PUBLISHED 09 December 2024

CITATION
Chang S, Yang T, Yin B, Zhang J, Ma L, Ding Y and Sui X (2024) Deformable multi-level feature network applied to nucleus segmentation. *Front. Microbiol.* 15:1519871. doi: 10.3389/fmicb.2024.1519871

Deformable multi-level feature network applied to nucleus segmentation

Shulei Chang, Tingting Yang, Bowen Yin, Jiayi Zhang, Liang Ma, Yanhui Ding and Xiaodan Sui*

School of Information Science and Engineering, Shandong Normal University, Jinan, China

Introduction: The nucleus plays a crucial role in medical diagnosis, and accurate nucleus segmentation is essential for disease assessment. However, existing methods have limitations in handling the diversity of nuclei and differences in staining conditions, restricting their practical application.

Methods: A novel deformable multi-level feature network (DMFNet) is proposed for nucleus segmentation. This network is based on convolutional neural network and divides feature processing and mask generation into two levels. At the feature level, deformable convolution is used to enhance feature extraction ability, and multi-scale features are integrated through a balanced feature pyramid. At the mask level, a one-stage framework is adopted to directly perform instance segmentation based on location.

Results: Experimental results on the MoNuSeg 2018 dataset show that the mean average precision (mAP) and mean average recall (mAR) of DMFNet reach 37.8% and 47.4% respectively, outperforming many current advanced methods. Ablation experiments verified the effectiveness of each module of the network.

Discussion: DMFNet provides an effective solution for nucleus segmentation and has important application value in medical image analysis.

KEYWORDS

nucleus segmentation, pathology images, deep learning, convolutional neural network, deformable multi-level feature network

1 Introduction

The nucleus plays an important role in the examination of hematoxylin and eosin stained tissue sections.Nuclear morphometric features and appearance, including the color of the surrounding cytoplasm, also help in identifying various types of cells, such as epithelial (glandular), stromal, or inflammatory cells, which in turn, provide an understanding of the glandular structure and disease presentation at low power (Kumar et al., 2017). In disease diagnosis, nuclear characteristics are key indicators. For example, abnormal morphological and structural changes in cancer cell nuclei, such as nuclear enlargement and nuclear-cytoplasmic ratio imbalance, can assist doctors in determining the type and stage of cancer. Moreover, nuclear segmentation can also contribute to pathological research by enabling the understanding of cellular level changes during the development of diseases.Therefore, accurate nucleus segmentation is critical in the field of medicine.

Owing to the importance of nuclear information in medicine, numerous researchers have proposed pathology image segmentation methods (Xinpeng et al., 2020; Liu et al., 2021), including level sets (Peifang et al., 2016), graphbased segmentation (Fuyong and Lin, 2016), mathematical morphologies (Wang et al., 2016), and pixel classification (Liu et al., 2019). However, such methods fail to generalize across a wide spectrum of tissue morphologies due to inter- and intra-nuclear color variations in crowded and chromatin sse nuclei (Kumar et al., 2017). Traditional methods face numerous limitations, and under these circumstances, machine learning techniques have gradually become a new hope for solving the nuclear segmentation problem due to their unique advantages.Techniques based on machine learning can provide better results for challenging cases of nucleus segmentation because they can be trained to recognize nucleus shapes and color variations (Yiming et al., 2018; Xieli et al., 2019).

However, automatic nucleus segmentation continues to be very challenging owing to variations in the nuclei, ambiguous borders, and differences in staining conditions. Nucleus segmentation tasks are challenging in three respects. Firstly, the shapes of the nuclei in the pathology images vary in shapes; however, the convolution kernel of the convolutional neural network (CNN) modules is a fixed geometric structure. In other words, the CNN modules do not possess the internal mechanism to handle nuclei of different shapes. Secondly, deep high-level features in the backbones have more semantic meanings, while shallow low-level features are more content descriptive (Zeiler and Rob, 2014). For example, high-level information can provide many semantic details, like staining conditions. Low-level information can provide content, such as the location of the nucleus. Thirdly, most instance segmentation methods based on CNN comprise two stages, which are complex and have room for improvement in accuracy.

Over the past dozen years, deep learning has emerged as a prominent category of machine learning algorithms, including natural language processing, computer vision, and more. One of the most representative models in deep learning models is CNN. In computer vision, different locations on images may correspond to objects with different scales or deformation (Jifeng et al., 2017); for example, fully convolutional networks (FCNs) (Jonathan et al., 2015) provide semantic segmentation with the ability of adaptive determination of scales or receptive field sizes for visual recognition tasks with fine localization; however, their performance warrants further improvements. Feature integration has led to the development of instance segmentation. FPN (Tsung-Yi et al., 2017) and PANet (Shu et al., 2018) integrate features via lateral connections to achieve excellent performance; however, they cannot merge shallow and deep information with each other. AdaptIS (Konstantin et al., 2019) predicts point proposals for classagnostic instance segmentation, and then generates a mask for the object located at this point. PolarMask (Xie et al., 2020) uses instance center classification and dense distance regression in a polar coordinate system to predict the contour of instances. These methods may be considered as a semidirect adigm. They are anchor-free and make the CNN simple; however, all of them require additional complex processing methods. ooTensorMask (Xinlei et al., 2019) operates in a dense sliding window and segments objects in fixed local patches, limited by patch scale. SOLO aims to segment instance masks directly, under the supervision of full instance mask annotations rather than in-box masks or additional pixel pairwise relations (Xinlong et al., 2020).

Recently, the most common instance segmentation method is the two-stage method. It has two approaches. The first one is “detect then segment”. It first detects the target to create bounding boxes and then divides the mask in each box. The second one is “label then cluster”. First, each pixel is predicted, and then the pixels of the same instance are grouped together.This approach is usually not as effective as the first approach. A typical example of a two-stage approach is Mask R-CNN, which uses a region proposal network (RPN) (Shaoqing et al., 2017) to obtain and classify candidate regions, which are then segmented using an FCN (Jonathan et al., 2015) model. Two stage methods achieve both step-wise and indirect object localization and mask generation, which either rely heavily on bounding box detection or clustering. On the contrary, one-stage instance segmentation methods can simultaneously achieve object localization and mask generation. SOLO (Xinlong et al., 2020) is one of the representative methods of one-stage instance segmentation methods, which takes an image as input and directly outputs instance masks and corresponding class probabilities, using a fully convolutional, frameless and groupless adigm.

Faisal et al. (2020) enhanced structured prediction capabilities for nucleus segmentation through conditional generative adversarial networks trained with synthetic and real data. Peter et al. (2019) formulated the nuclear segmentation task as the regression of intra-nuclei map distance to solve the joint segmentation of close nuclei. Similar to the nucleus segmentation task, Hao et al. (2016) proposed a deep contouraware network integrating multiple layers of contextual features to accurately segment glands from pathological images. Carsen et al. (2021) proposed a segmentation method called Cellpose that can accurately segment cells from various image types, with exciting results. For better generating bounding box proposals, Jingru et al. (2019) proposed a keypoint-based detector combined with cell instance segmentation. Oskar et al. (2019) segmented nuclei based on Mask R-CNN and used bounding boxes to detect nuclei instances. However, the shape of the nucleus tends to be oval, which presents an occlusal problem. This means that each bounding box may contain pixels representing two or more instances, which suggests that the bounding box may end up being suboptimal for kernel segmentation (Shengcong et al., 2020). Ortiz et al. (2020) proposed an instance segmentation method based on a recurrent residual network, which offers the advantages of improved segmentation accuracy and enhanced feature propagation stability. However, the method has some drawbacks, including high computational cost and training time, limited flexibility when handling complex scenarios, and sensitivity to the quality of input data. In recent years, methods based on Transformer have gradually emerged. Chen et al. (2021) combined the Transformer encoder with the U-Net architecture for medical image segmentation, which is especially suitable for medical image segmentation tasks. This method can effectively capture long-distance global dependencies and improve the segmentation accuracy. However, the computational resource consumption of Transformer is relatively large, resulting in long training time and high memory requirements. Cao et al. (2022) enhanced the segmentation performance of medical images by introducing Swin Transformer and utilizing its multi-scale characteristics, especially performing prominently in medical image segmentation tasks. However, when dealing with very large or complex images, the local windowing method of Swin Transformer may limit the ability to extract global information. He et al. (2023) enhanced the model's ability by introducing convolution operations in SwinUNETR. However, the training process may be relatively complex, involving a variety of techniques and adjustments, which may increase the implementation difficulty in practical applications.

In this study, we introduce a novel method called the deformable multi-level feature network (DMFNet) for nucleus segmentation. The DMFNet is based on a CNN using images of H&E stained tissue specimens. The DMFNet employs two levels to process the features and masks separately. To address the three challenges mentioned earlier, the DMFNet effectively combines deformable convolutional networks (DCNs) (Jifeng et al., 2017), balanced feature pyramid (BFP) (Jiangmiao et al., 2019), and segmenting objects by locations (SOLO) (Xinlong et al., 2020). Even though each of these components has been used in the past, we demonstrate that their combination in nuclei segmentation is superior to the existing standard methods. Thus, the main contributions of this study are as follows:

First, we use a novel module to replace the feature extraction module of a conventional CNN for dense spatial transformations; this can increase the transformation modeling capability. In the new module, the convolution kernel can be in various forms of deformation for free sampling.

Second, we integrate multi-level features, which are rescaled, integrated, refined, and strengthened to obtain balanced semantic features and refine the same. Finally, we use a one-stage network to directly distinguish instances by the center locations and object sizes instead of masks in boxes or pixel-pairwise relations. Nuclei segmentation by location renders the segment framework simple, and flexible. The DMFNet achieved the best performance on the MoNuSeg 2018 dataset, with its mAP and mAR approaching 37.8% and 47.4%, respectively.

2 Materials and methods

SOLO is a one-stage algorithm, and its backbone comprises a residual network (ResNet) (Kaiming et al., 2016) and FPN (Tsung-Yi et al., 2017). To mitigate the complex structure in the two-stage methods, our DMFNet is based on SOLO for nucleus segmentation. The proposed model consists of a feature level and a mask level, as shown in Figure 1. Specifically, the input image is first extracted with features in ResNet, which incorporates a deformable convolution (Jifeng et al., 2017). Then, the features are integrated into a balanced feature pyramid to obtain enhanced multi-level semantic features. Finally, a mask generation network predicts categories and generates instance masks simultaneously, thus achieving an effective instance segmentation under a one-stage network.

2.1 Feature extraction

At the feature level, a deformable residual network is trained for an efficient feature extraction. The standard convolution consists of the following two steps: (1) use a regular grid R for sampling on the input feature map x; and (2) perform a weighting operation. For example,

$$R = (-1, -1), (-1, 0), ..., (0, 1), (1, 1) \tag{1}$$

where R defines the size and dilation. Here, it defines a 3 × 3 kernel with a dilation of 1. Each position p_0 on the output feature map y, is calculated using the following formula:

$$y(p_0) = \sum_{p_n \in \mathbb{R}} w(p_n) \cdot x(p_0 + p_n) \tag{2}$$

where w is the weight of the sampled values, and p_n is an enumeration of the locations listed in R. In this network, the regular grid R is expanded by adding offsets, where $N = |R|$. The same position p_0 becomes:

$$y(p_0) = \sum_{n \in \mathbb{R}} w(p_n) \cdot x(p_0 + p_n + \delta p_n) \tag{3}$$

Now, the sampling location has become an irregular location. As the offset δp_n is usually a decimal number, Equation (3) is implemented via a bilinear interpolation, shown in Equation (4) below. Here, p defines an arbitrary location and q is an enumeration of all the integral spatial locations listed in feature map.

$$x(p) = \sum_{q} G(p, q) \cdot x(q) \tag{4}$$

where $G(.,.)$ denotes a bilinear interpolation kernel. It is divided into two one-dimensional kernels as follows:

$$G(p, q) = g(q_x, p_x) \cdot g(q_y, p_y) \tag{5}$$

where $g(a, b) = \max(0, 1 - |a - b|)$.

As illustrated in Figure 2, the offsets are obtained by applying a convolutional layer over the same input feature map. To learn the offsets, the gradients are back propagated using Equations 4, 5. The deformable network is integrated with the state-ofthe-art architecture ResNet to enhance the capability of the DMFNet for modeling the transformations. This possesses excellent feature extraction capability for nuclei of various shapes.

2.2 Feature integration

Feature integration occupies a crucial position in the field of deep learning. It focuses on merging and summarizing the feature information obtained from different network layers, diverse functional modules, or various feature extraction methods, ultimately creating a more comprehensive expression that accurately depicts the target features. This technique plays a

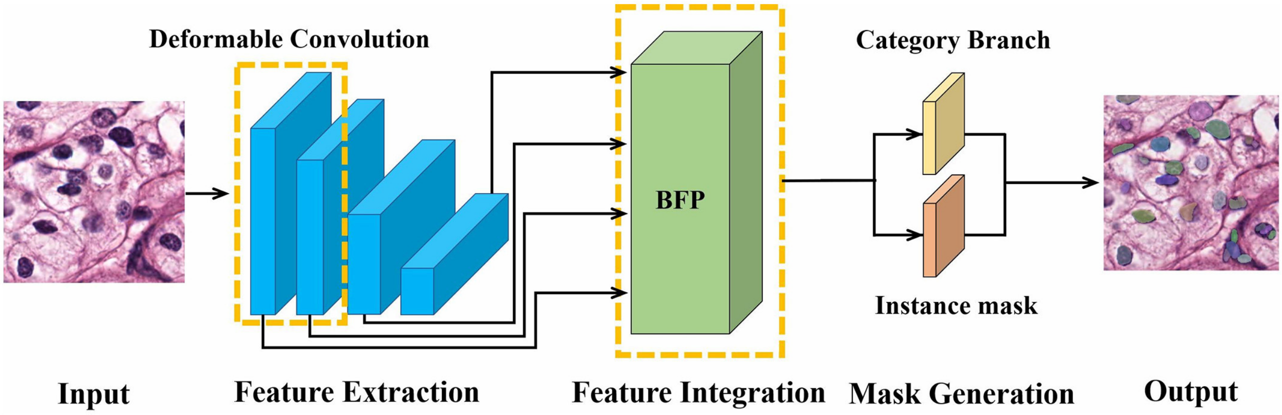


FIGURE 1
Illustration of our proposed DMFNet architecture.

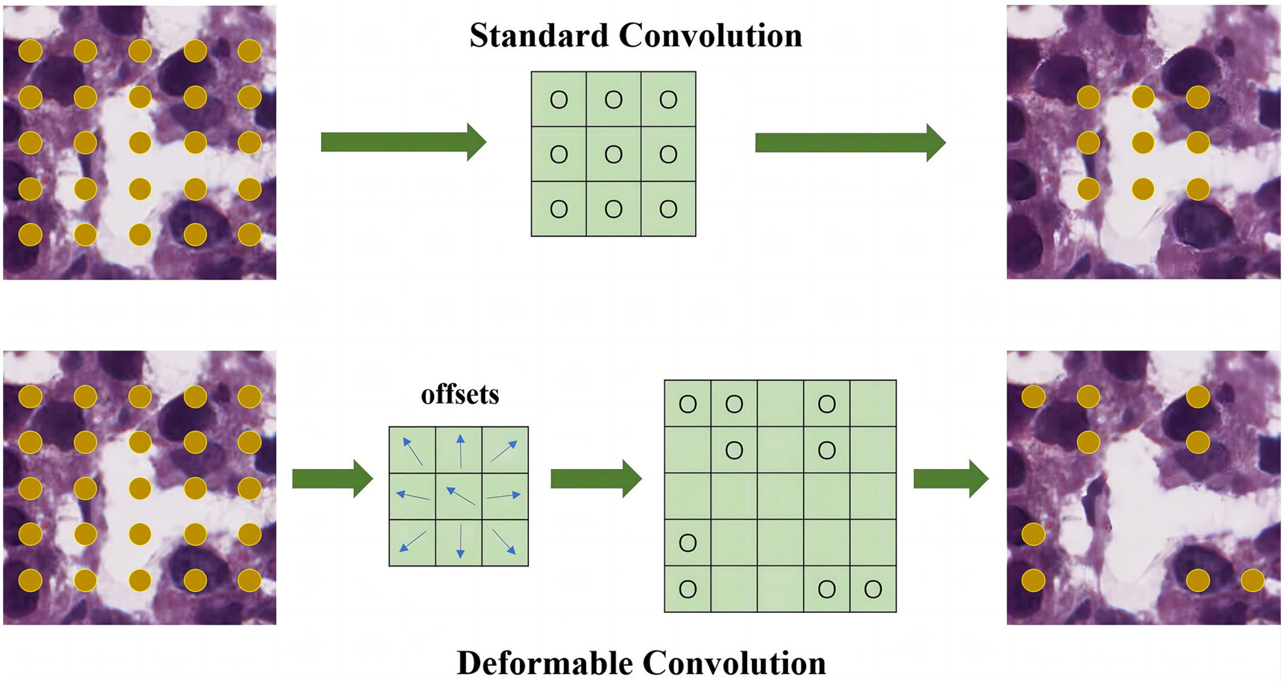


FIGURE 2
Feature extraction module for feature level. The upper part is the standard convolution, and the lower part is the deformable convolution.

significant role in enhancing the model's ability to understand and process complex data. The Balanced Feature Pyramid is an innovative structure specifically designed to optimize the feature integration process. In the ongoing evolution of deep learning models, while traditional feature pyramid networks are capable of capturing multi-scale features, they often face the challenge of imbalanced information distribution when merging features from different levels. For example, in tasks such as cell nucleus segmentation, this imbalance can lead to inaccurate and incomplete descriptions of nuclear features. In response to this challenge, the Balanced Feature Pyramid was developed, with its core mission being to address this issue. Through a series of unique designs and operations, it makes the feature integration process more efficient and precise, thus providing a higher-quality feature foundation for subsequent tasks, such as cell nucleus segmentation.

Next, we utilize the balanced feature pyramid in our DMFNet to strengthen the multi-level features. The essential purpose of this module is to strengthen the multi-level features using the same deeply integrated balanced semantic features. It consists of four steps, namely rescaling, integrating, refining, and strengthening (Jiangmiao et al., 2019). The structure of this module is shown in Figure 3. To obtain balanced semantic features, we first resize the

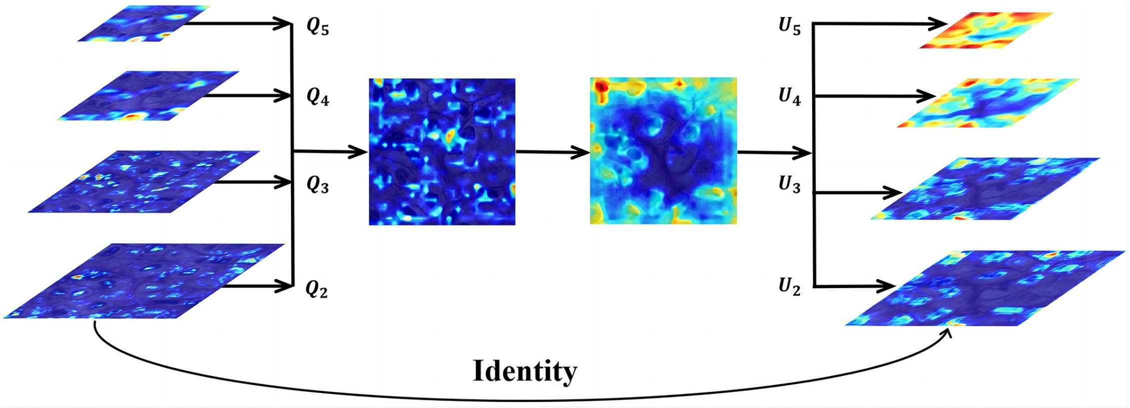


FIGURE 3
Feature integration module for feature level.Q_2, Q_3, Q_4, and Q_5 are multi-level features, and U_2, U_3, U_4, and U_5 are multi-level output signals.

multi-level features $\{Q_2, Q_3, Q_4, Q_5\}$ which have been generated by the FPN to the same size as Q_4, and merge them to obtain Q_{int} with interpolation and max-pooling. The balanced semantic features can be expressed as:

$$Q_{int} = \frac{1}{L}\sum_{l_{min}}^{l_{max}} Q_l \tag{6}$$

where L is the number of multi-level features; Q_l is the feature of resolution level l; and l_{min} and l_{max} are subtables representing the lowest and highest level indices, respectively.At this time, each resolution obtains equivalent information from the other resolutions.

Next, we use a non-local module (Xiaolong et al., 2018) to refine the balanced semantic features for more distinguishing features and better results. Non-local operations in deep neural networks are represented as:

$$U_i = \frac{1}{C(v)} \sum_{\forall j} f(v_i, v_j) g(v_j) \tag{7}$$

where i and j represent the indices of the output position and all possible associated positions, respectively; v indicates the input signal; U indicates the output signal, with the same size as v; and f (v_i, v_j) calculates the scalar between i and j. For example, the farther the distance between the positions of i and j, the smaller the value of the pairwise function f, which means that the position of j has less influence on i. $g(v_j)$ calculates the representation of the input signal at position j and $C(v)$ is the normalization parameter.

Finally, the refined features of the four levels are added to the original features through interpolation or pooling to enhance the original features. We effectively create the semantic features of different layers using the BFP. This offers a better accuracy for nucleus segmentation.

2.3 Mask generation

At the mask level, we further process the category prediction and instance mask generation. The pipeline for the same is shown in Figure 4.

We divide the picture into an $S \times S$ grid. The network output is divided into two branches, namely classification and mask branches. Simultaneous with the category prediction, each grid generates a corresponding instance mask. The size of the classification branch is $S \times S \times C$, where C is the number of categories. The mask branch size is $H \times W \times S^2$, where S^2 is the maximum number of instances predicted. When the center of the target object falls in the grid, the corresponding position of the classification branch and corresponding channel of the mask branch are responsible for the prediction of the object. For example, if the instance is allocated to the grid (i, j), then the channel $k = i \cdot S + j$ on the mask branch is responsible for predicting the mask of the target; each grid belongs to a single instance only. Finally, we use the non-maximum-suppression (NMS) algorithm to obtain the final results. Compared with a two-stage method, our one-stage method is simpler and can connect nucleus segmentation to a location to achieve better results.

The loss function includes two parts: category branch and mask branch. The loss function is as follows:

$$L_{DMF} = L_{\text{focal}} + \gamma L_{\text{m}} \tag{8}$$

The Sigmoid activation function output is used here. L_{focal} represents the category branch, and uses the traditional semantic segmentation loss function Focal Loss (Shaoqing et al., 2017) to measure the gap between the predicted category and the ground truth. L_m is the loss function of the mask branch, specifically expressed as:

$$L_{\text{m}} = \frac{1}{N} \sum_k \beta p_{\{i,j>0\}} L_{\text{dice}} \tag{9}$$

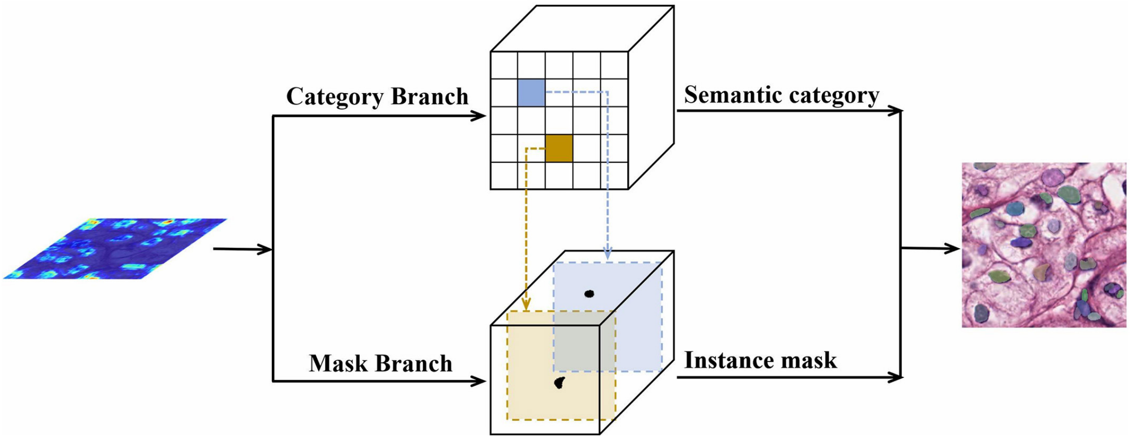


FIGURE 4
Mask generation network for mask level.

3 Results

3.1 Dataset

MonuSeg stands for Multi-organ Nucleus Segmentation, and the dataset was published at the official satellite event of MICCAI 2018.The MoNuSeg 2018 dataset contains 30 tissue images and 21,623 annotated nuclear boundaries, each image of size $1,000 \times 1,000$ pixels (Kumar et al., 2017). The dataset used H&E-stained tissue slides digitized at 40x magnification and contained nuclei of varying sizes from seven different organs. These organs include the bladder, liver, breast, kidney, colon, prostate, and stomach. We cropped each image into 16 patches, and the size of each patch was 250 × 250 pixels. Specifically, we generated 480 images, including 352 training, 32 validation, and 96 test images. Furthermore, we used data augmentation to augment the size of the datasets and reduce overfitting. Before the image is input into the model, it will go through a channel composed of different data enhancement methods. Each enhancement method is set with a certain probability value and different enhancement factors. In other words, each image will follow the data in the channel. Augmentation is randomly combined with a set probability. Numerous image transformation schemes that were used include brightness enhancement, contrast reduction, Gaussian noise, impulse noise, and Poisson noise. Some representative examples of data augmentation are shown in Figure 5.

3.2 Implementation details

The DMFNet was implemented in PyTorch and trained on an NVIDIA Tesla V100 GPU with 32 GB of video memory. During the training, the mini-batch strategy was used to iteratively train the DMFNet for 200 epochs, and each iteration used two samples as a batch, with a total of 35,200 iterations. The validation set was evaluated after each training epoch. The network used batch normalization (Sergey and Christian, 2015) for regularization every time the weight was updated, and the stochastic gradient descent was used to update the model parameters. The learning rate, weight decay, and momentum were set to 0.0025, 0.0001, and 0.9, respectively.

3.3 Evaluation metrics

In object detection, the intersection-over-union (IoU) metric, which is the ratio of the intersection to the union of the prediction bounding box generated by the network and the original ground truth bounding box, was used. The evaluation method of instance segmentation was very similar to the evaluation method of object detection, with the difference being that the IoU of the mask was calculated in lieu of the IoU of the bounding box. In this study, precision and recall under a specific IoU threshold were considered as the evaluation indicators, and the expressions for the same are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

where *TP* is the number of nuclei that are actually nuclei and are predicted by the model to be nuclei; *FP* is the number of nuclei that are actually background, but are predicted by the model; and *FN* is the number of nuclei that are actually nuclei, but not recognized as nuclei by the network. The *Precision* measures the proportion of samples that the model predicts as positive class (in nucleus segmentation, that is, predicted as nuclei) and are actually positive class among the samples predicted as positive class by the model. The *Recall* represents the proportion of samples that are actually positive class and are predicted as positive class by the model among the total number of actual positive class samples.The threshold of the IoU was calculated every 0.05 from 0.5 to 0.95 and the average precision (AP) was calculated every 0.05. The mean average precision (*mAP*) of all the results was used as the main indicator

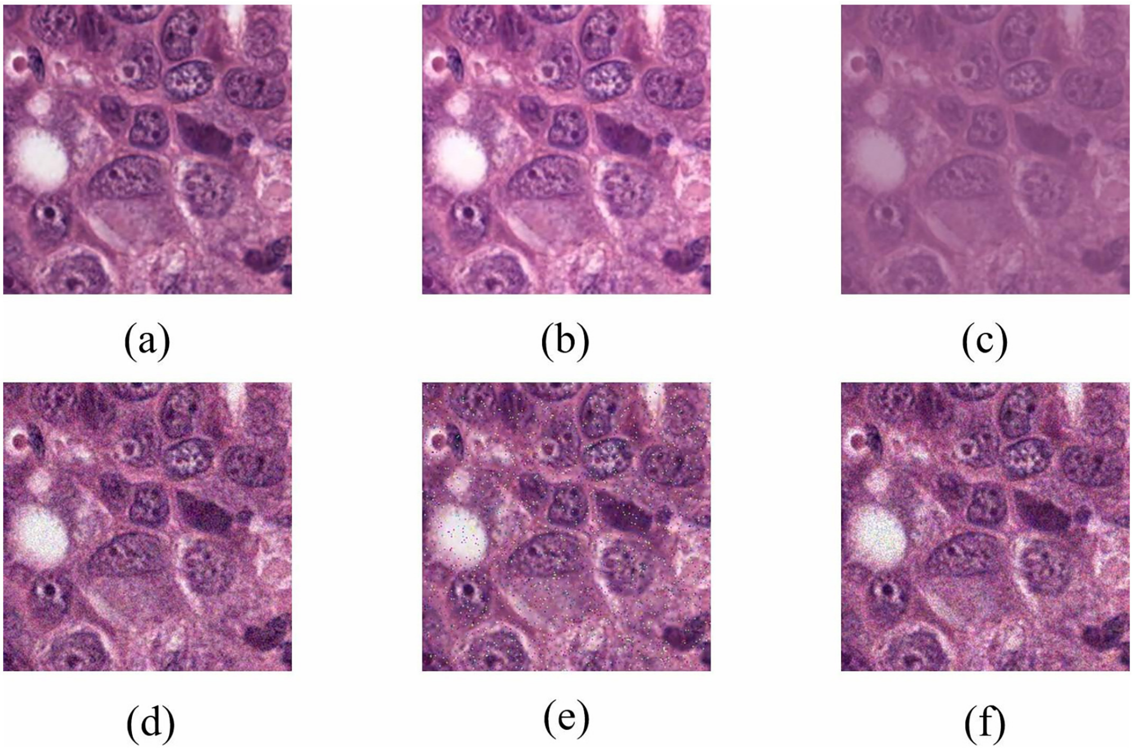


FIGURE 5
Representations of some image transformation schemes used. **(A)** Original. **(B)** Brightness enhancement. **(C)** Contrast reduction. **(D)** Gaussian noise. **(E)** Impulse noise. **(F)** Poisson noise.

to report the results of the DMFNet. The AP under a specific IoU threshold was calculated as follows:

$$AP_{IoU=k} = \frac{1}{101}\sum_{r\in R} p_{\text{interp}}(r) = \frac{1}{101}\sum_{r\in R} \max_{\tilde{r}\cdot\tilde{r}\geq r} p(\tilde{r}) \qquad (12)$$

where k indicates the threshold in $K:[0.5, 0.55, ..., 0.90, 0.95]$, r denotes the recall, and $R:[0, 0.01, 0.02, ..., 0.99, 1.0]$, with an interval of 0.01 and a total of 101 values. $p(\tilde{r})$ denotes the precision related to the recall rate $\tilde{r}$. To calculate the AP value at the ten thresholds, we considered the average value at the ten thresholds as the mean average precision (mAP). In addition, we also used the average recall (AR) as an evaluation metric, which was obtained by testing the mean $AR_{IoU} = k$ of more than 10 IoU thresholds, and a maximum of the top 100 predicted masks were given. Similarly, we also considered the average value at the ten thresholds as the mean average recall (mAR). In this study, the task of the model was to identify only one category; therefore, $AR_{IoU} = k$ at a specific threshold was equal to R in Equation 12. mAP is an important indicator for evaluating model performance, which measures the model's ability to accurately identify cell nuclei. A high mAP value indicates that the model performs well in accuracy and completeness. AR reflects the ability of the model to detect actual cell nuclei. In cell nucleus segmentation, high AR ensures that cell nucleus information is not missed, which can help detect diseased cells in early cancer screening in a timely manner. mAR measures the average performance of the model at different recall thresholds, calculating the average proportion of correctly predicted positive samples at multiple levels to the actual total number of positive samples. In addition, for mAP and mAR, we also used the following metric:

1. AP_{50}: AP value over a single threshold of IoU = 0.50.
2. AP_{75}: AP value over a single threshold of IoU = 0.75.
3. AR_{50}: AR value over a single threshold of IoU = 0.50.
4. AR_{75}: AR value over a single threshold of IoU = 0.75.

3.4 Ablation study

To verify the effectiveness of the feature extraction and feature integration modules, we used a network with and without those modules, embedded the same into the DMFNet, and then trained and evaluated them separately. To make an unbiased comparison, both these models used the same experiment and hyperparameter configuration. Table 1 summarizes the ablation studies on the effects of each module of the DMFNet.

As shown in Table 1, the DCN module was embedded in the SOLO, and the AP and AR scores of the segmented network were significantly improved, which means that the deformable feature extraction could better improve the transformation modeling capability of the model. This also proves that the deformable convolution had a higher accuracy for the irregular circle of the nucleus. When the BFP module was added, the mAP and mAR increased by 1.9% and 0.9%, respectively. This verifies that the multi-level feature integration played a key role. Specifically, the enhancement effect of the two modules was inconsistent; however, using the two modules at the same time yielded a better result, and it was proved that the DMFNet integrated the two modules effectively.

Next, we visualized the ablation study and analyzed the details of our model. More typical examples are provided in Figure 6, including different stained nuclei and their masks. A model without the modules would result in errors, such as a small number of segmented nuclei and a lack of distinction between adjacent nuclei. The reason for this behavior was that the network did not completely learn the overall features and, hence, did not fully identify the edges of the features. In contrast, the DMFNet could segment more nuclei, correctly segment adjacent nuclei, and also segment each nucleus more completely. This means that our model offers advantages in nuclei edge extraction and global feature integration, and it is also proved that the two modules played a significant role.

TABLE 1 Effects of each module of the DMFNet (%).

DCN	*BFP*	*mAP*	AP_{50}	AP_{75}	*mAR*	AR_{50}	AR_{75}
		29.7	64.6	24.8	38.6	72.3	38.4
✓		34.9 $\uparrow_{5.2}$	70.1 $\uparrow_{5.5}$	32.5 $\uparrow_{7.7}$	43.0 $\uparrow_{4.4}$	77.3 $\uparrow_{5.0}$	46.6 $\uparrow_{8.2}$
	✓	31.6 $\uparrow_{1.9}$	67.7 $\uparrow_{3.1}$	26.1 $\uparrow_{1.3}$	39.5 $\uparrow_{0.9}$	73.9 $\uparrow_{1.6}$	40.0 $\uparrow_{1.6}$
✓	✓	37.8 $\uparrow_{8.1}$	77.8 $\uparrow_{13.2}$	33.6 $\uparrow_{8.8}$	47.4 $\uparrow_{8.8}$	85.3 $\uparrow_{13.0}$	49.0 $\uparrow_{10.6}$

3.5 Comparisons

To verify the effectiveness of the DMFNet on nuclei segmentation, we compared the proposed model with a few state-of-the-art methods, with the same experimental configuration. Table 2 lists the instance segmentation results of each of these methods. Compared to the existing baseline methods, for example, Mask R-CNN (Kaiming et al., 2017), TensorMask (Xinlei et al., 2019), and PolarMask (Xie et al., 2020), the proposed method yielded better results than the state-of-the-art methods. The DMFNet achieved the best performance on the MoNuSeg 2018

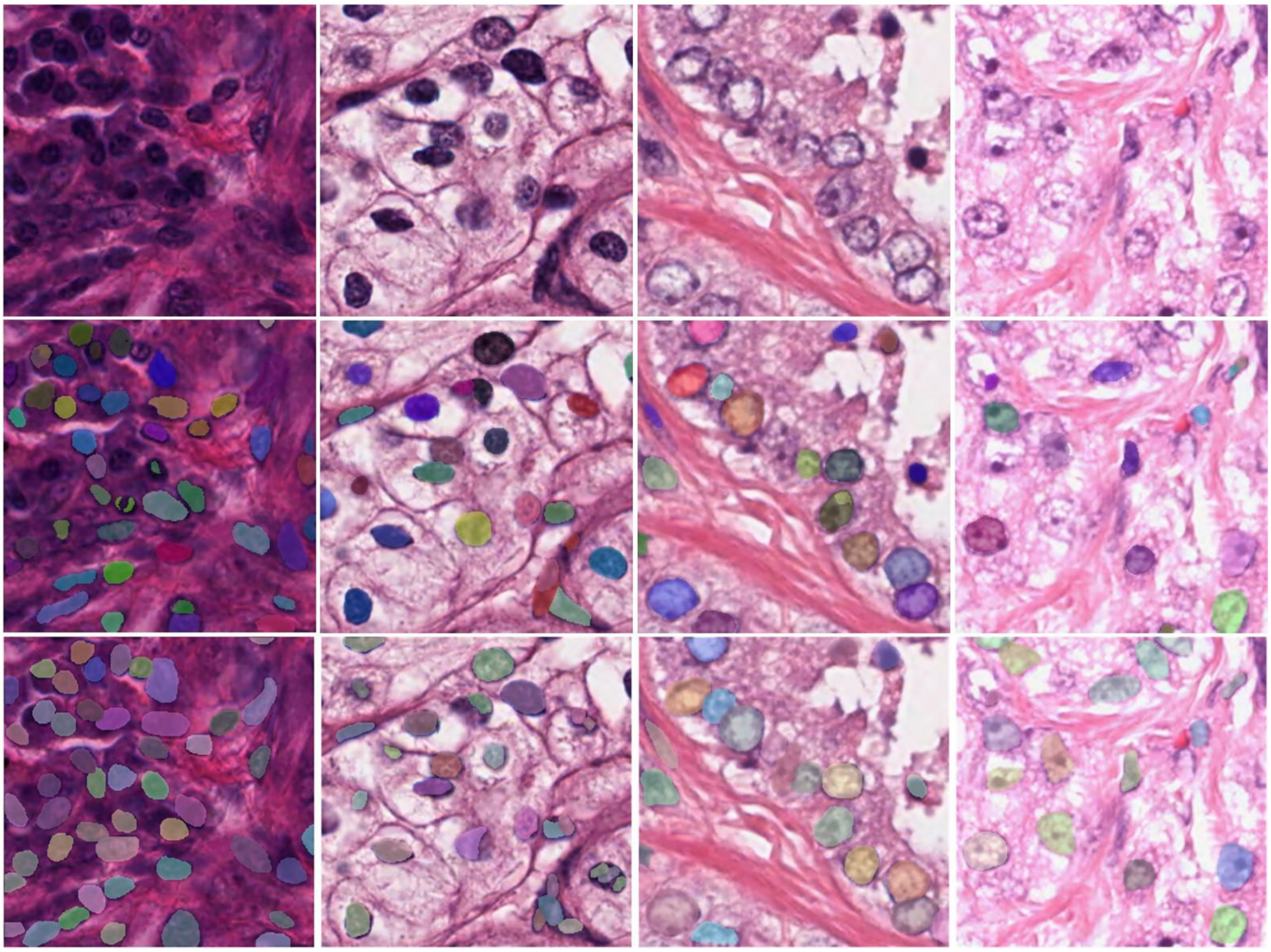


FIGURE 6
Visualization comparison between the DMFNets (with or without the two feature processing modules) and examples of segmentations model; the first row contains the original tissue images; the second row contains the images from the DMFNet without the two modules; while the third row contains the images from our proposed DMFNet.

TABLE 2 Effects of each module of the DMFNet (%).

Method	mAP	AP_{50}	AP_{75}	mAR	AR_{50}	AR_{75}
PolarMask	21.7	52.4	14.5	31.8	64.1	29.4
Mask R-CNN	30.3	67.8	24.1	40.9	77.4	40.2
TensorMask	31.4	67.4	27.1	39.8	74.0	39.7
SOLO	29.7	64.6	24.8	38.6	72.3	38.4
DMFNet	37.8	77.8	33.6	47.4	85.3	49.0

dataset, with its mAP and mAR approaching 37.8% and 47.4%, respectively. Furthermore, when the IoU threshold was 0.50 and 0.75, our method yielded a better performance, and both AP and AR were the best.

Figure 7 shows the segmentation P-R curves (smoothed) of five models with 10 IoU thresholds ranging from 0.5 to 0.75 with an interval of 0.05. Figure 7 shows that with a change in the IoU threshold, the trend of the model was approximately the same, and the corresponding P-R curve gradually approached the coordinate axis; however, the closing speed of the P-R curve of the proposed model was slower than that of the other models. This proves that the prediction result of the proposed DMFNet had a higher score and better overall quality.

4 Discussion

Changes in nuclear morphology are closely related to the growth state of tumors. For nearly 150 years, changes in nuclear morphology have been the gold standard for cancer diagnosis, so it is critical to understand single nuclear instances. With the development of digital pathology, the emergence of electronic pathology pictures helps pathologists get rid of microscopes, but it is very cumbersome and difficult to browse and observe pathological nuclei on the computer. The main method for pathologists to quantify nuclei is subjective estimation, which cannot be done precise quantification.

In digital pathology research, instantiating nuclei can help pathologists understand the structure of individual nuclei. Quantitative analysis of the spatial distribution of nuclear clusters and the number of mitoses, which are key factors in cancer diagnosis and prognosis. In practice, there are independent nuclei and clusters of nuclei that are clustered together. Although the semantic segmentation of nuclei has an excellent effect and can extract nuclei from pathological images, it is not suitable for independent study of nuclei clusters or adjacent nuclei. Scene of the nucleus in a cell.

For the above reasons, we propose a model DMFNet to segment pathological image nuclei from the perspective of instance segmentation. DMFNet analyzes the process of nucleus instance segmentation at both feature and segmentation levels. At the feature level, we added a feature extraction module for the diverse morphological characteristics of nuclei, which enhanced the model's transformation and modeling capabilities to better sample target instances. In nucleus segmentation, both detail features and semantic features play a very important role, so we propose to use the feature integration module to integrate and enhance the two features. At the mask level, we replace the traditional methods using anchor boxes or clustering with a one-stage location-based instance segmentation method, making the model simpler. Experiments show that the method in this paper can effectively improve the accuracy of nucleus instance segmentation.

In summary, both feature extraction and feature integration modules can improve the accuracy of nucleus segmentation; moreover, their combination can further improve the performance, which not only validates that their combination is suitable for the task of nucleus segmentation, but also shows that they differ from feature level Aspects enhance the segmentation task. Therefore, we can use these two modules simultaneously in the application scenario of nucleus segmentation to achieve optimal segmentation results.

In the field of medical image processing, the development of automatic annotation technologies is of critical significance for improving the efficiency of clinical applications. The DMFNet model in this study demonstrates remarkable potential in the automatic annotation of medical images. By achieving precise segmentation of cell nuclei, the model can automatically identify and annotate key information such as the location and boundaries of the nuclei. This greatly reduces the workload associated with manual annotation. Traditional manual annotation methods require significant time and effort from pathologists, especially when handling large-scale medical image data. In contrast, the DMFNet model can rapidly and accurately perform the annotation task. For example, annotating pathological slide images containing numerous cell nuclei may take several hours or even days manually, while the DMFNet model can complete the preliminary annotation in a much shorter time, achieving high accuracy. This not only improves annotation efficiency but also provides timely and reliable data support for subsequent clinical diagnosis and research, thereby promising to enhance overall clinical application efficiency and provide robust support for early disease diagnosis and precision treatment.

In terms of practical clinical applications, this nuclear cell segmentation model holds great potential. In the process of cancer diagnosis, accurate nuclear cell segmentation is a crucial step. Currently, the incidence rate of cancer remains at a relatively high level, and early diagnosis is of great significance for improving patients' survival rates and quality of life. By enhancing the accuracy of nuclear cell instance segmentation, this model can provide pathologists with more accurate and detailed information about nuclear cells. For example, in the diagnosis of common cancers such as breast cancer and lung cancer, pathologists can utilize this model to observe the morphology, size, and distribution of nuclear cells more clearly, thereby making a more accurate judgment on the benign or malignant nature of the tumor. This is of great importance for the early detection of tiny tumors, the determination of cancer staging, and the formulation of personalized treatment plans. Moreover, during the follow-up after cancer treatment, this model can also be used to monitor changes in the morphology of nuclear cells, enabling the timely detection of signs of cancer recurrence.

We analyzed the problem of nuclei segmentation in pathological images and proposed a model for nuclei segmentation,

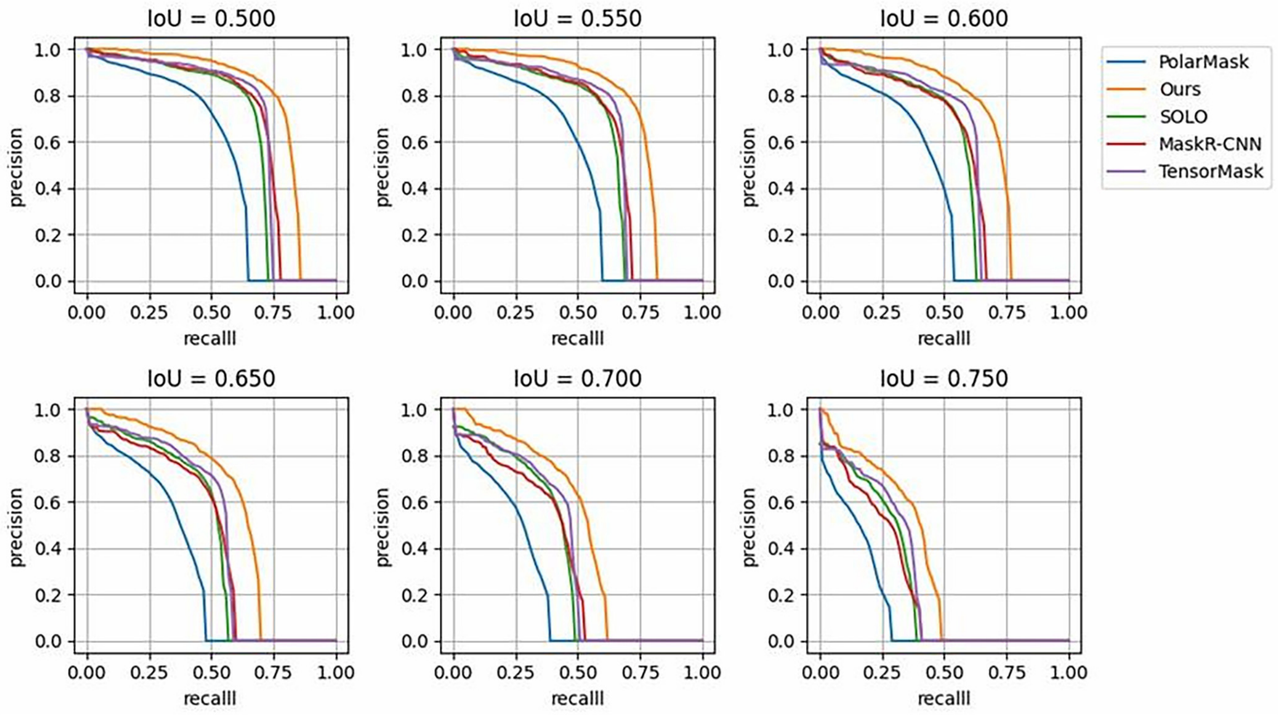


FIGURE 7
P-R Curves of state-of-the-art models over different IoU thresholds, compared with those of the proposed DMFNet.

but with the change of datasets and the development of medical images, the model still needs to be further optimized. Although the nucleus segmentation dataset we used is a well received work in recent years, in the process of analyzing the edges, it is found that the labeling of the edges is not fine enough and the amount of training data is relatively small.Therefore, transfer learning should be performed in combination with newly released datasets to improve the robustness and accuracy of the model. We only segment the nuclei in the pathological images without considering the types of nuclei. In tumor tissues, the cells included not only cancer cells, but also stromal cells, lymphocytes, macrophages, etc. Recent studies have shown that tumor cell nuclei Stromal cell interactions are involved in tumor progression and metastasis. In the following work, the fine-grained classification of the research cells is also needed, and the cells are divided into tumor cells, stromal cells, lymphocytes, etc.

5 Conclusions

In this study, we proposed an innovative model, DMFNet, which holds significant value in clinical applications. This model is mainly used for nuclear segmentation from digital pathology images of different organs, a function that is of great significance for the clinical diagnosis and treatment of diseases such as cancer. In clinical practice, the accuracy of nuclear segmentation is crucial for determining the nature and development stage of tumors. We conducted a detailed analysis of the nuclear segmentation process and made targeted improvements to key components such as feature extraction, fusion, and template generation. Through these enhancements, DMFNet effectively combines DCN, BFP, and SOLO, significantly improving the performance of the segmentation network.

However, the DMFNet model still has certain limitations. At the methodological level, although it exhibits good performance on the existing dataset, with the development of medical imaging technology and the emergence of new datasets, the model may face adaptability issues. From the perspective of network structure, although the performance has been improved through module combination, when dealing with large-scale, high-resolution pathological images, the problem of excessive consumption of computational resources is rather prominent, which affects the running efficiency of the model to a certain extent. In terms of feature extraction, the expression of features such as the morphology and texture of the cell nucleus is not rich enough, making it difficult to capture some subtle but critical pathological features, which may thus affect the accuracy of segmentation. In the process of feature fusion, the fusion method of different hierarchical features is not optimal, resulting in information loss or redundancy, leading to the underutilization of some features. The template generation process is not flexible enough in adapting to the diverse morphological distributions of cell nuclei, and the processing ability of the model is limited when facing complex pathological situations.

In the future, we will further improve the model from the directions of enhancing model performance (such as optimizing feature extraction, fusion, and template generation), expanding application scenarios (to more organ diseases and integrating with clinical processes), and increasing model interpretability (visualization and constructing explanatory models). When the DMFNet was applied to the MoNuSeg 2018 dataset, the experimental results clearly demonstrated the performance advantages of this method over some existing methods. This implies that in actual clinical scenarios, pathologists can utilize this model to more accurately extract nuclear information from pathological images. Meanwhile, during the follow-up of diseases, this model also helps in continuously monitoring the changes of cell nuclei, enabling the timely detection of disease progression or recurrence.

In summary, the DMFNet model provides important support for clinical diagnosis and treatment and has remarkable clinical value.

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

SC: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. TY: Conceptualization, Investigation, Software, Writing – original draft. BY: Formal analysis, Visualization, Writing – original draft. JZ: Data curation, Formal analysis, Writing – original draft. LM: Validation, Writing – original draft. YD: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing. XS: Conceptualization, Formal analysis, Funding acquisition, Methodology, 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 National Nature Science Foundation of China (62302279 and 62272283); the Natural Science Foundation of Shandong Province (ZR2021QF135); the "Young Innovation Team Program" of Shandong Provincial University (2022KJ250); New Twentieth Items of Universities in Jinan (2021GXRC049).

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

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M. (2022). "Swin-unet: Unet-like pure transformer for medical image segmentation,"? in *European Conference on Computer Vision* (Cham: Springer), 205–218. doi: 10.1007/978-3-031-25066-8_9

Carsen, S., Tim, W., Michalis, M., and Marius, P. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* 18, 100–106. doi: 10.1038/s41592-020-01018-x

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y. (2021). Transunet: transformers make strong encoders for medical image segmentation. *arXiv* [preprint] arXiv,:2102.04306. doi: 10.48550/arXiv.2102.04306

Faisal, M., Daniel, B., Richard, C., Gregory, N. M., Kevan, J. S., Alexander, B., et al. (2020). Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans. Med. Imag.* 39, 3257–3267. doi: 10.1109/TMI.2019.2927182

Fuyong, X., and Lin, Y. (2016). Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Rev. Biomed. Eng.* 9, 234–263. doi: 10.1109/RBME.2016.2515127

Hao, C., Xiaojuan, Q., Lequan, Y., and Pheng-Ann, H. (2016). "DCAN: Deep contour-aware networks for accurate gland segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 2487–2496.

He, Y., Nath, V., Yang, D., Tang, Y., Myronenko, A., and Xu, D. (2023). "Swinunetr-v2: Stronger swin transformers with stagewise convolutions for 3d medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer).

Jiangmiao, P., Kai, C., Jianping, S., Huajun, F., Wanli, O., and Dahua, L. (2019). "Libra r-cnn: Towards balanced learning for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE), 821–830.

Jifeng, D., Haozhi, Q., Yuwen, X., Yi, L., Guodong, Z., Han, H., and Yichen, W. (2017). "Deformable convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Venice: IEEE), 764–773.

Jingru, Y., Pengxiang, W., Qiaoying, H., Hui, Q., Bo, L., Daniel, J., et al. (2019). "Multi-scale cell instance segmentation with keypoint graph based bounding boxes," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*, eds. S. Dinggang, L. Tianming, E. Caroline, Z. Sean, Y. Pew-Thian, and K. Ali (Cham. Springer International Publishing), 369–377.

Jonathan, L., Evan, S., and Trevor, D. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA: IEEE), 640–651.

Kaiming, H., Georgia, G., Piotr, D., and Ross, G. (2017). "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Venice: IEEE), 386–397.

Kaiming, H., Xiangyu, Z., Shaoqing, R., and Jian, S. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 770–778.

Konstantin, S., Olga, B., and Anton, K. (2019). "AdaptiS: Adaptive instance selection network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul: IEEE), 7354–7362. doi: 10.1109/ICCV.2019.00745

Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., and Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* 36, 1550–1560. doi: 10.1109/TMI.2017.2677499

Liu, T., Guo, Q., Lian, C., Ren, X., Liang, S., Yu, J., Niu, L., Sun, W., and Shen, D. (2019). Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks. *Med. Image Analy.* 58:101555. doi: 10.1016/j.media.2019.101555

Liu, X., Wang, S., Zhang, Y., Liu, D., and Hu, W. (2021). Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning. *Neurocomputing* 452, 576–591. doi: 10.1016/j.neucom.2020.07.143

Ortiz, R., de Medeiros, G., Peters, A. H., Liberali, P., and Rempfler, M. (2020). "RDCNet: Instance segmentation with a minimalist recurrent residual network," in *Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020* (Lima, Peru: Springer), 434–443.

Oskar, V. A., Ullah, A. S., and Juho, K. (2019). "Mask-RCNN and U-Net ensembled for nuclei segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (Venice: IEEE), 208–212.

Peifang, G., Alan, E., and Prabir, B. (2016). "Segmentation of nuclei in digital pathology images," in *2016 IEEE 15th international conference on cognitive informatics & cognitive computing (ICCI* CC)* (Palo Alto, CA: IEEE), 547–550.

Peter, N., Marick, L., Fabien, R., and Thomas, W. (2019). Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans. Med. Imag.* 38, 448–459. doi: 10.1109/TMI.2018.2865709

Sergey, I., and Christian, S. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, eds. B. Francis, and B. David (Lille: PMLR), 448–456.

Shaoqing, R., Kaiming, H., Ross, G., and Jian, S. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Analy. Mach. Intellig.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Shengcong, C., Changxing, D., and Dacheng, T. (2020). "Boundary-assisted region proposal networks for nucleus segmentation," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020*, eds. A. Purang, S. Danail, M. Diana, Z. S. Kevin, R. Daniel, and J. Leo (Cham: Springer International Publishing), 279–288.

Shu, L., Lu, Q., Haifang, Q., Jianping, S., and Jiaya, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (San Francisco, CA: IEEE), 8759–8768.

Tsung-Yi, L., Piotr, D., Ross, G., Kaiming, H., Bharath, H., and Serge, B. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: IEEE), 936–944.

Wang, P., Hu, X., Li, Y., Liu, Q., and Zhu, X. (2016). Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Proc.* 122, 1–13. doi: 10.1016/j.sigpro.2015.11.011

Xiaolong, W., Ross, G., Abhinav, G., and Kaiming, H. (2018). "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Salt Lake City, UT: IEEE), 7794–7803.

Xie, E., Sun, P., Song, X., Wang, W., Liu, X., Liang, D., et al. (2020). "Polarmask: single shot instance segmentation with polar representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12193–12202.

Xieli, L., Yuanyuan, W., Qisheng, T., Zhen, F., and Jinhua, Y. (2019). Dual U-Net for the segmentation of overlapping glioma nuclei. *IEEE Access* 7, 84040–84052. doi: 10.1109/ACCESS.2019.2924744

Xinlei, C., Ross, G., Kaiming, H., and Piotr, D. (2019). "TensorMask: a foundation for dense object segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul: IEEE), 2061–2069.

Xinlong, W., Tao, K., Chunhua, S., Yuning, J., and Lei, L. (2020). "Solo: Segmenting objects by locations," in *Computer Vision-ECCV 2020*, eds. V. Andrea, B. Horst, B. Thomas, and F. Jan-Michael (Cham: Springer International Publishing), 649–665.

Xinpeng, X., Jiawei, C., Yuexiang, L., Linlin, S., Kai, M., and Yefeng, Z. (2020). "Instance-aware self-supervised learning for nuclei segmentation," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020*, eds. P. Abolmaesumi, S. Danail, M. Diana, M., Z. S. Kevin, R. Daniel, and J. Leo (Cham: Springer International Publishing), 341–350.

Yiming, L., Pengcheng, Z., Qingche, S., Andi, L., Peng, Z., and Zhiguo, G. (2018). Automatic segmentation of cervical nuclei based on deep learning and a conditional random field. *IEEE Access* 6, 53709–53721. doi: 10.1109/ACCESS.2018.2871153

Zeiler, M. D., and Rob, F. (2014). "Visualizing and understanding convolutional networks," in *Computer Vision-ECCV 2014*, eds. F. David, P. Tomas, S. Bernt, and T. Tinne (Cham. Springer International Publishing), 818–833.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 13 December 2024
DOI 10.3389/fmicb.2024.1509915

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Baolin Wu,
Huaxi MR Research Center (HMRRC), China
Rongrong Cao,
Soochow University, China
Kening Sun,
General Hospital of Ningxia Medical University, China
Xipeng Long,
Zhejiang University, China

*CORRESPONDENCE
Lina Wu
✉ 1066813620@qq.com

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

RECEIVED 11 October 2024
ACCEPTED 27 November 2024
PUBLISHED 13 December 2024

CITATION
Huang B, Gao Y and Wu L (2024) Assessment of body composition and prediction of infectious pancreatic necrosis via non-contrast CT radiomics and deep learning.
Front. Microbiol. 15:1509915.
doi: 10.3389/fmicb.2024.1509915

Assessment of body composition and prediction of infectious pancreatic necrosis via non-contrast CT radiomics and deep learning

Bingyao Huang[1†], Yi Gao[2†] and Lina Wu[2*]

[1]Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China,
[2]Department of Laboratory Medicine, Shengjing Hospital of China Medical University, Shenyang, China

Aim: The current study aims to delineate subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), the sacrospinalis muscle, and all abdominal musculature at the L3–L5 vertebral level from non-contrast computed tomography (CT) imagery using deep learning algorithms. Subsequently, radiomic features are collected from these segmented images and subjected to medical interpretation.

Materials and methods: This retrospective analysis includes a cohort of 315 patients diagnosed with acute necrotizing pancreatitis (ANP) who had undergone comprehensive whole-abdomen CT scans. The no new net (nnU-Net) architecture was adopted for the imagery segmentation, while Python scripts were employed to derive radiomic features from the segmented non-contrast CT images. In light of the intrinsic medical relevance of specific features, two categories were selected for analysis: first-order statistics and morphological characteristics. A correlation analysis was conducted, and statistically significant features were subjected to medical scrutiny.

Results: With respect to VAT, skewness ($p = 0.004$) and uniformity ($p = 0.036$) emerged as statistically significant; for SAT, significant features included skewness ($p = 0.023$), maximum two-dimensional (2D) diameter slice ($p = 0.020$), and maximum three-dimensional (3D) diameter ($p = 0.044$); for the abdominal muscles, statistically significant metrics were the interquartile range (IQR; $p = 0.023$), mean absolute deviation ($p = 0.039$), robust mean absolute deviation ($p = 0.015$), elongation ($p = 0.025$), sphericity ($p = 0.010$), and surface volume ratio ($p = 0.014$); and for the sacrospinalis muscle, significant indices comprised the IQR ($p = 0.018$), mean absolute deviation ($p = 0.049$), robust mean absolute deviation ($p = 0.025$), skewness ($p = 0.008$), maximum 2D diameter slice ($p = 0.008$), maximum 3D diameter ($p = 0.005$), sphericity ($p = 0.011$), and surface volume ratio ($p = 0.005$).

Conclusion: Diminished localized deposition of VAT and SAT, homogeneity in the VAT and SAT density, augmented SAT volume, and a dispersed and heterogeneous distribution of abdominal muscle density are identified as risk factors for infectious pancreatic necrosis (IPN).

KEYWORDS

pancreatitis, infectious pancreatic necrosis, artificial intelligence, adipose tissue, musculature

1 Introduction

Acute necrotizing pancreatitis (ANP), a grave complication of acute pancreatitis (AP), arises from the aberrant activation of pancreatic digestive enzymes, resulting in tissue necrosis. This necrotic tissue creates an environment conducive to bacterial proliferation, which frequently precipitates infectious pancreatic necrosis (IPN), occurring in approximately 30% of ANP patients and manifesting a mortality rate as high as 30%. Consequently, the early and precise diagnosis and treatment of IPN are of paramount importance. Given the intimate connection between the pathophysiology of ANP and systemic metabolism, recent research has focused on the roles of muscle and adipose tissue in ANP and its complications (Yee et al., 2021; Zhang X. et al., 2024).

In unraveling the intricate pathogenesis of IPN, it is essential to consider the potential contributions of body composition, particularly with respect to muscle and fat. Skeletal muscle, a fundamental component of the human body, not only underpins essential motor functions but is also intimately associated with an individual's metabolic state, inflammatory responses, and long-term clinical outcomes. Adipose tissue, especially visceral fat, is recognized as a "metabolically active" entity that secretes various bioactive molecules involved in regulating energy metabolism, inflammatory reactions, and immune functions. Recent evidence suggests that alterations in muscle and fat content and distribution may exert direct or indirect influences on the progression and outcomes of AP (Fu et al., 2023; Dawra et al., 2023). Nonetheless, a comprehensive understanding of the specific roles these changes play in the development of IPN remains elusive. This study seeks to bridge this gap by examining the distinctive alterations in muscle and fat among IPN patients, thereby offering a novel perspective on the multifaceted disease trajectory of IPN. While the previous studies have acknowledged the significance of assessing body composition, they have been mainly confined to macroscopic observations and correlative analyses, with a limited exploration of the precise mechanisms by which muscle and fat impact IPN. Hence, this study will utilize CT scans to meticulously evaluate the distribution of muscle and fat in IPN patients. This methodological approach is anticipated to elucidate the intrinsic correlation between muscle and fat status and the pathophysiological underpinnings of IPN, providing a scientific foundation for early diagnosis, therapeutic strategy development, and prognostic enhancement. The objective is to furnish new theoretical insights and practical guidance for precision medicine and IPN management.

Computed tomography (CT), as a prevalent imaging technique, offers distinct advantages in assessing body composition. It not only delineates the distribution of muscles and fats with clarity but also quantitatively analyzes critical parameters, such as muscle area and fat content, through precise measurement tools, thereby affording clinicians a wealth of morphological and functional information (Zhang R. et al., 2023; Vogele et al., 2023). In the evaluation of pancreatitis, CT accurately portrays pancreatic morphological alterations, necrotic regions, and the spread of inflammation, playing a pivotal role in diagnosing ANP and monitoring disease progression (Balthazar et al., 1990). Moreover, CT has proven to be particularly adept at assessing muscle and fat, enabling the exact measurement of their distribution and proportions, which is instrumental in evaluating patients' nutritional status, inflammatory responses, and disease prognoses (Hou et al., 2024). The advent of artificial intelligence technology, particularly the extensive application of deep learning algorithms in medical image processing, has introduced a novel perspective and set of tools for exploring the complex pathological mechanisms of AP (Zhang C. et al., 2024; Yin et al., 2024). By leveraging deep learning algorithms, we can uncover the profound features embedded within the vast repository of CT image data (Zhang R. et al., 2023), which may be intimately linked to the pathological changes of IPN, thereby facilitating early prediction and precise treatment of the condition. Despite the limitations of non-enhanced CT in the traditional visual diagnosis of pancreatic diseases due to the lack of contrast, its amalgamation with radiomics technology has yielded promising diagnostic outcomes (Koç and Taydaş, 2020; Janisch et al., 2022; Cao et al., 2023).

This study explores the utility of body composition assessment based on non-contrast CT in ANP patients and harnesses deep learning and radiomics techniques to delve into the potential connections between body components and the onset of IPN. Additionally, we aspire to provide novel insights and strategies for the early detection and personalized treatment of the disease by examining the interplay with body composition.

2 Materials and methods

2.1 Patients

This study was conducted in accordance with the Declaration of Helsinki and received ethical approval from the Ethics Committee of Shengjing Hospital at China Medical University, with a waiver of informed consent for participants (ethical approval number: 2024PS1480K). As depicted in Figure 1, we conducted a retrospective analysis of data from patients who were diagnosed with ANP and admitted to our institution between March 2019 and August 2024 and underwent CT scans within a week of symptom onset. Inclusion criteria included CT scans performed within 1 week of admission. Exclusion criteria included: (1) pregnancy; (2) age below 18 years; (3) concurrent malignancy; and (4) non-whole-abdomen CT scans, poor image quality, or incomplete clinical data that could compromise the accuracy and reliability of the assessment outcomes.

Clinical data collected included age, sex, IPN status, diabetes, hypertension, hyperlipidemia, hypoxemia, coronary heart disease, gallstone pancreatitis, mechanical ventilation, and hospital stay duration.

2.2 CT image acquisition

Patients were subjected to whole-abdomen CT imaging within 1 week following admission. All scans were performed with the patients in a supine position during inhalation using a (1) Philips Brilliance ICT 256-slice spiral CT scanner (Philips Healthcare). (2) Python package PyRadiomics version 3.0.1 (Python Software Foundation). (3) Statistical Package for the Social Sciences (SPSS) version 26.0 (IBM Corp). The scanning field extended from the diaphragmatic dome to the pubic symphysis. Scan parameters were set as follows: tube voltage at 120 kV, tube current adjusted to automatic milliamperage, matrix size of 512 × 512, a pitch of 1, with routine images at a slice thickness of 3.0 mm, and thin-section images at 1.0 mm intervals.

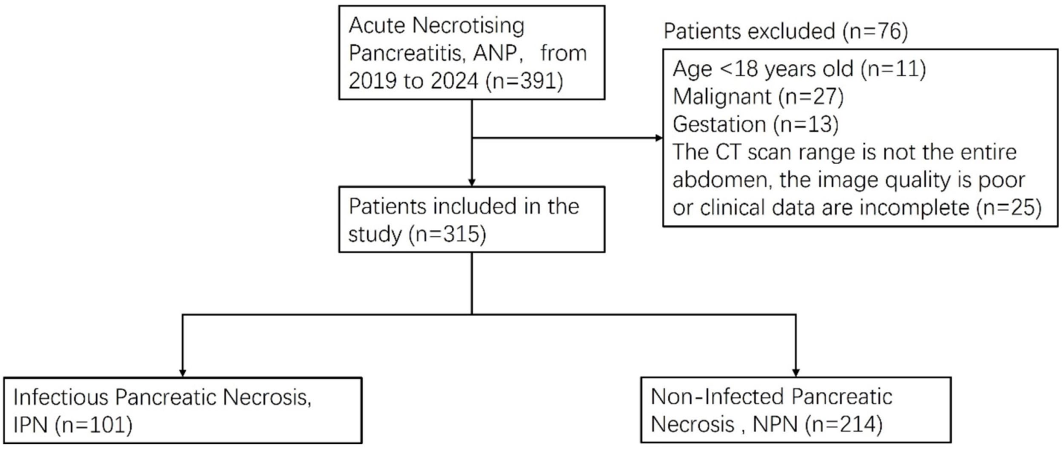


FIGURE 1
Patients' enrollment and exclusion process in the infectious pancreatic necrosis (IPN) database.

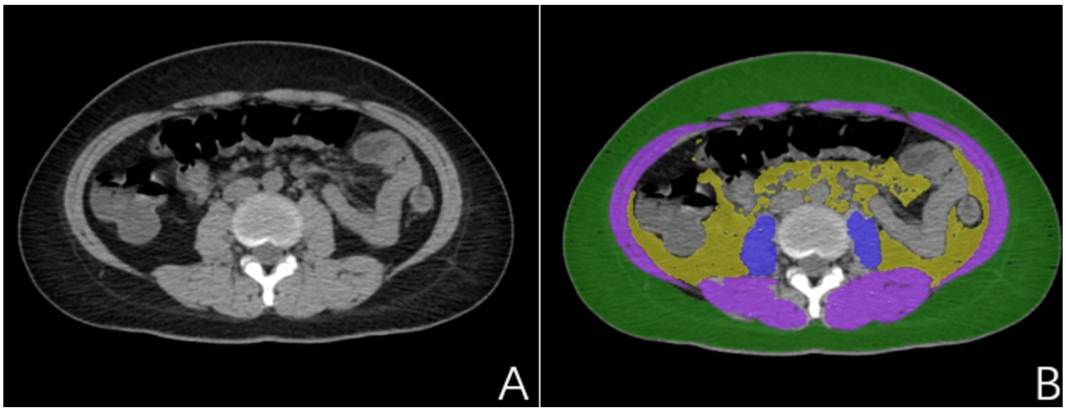


FIGURE 2
(A) Non-pancreatitis crowd image; **(B)** Delineated subcutaneous adipose tissue (SAT) (green), visceral adipose tissue (VAT) (yellow), sacrospinalis (blue), and abdominal muscles (blue and purple).

2.3 Data annotations

The present study utilized a stringent data annotation protocol to guarantee precision and consistency. To ensure the model's generalizability and to reduce the interference of pancreatitis on the delineation of visceral fat, we employed CT images from a distinct cohort of individuals without pancreatitis for region of interest (ROI) annotation. Two experienced radiologists, each with over 5 years of expertise in diagnostic imaging and unaware of the study's aims, initially demarcated the subcutaneous fat (SAT), visceral fat (VAT), sacrospinalis, and all abdominal muscles at the L1–S1 vertebral level (Figure 2). To further bolster the reliability of the data annotation, a seasoned diagnostic radiologist with over 15 years of experience, also unfamiliar with the study's objectives, was brought in to scrutinize the ROIs. This senior radiologist, well-versed in medical imaging and rich in clinical diagnostic acumen, meticulously reviewed and corrected the annotations made by the junior physicians; during the review, the senior radiologist engaged in profound discussions with the junior physicians regarding controversial or unclear areas. However, these discussions were restricted to technical matters and excluded any discourse on the patient's clinical conditions or the study's hypotheses. Consensus was achieved through negotiation to ensure that each ROI annotation was exact and precise. This blinded data annotation process helped minimize the influence of subjective bias on the study's outcomes.

2.4 Segmentation network

For the segmentation model, we adopted a 5-fold cross-validation approach with a data partition ratio of 5:1. Specifically, the dataset was initially randomly divided into five subsets, with four subsets (80%) serving as the training set and the remaining subset (20%) as the validation set. This procedure was replicated 5 times to ensure that each data point was used as a validation set exactly once. This ensures that we can maximize data utilization while more accurately evaluating the model's generalization ability.

A neural network architecture was employed for accurate image segmentation. The process began with the segmentation of the L3–L5 vertebrae. For ROI regions with smaller initial segmented areas and errors, erosion processing was applied to enhance their accuracy. Following this, connected component analysis was conducted, revealing that the segmentation accuracy of the L4 vertebra was the highest. Using this information, we expanded one connected component upward and one downward to accurately identify the L3–L5 vertebral region. Subsequently, the muscles and fat within this region were delineated.

Our segmentation network was constructed based on the nnU-Net architecture (Isensee et al., 2021). nnUNet can automatically perform preprocessing based on the characteristics of the dataset. Additionally, it offers various architectures that handle 3D matrices effectively, making it highly suitable for CT images. Our segmentation network exclusively used non-enhanced CT images as a data source.

2.5 Evaluation of ANP and IPN

For the assessment of ANP (Figures 3A–C), all case evaluations were conducted by two radiologists specializing in imaging diagnostics, each with over 8 years of diagnostic experience, who reviewed all imaging studies performed during the inpatient stay of the enrolled patients, devoid of any clinical information and adverse outcomes. Regarding the assessment of IPN (Figures 3D–F), an abdominal CT specialist with over a decade of diagnostic experience evaluated the cases by synthesizing clinical data, imaging findings, and laboratory test results. It is pertinent to note that IPN was defined as the initial percutaneous catheter drainage or surgical retrieval, yielding a positive culture or the observation of extraluminal gas on CT scans. Pancreatic necrosis refers to areas within the pancreatic parenchyma that exhibit hypoattenuation or lack of enhancement on CT imaging. Peripancreatic necrosis is characterized by collections containing varying amounts of fluid and necrotic tissue associated with necrotizing pancreatitis, and it can be diagnosed when non-liquid components of non-enhancing areas are visualized on CT scans.

2.6 Radiomics feature selection

Utilizing the Python package PyRadiomics version 3.0.1, we conducted the extraction of radiomic features from non-contrast CT images that had been segmented automatically. In recognition of the intrinsic clinical relevance of specific attributes, we narrowed our focus to two principal categories: first-order and shape characteristics. The first-order features encapsulate the distribution patterns of pixel intensities within the images (refer to Table 1, items 1–18), whereas the shape features concentrate on delineating the geometric attributes of the imaged structures (refer to Table 1, items 19–32).

2.7 Statistical analysis

Statistical computations were executed utilizing Statistical Package for the Social Sciences (SPSS) version 26.0. Quantitative data were expressed in terms of mean deviation $\left(\overline{x} \pm s\right)$ and were subjected to comparison via *t*-tests. Qualitative data were represented in frequencies and were evaluated using the χ^2/Fisher exact tests, as appropriate. In radiomic data, the *t*-tests were employed for data that exhibited a normal distribution. In contrast, the Mann–Whitney *U* test was utilized for data that did not conform to a normal distribution. $p < 0.05$ was deemed indicative of statistical significance.

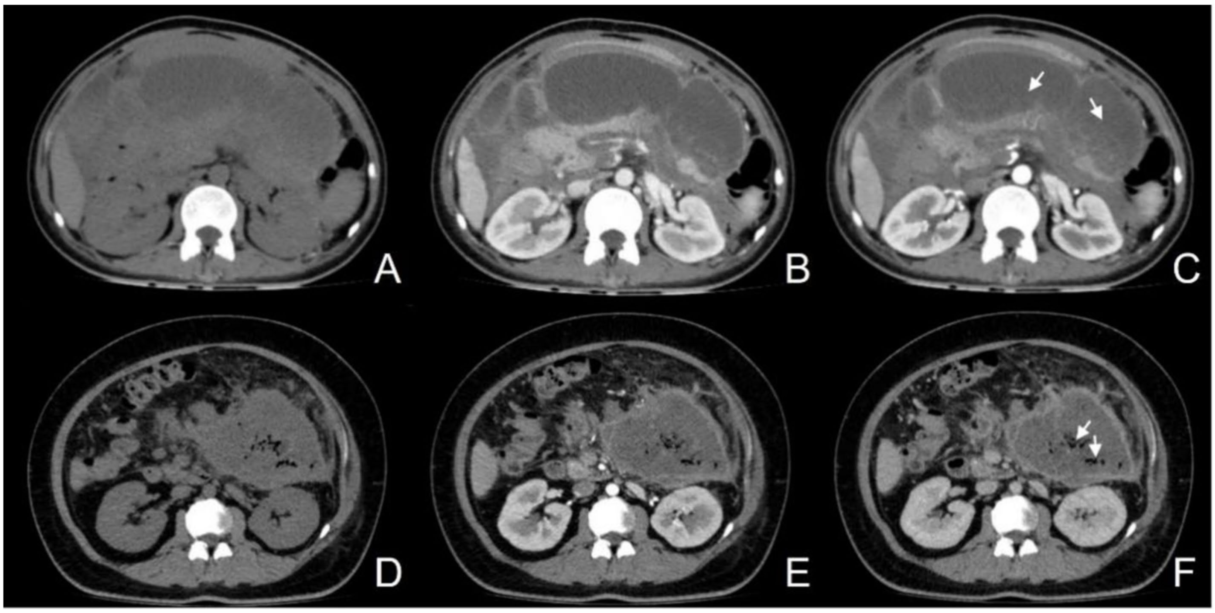


FIGURE 3
(A–C) Female, 50-year-old, acute necrotizing pancreatitis (ANP), showing a slightly hypodense lesion without enhancement after contrast (**C**, arrow); **(D–F)** Female, 31-year-old, infectious pancreatic necrosis (IPN), with scattered gas within the necrosis (**F**, arrow).

TABLE 1 Radiomics features and their mathematical explanations used in this study.

Number	Feature	Mathematical explanation
1	10 percentile	The 10th percentile value of the voxel intensities within the region of interest (ROI). This is the value below which 10% of the data falls.
2	90 percentile	The 90th percentile value of the voxel intensities within the ROI. This is the value below which 90% of the data falls.
3	Energy	A measure of the magnitude of voxel values in an image. It is the sum of the squares of the voxel values.
4	Entropy	Specifies the uncertainty/randomness in the image values. It measures the average amount of information required to encode the image values.
5	IQR	The difference between the 75th percentile (Q3) and the 25th percentile (Q1) of the voxel intensities within the ROI. It is a measure of the spread of the middle 50% of the data.
6	Kurtosis	A measure of the "peakedness" of the distribution of voxel intensities within the ROI. A high kurtosis value indicates a sharp peak and heavy tails, while a low kurtosis value indicates a flat distribution.
7	Maximum	The maximum voxel intensity value within the ROI.
8	Mean	The average voxel intensity value within the ROI.
9	MAD	The average of the absolute differences between the individual voxel intensities and the mean voxel intensity.
10	Median	The middle value of the voxel intensities within the ROI, such that half of the data is above and half is below this value.
11	Minimum	The minimum voxel intensity value within the ROI.
12	Range	The difference between the maximum and minimum voxel intensity values within the ROI.
13	Robust MAD	A measure of the spread of the data that is less sensitive to outliers than the MAD.
14	Root mean squared	The square root of the mean of the squares of the differences between the individual voxel intensities and the mean voxel intensity.
15	Skewness	A measure of the asymmetry of the distribution of voxel intensities within the ROI. A positive skewness indicates a tail on the right side of the distribution, while a negative skewness indicates a tail on the left side.
16	Total energy	The energy feature scaled by the volume of the voxel in cubic mm. It takes into account both the magnitude of the voxel values and the size of the ROI.
17	Uniformity	A measure of the homogeneity of the voxel intensities within the ROI. It is the sum of the squares of each voxel intensity value divided by the square of the sum of the voxel intensity values.
18	Variance	The average of the squared differences between the individual voxel intensities and the mean voxel intensity. It measures the spread of the data around the mean.
19	Elongation	The elongation of the ROI shape is a measure of the relationship between the two largest principal components of the ROI.
20	Flatness	The flatness of the ROI shape is a measure of the relationship between the largest and smallest principal components of the ROI.
21	Least axis length	The length of the smallest principal axis of the ROI.
22	Major axis length	The length of the largest principal axis of the ROI.
23	Maximum 2D diameter column	The maximum 2D diameter of the ROI in the column direction (typically the y-axis in an image).
24	Maximum 2D diameter row	The maximum 2D diameter of the ROI in the row direction (typically the x-axis in an image).
25	Maximum 2D diameter slice	The maximum 2D diameter of the ROI in the slice direction (typically the z-axis in a 3D image).
26	Maximum 3D diameter	The maximum 3D diameter of the ROI is the largest Euclidean distance between any two points on the surface of the ROI.
27	Mesh volume	The volume of the ROI is calculated from the triangular mesh that represents the surface of the ROI.
28	Minor axis length	The length of the second-largest principal axis of the ROI
29	Sphericity	A measure of how spherical the ROI is. It is the ratio of the surface area of a sphere with the same volume as the ROI to the actual surface area of the ROI.
30	Surface area	The surface area of the ROI.
31	Surface volume ratio	The ratio of the surface area of the ROI to its volume. A lower value indicates a more compact, spherical shape.
32	Voxel volume	The volume of a single voxel within the ROI.

The overall study workflow is shown in Figure 4.

3 Results

3.1 Clinical characteristics

The present study enrolled a total of 315 patients. Table 2 delineates the clinical attributes of the participants within the IPN and non-infected pancreatic necrosis (NPN) cohorts.

A comparative analysis of the clinical characteristics between the IPN ($n = 101$) and NPN ($n = 214$) groups revealed no significant disparities with respect to sex, diabetes, hypertension, hyperlipidemia, and coronary heart disease. However, patients in the IPN cohort were notably older ($p = 0.038$), exhibited a more significant requirement for mechanical ventilation ($p < 0.001$), and experienced significantly prolonged hospital admissions ($p < 0.001$) compared to their NPN counterparts. While a higher incidence of hypoxemia was observed among IPN patients, this discrepancy did not reach statistical significance. These observations proffer critical insights for subsequent inquiries into the pathophysiology and prognostic determinants of IPN.

3.2 Segmentation results

Following a comprehensive assessment across five validation datasets, utilizing a 5-fold cross-validation approach with each model undergoing 100 training epochs, the model exhibiting superior performance was designated as the definitive model (Figures 5–7). The optimal model achieved an accuracy of 0.91 in segmenting the L3–L5 vertebral region within the validation dataset.

Within the L3–L5 vertebral range, we executed segmentation of the SAT, VAT, sacrospinalis, and all abdominal musculature (Figure 8).

3.3 Correlation between body composition and infectious pancreatic necrosis

We conducted a detailed analysis of the first-order and shape characteristics of the segmented VAT, SAT, sacrospinalis, and all abdominal muscles. For the statistically significant features, we attempted to provide explanations and presented them in the form of box plots (Figure 9).

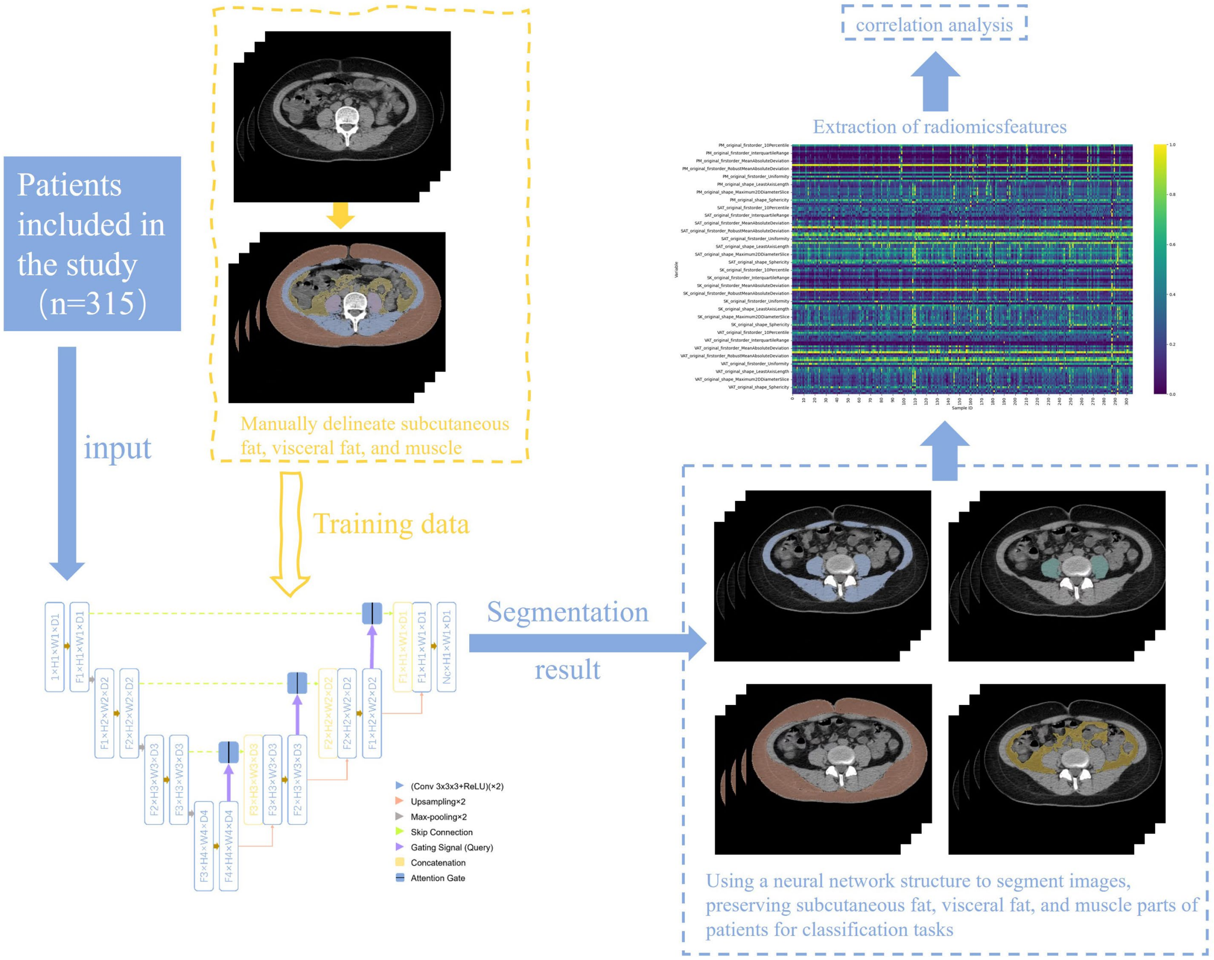


FIGURE 4
Study flowchart.

TABLE 2 **Comparison of clinical characteristics between infectious pancreatic necrosis (IPN) and non-infected pancreatic necrosis (NPN) groups.**

Clinical characteristics		IPN (*n* = 101)	NPN (*n* = 214)	*p*
Sex (*n*)	Male	70 (69.3%)	143 (66.8%)	0.66
	Female	31 (30.7%)	71 (33.2%)	
Age (years)		47.91 ± 15.45	44.14 ± 13.85	0.038
Diabetes mellitus (*n*)	Yes	44 (43.6%)	98 (45.8%)	0.71
	No	57 (56.4%)	116 (54.2%)	
Hypertension (*n*)	Yes	35 (34.7%)	61 (28.5%)	0.269
	No	66 (65.3%)	153 (71.5%)	
Hyperlipidemia (*n*)	Yes	44 (43.6%)	108 (50.5%)	0.252
	No	57 (56.4%)	106 (49.5%)	
Coronary heart disease (n)	Yes	5 (5.0%)	8 (3.7%)	0.762
	No	96 (95.0%)	206 (96.3%)	
Mechanical ventilation (*n*)	Yes	32 (31.7%)	19 (8.9%)	<0.001
	No	69 (68.3%)	195 (91.1%)	
Biliary pancreatitis (*n*)	Yes	13 (12.9%)	26 (12.1%)	0.856
	No	88 (87.1%)	188 (87.9%)	
Concurrent hypoxemia (*n*)	Yes	31 (30.7%)	49 (22.9%)	0.138
	No	70 (69.3%)	165 (77.1%)	
Length of hospital stay (days)		35.83 ± 33.06	19.71 ± 19.00	<0.001

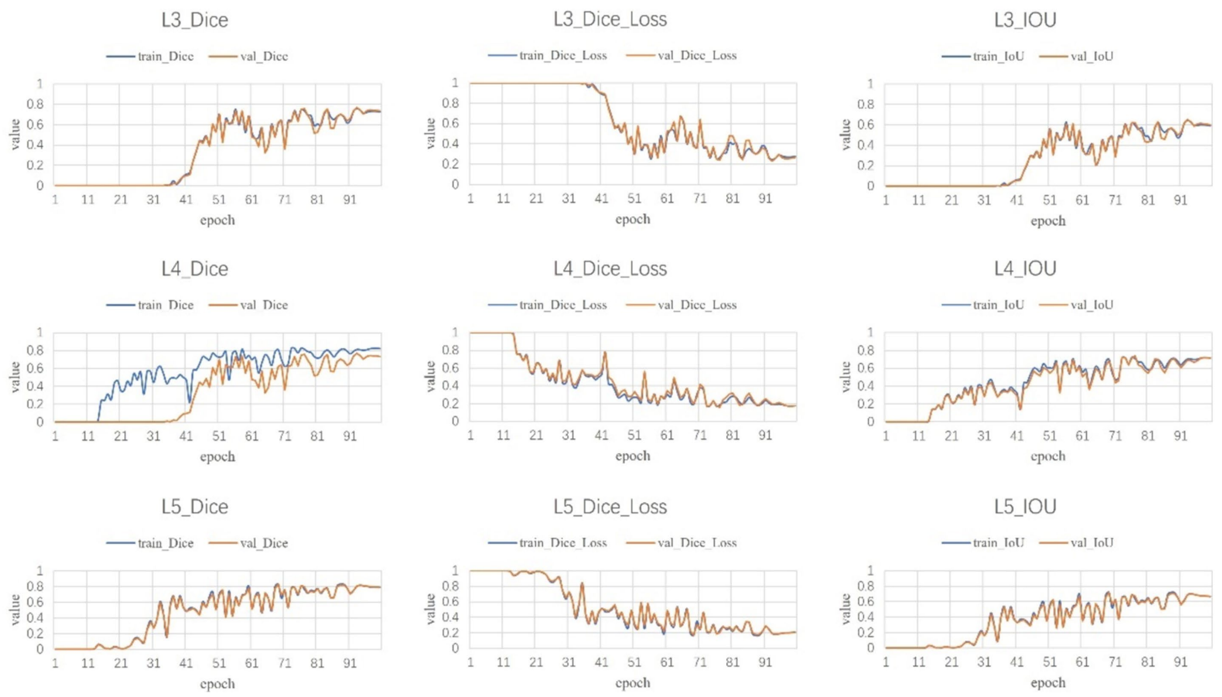


FIGURE 5
L3–L5 segmentation results.

3.3.1 VAT

As shown in Table 3, for the first-order VAT features, skewness ($p = 0.004$) and uniformity ($p = 0.036$) were statistically significant; the other features were not.

The skewness feature of VAT is correlated with the symmetry of VAT pixel distribution, with higher skewness values indicating excessive fat accumulation in certain areas and relatively less in others; within this set of features, skewness is 0.60 in the negative group, and lower in the positive group (0.48), suggesting that VAT in IPN patients exhibits less localized accumulation.

The uniformity feature pertains to whether VAT is evenly distributed within the ROI, with values closer to 1

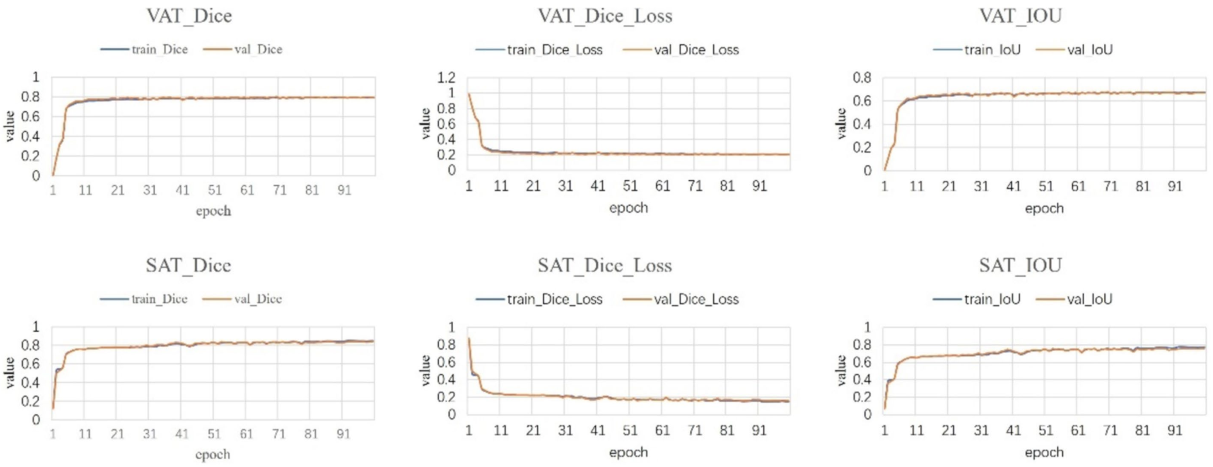


FIGURE 6
Fat segmentation results.

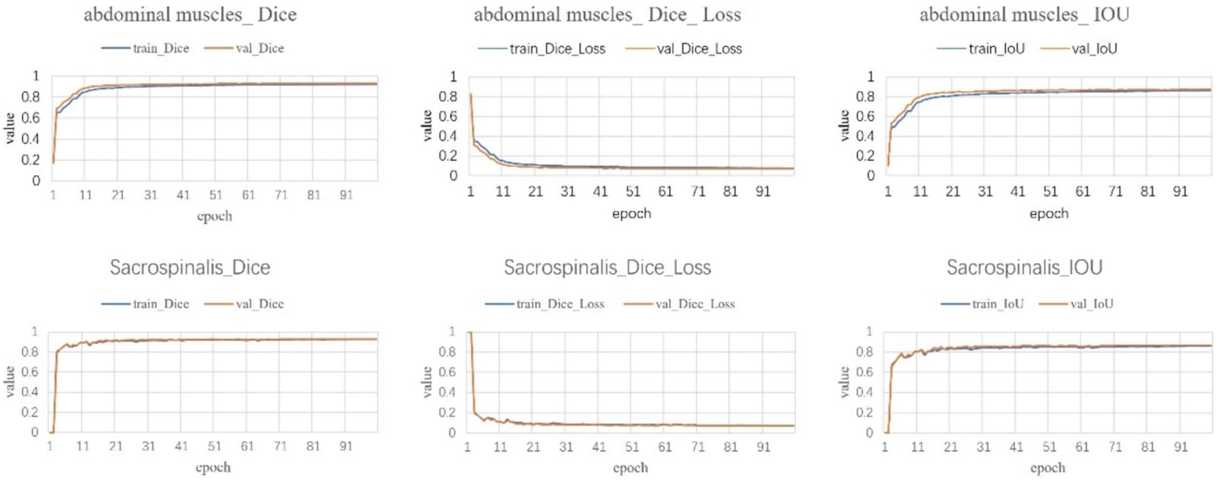


FIGURE 7
Muscle segmentation results.

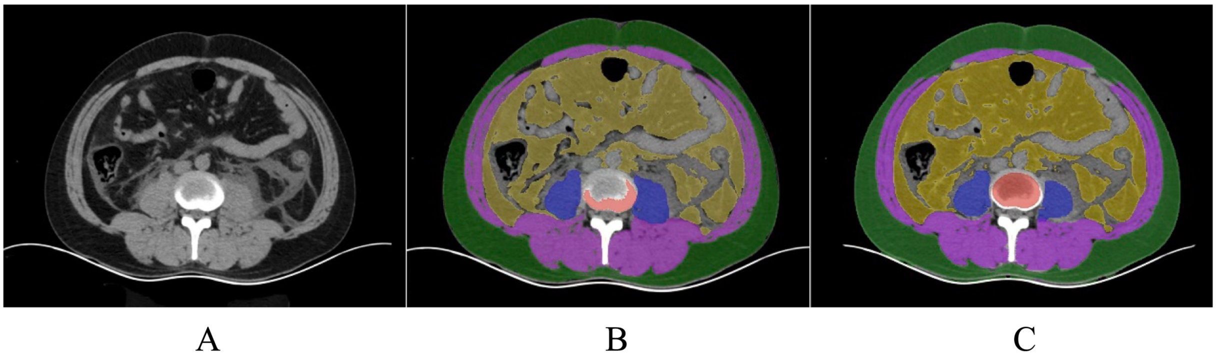


FIGURE 8
(A) Original image of a patient with pancreatitis. **(B)** A senior radiologist with over 15 years of experience manually delineates the ROI for assessing the effectiveness of segmenting human body components in patients with pancreatitis under human visualization (using a model not trained for pancreatitis). **(C)** Model-segmented ROI.

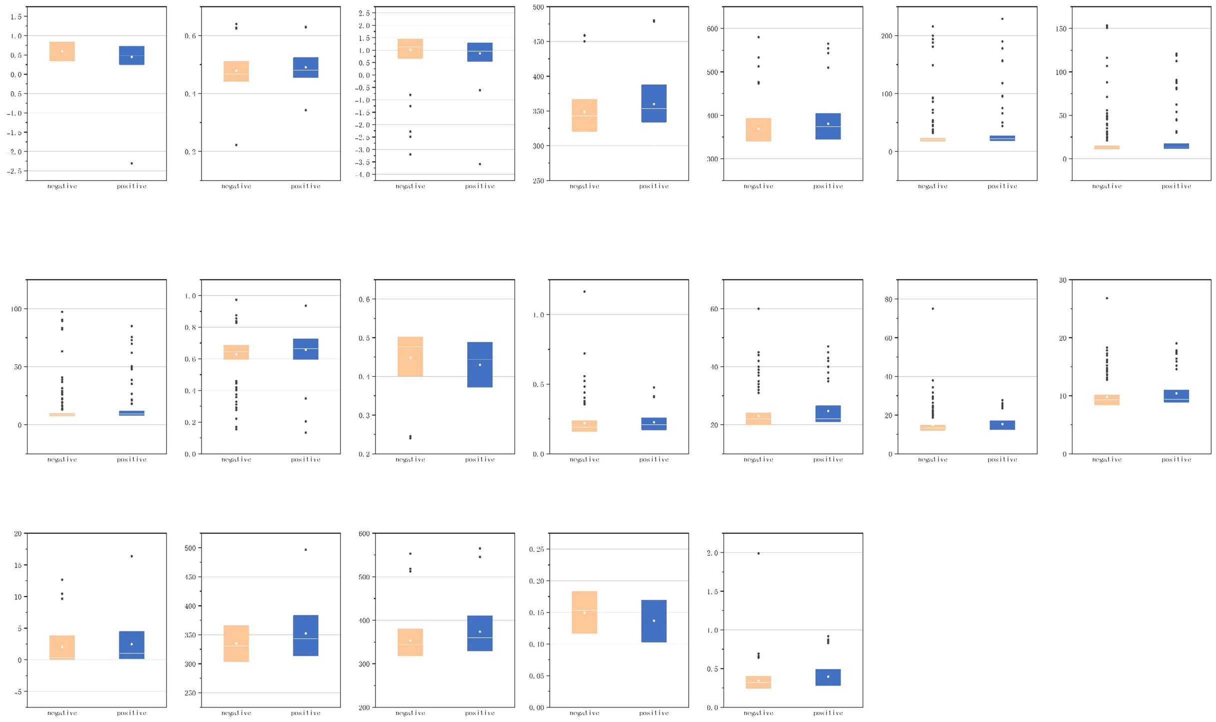


FIGURE 9
Radiomic features with statistical significance in body composition.

indicating a more uniform texture and values farther from 1 suggesting greater heterogeneity. The negative group exhibits slightly lower uniformity than the positive group (0.47 compared to 0.48), indicating that the VAT density in IPN patients is more uniform.

3.3.2 SAT

As shown in Table 4, for the first-order features of SAT, skewness ($p = 0.023$) showed statistical significance; among the shape features, maximum two-dimensional (2D) diameter slice ($p = 0.020$) and maximum three-dimensional (3D) diameter ($p = 0.044$) were statistically significant; the other features were not.

The skewness feature of SAT is higher in the negative group (1.14) compared to the positive group (0.96), indicating that SAT in IPN patients has less localized accumulation.

The maximum 2D diameter slice measures the maximum diameter of the SAT area in a 2D image, reflecting the extent of SAT expansion in the axial plane; the average maximum 2D diameter slice in the positive group (362.97) is greater than that in the negative group (350.19). The maximum 3D diameter reflects the overall size and shape of SAT in 3D space. Similarly, the maximum 3D diameter in the positive group (374.06) is larger than that in the negative group (358.72). Both features suggest that the volume of subcutaneous adipose tissue in IPN patients is larger relative to NPN patients.

3.3.3 Abdominal muscles

As shown in Table 5, for the abdominal muscles' first-order features, interquartile range ($p = 0.023$), mean absolute deviation ($p = 0.039$), and robust mean absolute deviation ($p = 0.015$) were statistically significant; among the shape features, elongation ($p = 0.025$), sphericity ($p = 0.010$), and surface volume ratio ($p = 0.014$) were statistically significant; and the other features were not.

The interquartile range (IQR) of the abdominal muscles reflects their stability or variability under different conditions, describing the degree of dispersion in data distribution. It represents the range of the middle 50% of the data. The IQR in the negative group is less than that in the positive group (20.00 vs. 22.00). This may indicate that in patients who develop infectious pancreatic necrosis, the signal intensity distribution of the abdominal muscles is more dispersed.

Mean absolute deviation (MAD) is another statistical measure of data distribution dispersion, which quantifies the average distance of data points from the mean. The robust mean absolute deviation is a more robust version of MAD, insensitive to outliers. The findings of these two features are consistent, indicating that the signal intensity distribution of the abdominal muscles in patients who develop infectious pancreatic necrosis may be more uneven.

Elongation reflects the extent of longitudinal extension of muscle fibers or muscle blocks, with the elongation rate in the positive group being greater than that in the negative group (0.67 vs. 0.64). This may suggest that the shape of the abdominal muscles in patients who develop infectious pancreatic necrosis is more elongated in one direction.

Sphericity measures the similarity of the shape of an ideal sphere, reflecting the compactness and regularity of the shape of muscle fibers or muscle blocks; the sphericity in the negative group is greater than that in the positive group (0.48 vs. 0.44). This may imply that the shape of the abdominal muscles in patients who do not develop infectious

TABLE 3 Correlation between visceral adipose tissue (VAT) radiomic features and infectious pancreatic necrosis (IPN).

Feature	Negative	Positive	Statistical methods	Statistic	*p*
10 Percentile	−113.00 (−118.00, −107.00)	−112.27	Mann–Whitney U	9057.5	0.087
90 Percentile	−50.00 (−58.00, −44.00)	−50.75	Mann–Whitney U	9560.5	0.308
Energy	5559141605.50 (3540920150.50, 7802701555.75)	5517645106.00 (3029784775.75, 8244608979.00)	Mann–Whitney U	10,517	0.766
Entropy	1.26 (1.19, 1.32)	1.23 (1.17, 1.30)	Mann–Whitney U	11,721	0.05
IQR	31.00 (28.00, 35.00)	31.73	Mann–Whitney U	10,240	0.935
Kurtosis	3.05 (2.65, 3.72)	2.85 (2.55, 3.43)	Mann–Whitney U	11,718	0.051
Maximum	30.00 (16.00, 41.75)	21.50 (12.00, 41.25)	Mann–Whitney U	11,696.5	0.054
Mean	−84.93	−84.93	*t*-test	1.78111074	0.076
MAD	18.87 (17.33, 20.14)	18.47 (17.20, 20.04)	Mann–Whitney U	10,971	0.356
Median	−89.00 (−96.00, −82.00)	−88.61	Mann–Whitney U	8,889	0.052
Minimum	−233.00 (−301.50, −187.25)	−220.50 (−282.00, −175.00)	Mann–Whitney U	9,205	0.132
Range	268.00 (219.00, 337.25)	247.50 (202.75, 317.75)	Mann–Whitney U	11,636	0.066
Robust MAD	13.40 (11.99, 14.55)	13.24 (11.79, 14.80)	Mann–Whitney U	10,482	0.803
Root mean squared	88.78 (82.43, 94.47)	88.32	Mann–Whitney U	11,603	0.073
Skewness	0.6	0.48 (0.26, 0.72)	Mann-–Whitney U	12,390	0.004
Total energy	12114180613.50 (8090181675.25, 17111404744.75)	12041332397.50 (6156355601.50, 18040494741.50)	Mann–Whitney U	10,499	0.785
Uniformity	0.47 (0.44, 0.51)	0.48 (0.46, 0.52)	Mann–Whitney U	8,779	0.036
Variance	556.66 (488.00, 616.44)	527.72 (467.94, 606.80)	Mann–Whitney U	11,270	0.182
Elongation	0.61 (0.54, 0.68)	0.61 (0.56, 0.67)	Mann–Whitney U	9,754	0.452
Flatness	0.40 (0.35, 0.45)	0.40 (0.36, 0.47)	Mann–Whitney U	9,526	0.287
Least axis length	110.84 (104.00, 121.47)	111.99 (105.00, 140.59)	Mann–Whitney U	9,345	0.189
Major axis length	285.63 (264.87, 311.07)	291.49 (277.58, 313.18)	Mann–Whitney U	9,020	0.078
Maximum 2D diameter column	288.81 (268.45, 308.03)	292.30 (275.25, 314.29)	Mann–Whitney U	9,355	0.193
Maximum 2D diameter row	205.30 (182.48, 229.82)	208.38 (187.62, 233.90)	Mann–Whitney U	9,794	0.486
Maximum 2D diameter slice	276.66 (257.45, 300.00)	284.43 (267.16, 301.85)	Mann–Whitney U	9,070	0.09
Maximum 3D diameter	295.28 (272.07, 315.35)	296.97 (279.89, 326.37)	Mann–Whitney U	9,335	0.184
Mesh volume	1594888.51 (1116303.11, 2059782.88)	1565850.12 (994337.63, 2122169.49)	Mann–Whitney U	10,245	0.94
Minor axis length	176.57 (154.86, 198.44)	177.83 (157.36, 205.76)	Mann–Whitney U	9,538	0.294
Sphericity	0.19 (0.16, 0.21)	0.18	Mann–Whitney U	11,307	0.166
Surface area	348265.70 (276579.17, 415143.35)	345724.42 (280669.46, 450891.29)	Mann–Whitney U	9,911	0.593
Surface volume ratio	0.23 (0.19, 0.29)	0.24 (0.20, 0.31)	Mann–Whitney U	9,719	0.424
Voxel volume	1595786.56 (1117477.57, 2062007.94)	1577308.92 (998146.19, 2124513.29)	Mann–Whitney U	10,240	0.935

pancreatic necrosis is closer to spherical. In contrast, the shape of the abdominal muscles in patients who develop infectious pancreatic necrosis may become less regular due to inflammation or other pathological changes.

The surface volume ratio reflects the ratio of muscle surface area to volume, with the positive group having a higher surface volume ratio than the negative group (0.21 vs. 0.19). This may indicate that the abdominal muscles of patients who develop

TABLE 4 **Correlation between subcutaneous adipose tissue (SAT) radiomic features and infectious pancreatic necrosis (IPN).**

Feature	Negative	Positive	Statistical methods	Statistic	*p*
10 Percentile	−116.00 (−121.00, −111.00)	−116.00 (−120.00, −110.00)	Mann–Whitney U	9,656	0.453
90 Percentile	−61.00 (−72.00, −52.00)	−60.59	Mann–Whitney U	9,035	0.107
Energy	7382051504.50 (4423188997.75, 10268867202.75)	7420446548.00 (4086489735.00, 10569815384.00)	Mann–Whitney U	10,304	0.883
Entropy	1.23 (1.13, 1.32)	1.23	Mann–Whitney U	9,958	0.741
IQR	25.00 (21.00, 31.00)	27.00 (22.00, 31.50)	Mann-–Whitney U	8,953.5	0.084
Kurtosis	5.48 (4.11, 7.55)	5.12 (3.45, 6.55)	Mann–Whitney U	11,508	0.069
Maximum	57.50 (35.25, 125.75)	48.00 (20.00, 104.00)	Mann–Whitney U	11,139	0.192
Mean	−93.04 (−98.62, −87.12)	−92.63 (−97.19, −82.42)	Mann–Whitney U	9,097	0.127
MAD	17.07 (14.82, 19.31)	17.39	Mann–Whitney U	9,674	0.469
Median	−98.00 (−103.75, −92.25)	−97.00 (−101.00, −87.00)	Mann–Whitney U	8,987.5	0.093
Minimum	−301.00 (−416.00, −195.50)	−296.00 (−383.00, −203.50)	Mann–Whitney U	9,966.5	0.75
Range	392.50 (243.25, 536.25)	350.00 (237.50, 511.00)	Mann–Whitney U	10,705	0.482
Robust MAD	10.93 (9.08, 13.27)	11.4	Mann–Whitney U	9,067	0.117
Root mean squared	95.87 (90.82, 100.83)	95.92 (86.10, 99.67)	Mann–Whitney U	11,329	0.117
Skewness	1.14 (0.67, 1.44)	0.96 (0.55, 1.27)	Mann–Whitney U	11,835	0.023
Total energy	15276971006.50 (10189081819.00, 22145540431.25)	15458624133.00 (10131235701.00, 22930332683.00)	Mann–Whitney U	10,199	0.998
Uniformity	0.47 (0.45, 0.52)	0.47 (0.44, 0.50)	Mann–Whitney U	10,760	0.435
Variance	508.17 (410.41, 646.90)	516.61 (397.99, 650.74)	Mann–Whitney U	10,310	0.876
Elongation	0.78 (0.73, 0.83)	0.79 (0.72, 0.84)	Mann–Whitney U	10,009	0.795
Flatness	0.26 (0.24, 0.30)	0.27 (0.24, 0.34)	Mann–Whitney U	9,773	0.557
Least axis length	112.77 (105.52, 124.29)	113.79 (105.34, 146.02)	Mann–Whitney U	9,525	0.352
Major axis length	433.34 (402.80, 460.00)	434.86	Mann–Whitney U	9,442	0.295
Maximum 2D diameter column	350.19 (334.44, 383.90)	362.97 (338.26, 397.18)	Mann–Whitney U	8,978	0.091
Maximum 2D diameter row	283.24 (263.04, 305.95)	286.41 (261.52, 314.55)	Mann–Whitney U	9,870	0.651
Maximum 2D diameter slice	342.91 (320.82, 366.55)	353.30 (334.67, 385.09)	Mann–Whitney U	8,521	0.02
Maximum 3D diameter	358.72 (340.51, 392.78)	374.06 (345.72, 403.56)	Mann–Whitney U	8,742	0.044
Mesh volume	1653557.96 (1235842.07, 2195203.36)	1680448.40 (1204551.02, 2591183.65)	Mann–Whitney U	9,960	0.743
Minor axis length	337.98 (312.61, 373.83)	338.29	Mann–Whitney U	9,658	0.455
Sphericity	0.25	0.25	*t*-test	1.448543564	0.148
Surface area	255759.30 (220992.29, 311631.51)	267878.28 (219325.30, 333155.36)	Mann–Whitney U	9,592	0.402
Surface volume ratio	0.16 (0.13, 0.21)	0.17 (0.12, 0.22)	Mann–Whitney U	9,590	0.4
Voxel volume	1654703.97 (1238486.14, 2195450.68)	1680171.98 (1205763.96, 2590036.86)	Mann–Whitney U	9,973	0.757

IQR, interquartile range; MAD, mean absolute deviation.

infectious pancreatic necrosis have a relatively larger surface area in proportion to volume, possibly reflecting tissue edema due to the inflammatory response.

3.3.4 Sacrospinalis

As shown in Table 6, for the first-order features of sacrospinalis, interquartile range ($p = 0.018$), mean absolute deviation

TABLE 5 Correlation between abdominal muscles radiomic features and infectious pancreatic necrosis (IPN).

Feature	Negative	Positive	Statistical methods	Statistic	*p*
10 Percentile	30.00 (25.00, 35.00)	28.50 (21.00, 35.00)	Mann–Whitney U	11,216	0.207
90 Percentile	69.00 (64.00, 75.00)	69.00 (62.00, 78.00)	Mann–Whitney U	10,101.5	0.785
Energy	239309670.50 (155707966.75, 357551615.00)	241836573.50 (122022174.50, 368706009.75)	Mann–Whitney U	10,375	0.918
Entropy	1.04 (0.99, 1.09)	1.04 (0.99, 1.15)	Mann–Whitney U	9,891	0.574
IQR	20.00 (18.00, 23.00)	22.00 (19.00, 27.00)	Mann–Whitney U	8,649	0.023
Kurtosis	4.55 (3.81, 8.97)	4.80 (3.84, 9.29)	Mann–Whitney U	10,139	0.825
Maximum	147.50 (115.00, 351.75)	171.00 (119.75, 510.00)	Mann–Whitney U	9,185	0.125
Mean	49.46 (44.72, 54.10)	50.00 (42.57, 54.91)	Mann–Whitney U	10,608	0.672
MAD	12.64 (11.20, 14.75)	13.21 (11.66, 17.42)	Mann–Whitney U	8,799	0.039
Median	50.00 (45.25, 54.00)	50.50 (43.00, 55.00)	Mann–Whitney U	10,501.5	0.782
Minimum	−63.50 (−75.00, −50.00)	−64.00 (−82.50, −51.50)	Mann–Whitney U	10,818	0.476
Range	213.00 (179.25, 429.00)	247.50 (184.75, 672.50)	Mann–Whitney U	9,092	0.096
Robust MAD	8.58 (7.67, 9.77)	9.19 (8.12, 11.61)	Mann–Whitney U	8,527	0.015
Root mean squared	52.48 (47.75, 56.90)	53.14 (45.53, 58.53)	Mann–Whitney U	10,322	0.976
Skewness	−0.29 (−0.51, 0.31)	−0.23 (−0.47, 0.72)	Mann–Whitney U	9,782	0.476
Total energy	504705905.60 (331904476.82, 745238186.75)	505832574.90 (277667575.85, 814704451.82)	Mann–Whitney U	10,407	0.883
Uniformity	0.51 (0.49, 0.54)	0.50 (0.49, 0.55)	Mann–Whitney U	10,606	0.674
Variance	274.30 (214.07, 396.96)	295.51 (227.77, 519.59)	Mann–Whitney U	8,972	0.067
Elongation	0.64 (0.60, 0.69)	0.67 (0.60, 0.72)	Mann–Whitney U	8,669	0.025
Flatness	0.23 (0.20, 0.26)	0.23 (0.21, 0.28)	Mann–Whitney U	9,539	0.295
Least axis length	43.31 (36.50, 49.07)	42.02 (36.22, 51.25)	Mann–Whitney U	10,513	0.77
Major axis length	178.21 (166.32, 199.38)	175.26 (166.77, 199.52)	Mann–-Whitney U	10,828	0.467
Maximum 2D diameter column	149.23 (140.63, 186.94)	147.74 (138.45, 201.81)	Mann–Whitney U	10,701	0.581
Maximum 2D diameter row	117.22 (109.44, 134.28)	118.61 (107.65, 148.15)	Mann–Whitney U	9,896	0.578
Maximum 2D diameter Slice	140.85 (132.95, 159.86)	139.52 (131.26, 163.66)	Mann–Whitney U	10,818	0.476
Maximum 3D diameter	163.09 (153.48, 222.48)	161.51 (152.06, 226.11)	Mann–Whitney U	10,598	0.682
Mesh volume	170053.90 (124849.08, 226962.00)	158223.70 (114210.53, 203752.58)	Mann–Whitney U	11,066	0.292
Minor axis length	114.68 (107.51, 125.67)	115.75 (105.68, 144.08)	Mann–Whitney U	9,822	0.511
Sphericity	0.48 (0.40, 0.50)	0.44 (0.37, 0.49)	Mann–Whitney U	12,166	0.01
Surface area	32950.71 (26557.99, 38100.90)	32240.91 (25502.50, 39481.43)	Mann–Whitney U	10,503	0.78
Surface volume ratio	0.19 (0.16, 0.24)	0.21 (0.17, 0.26)	Mann–Whitney U	8,516	0.014
Voxel volume	170378.76 (125078.19, 227293.64)	158435.95 (114481.55, 203957.85)	Mann–Whitney U	11,068	0.29

IQR, interquartile range; MAD, mean absolute deviation.

($p = 0.049$), robust mean absolute deviation ($p = 0.025$), and skewness ($p = 0.008$) exhibited statistical significance; among the shape features, maximum 2D diameter slice ($p = 0.008$), maximum 3D diameter ($p = 0.005$), sphericity ($p = 0.011$), and surface volume ratio ($p = 0.005$) were statistically significant; the other features were not.

The interquartile range of the sacrospinalis is equal in the median for both groups, but the confidence interval for the positive group is

TABLE 6 Correlation between sacrospinalis radiomic features and infectious pancreatic necrosis (IPN).

Feature	Negative	Positive	Statistical methods	Statistic	*p*
10 Percentile	21.00 (16.00, 25.00)	21.22	Mann–Whitney U	11,298	0.169
90 Percentile	63.00 (58.00, 69.00)	63.50 (57.75, 72.50)	Mann–Whitney U	9,998	0.678
Energy	1249152060.00 (764619051.75, 1994064418.25)	1187720175.50 (668420562.50, 2087340047.00)	Mann–Whitney U	10,681	0.6
Entropy	0.99 (0.86, 1.04)	1.00 (0.85, 1.09)	Mann–Whitney U	9,841	0.528
IQR	22.00 (20.00, 24.00)	22.00 (21.00, 26.25)	Mann–Whitney U	8,587.5	0.018
Kurtosis	10.42 (5.55, 61.15)	23.54 (5.97, 66.70)	Mann–Whitney U	9,273	0.157
Maximum	526.50 (382.25, 756.25)	528.50 (402.00, 796.25)	Mann–Whitney U	9,686.5	0.398
Mean	43.95 (38.19, 48.39)	43.25 (37.46, 49.67)	Mann–Whitney U	10,389	0.903
MAD	13.25 (12.16, 14.78)	13.39 (12.62, 16.81)	Mann–Whitney U	8,868	0.049
Median	45.00 (39.00, 49.00)	45.2	Mann–Whitney U	10,492.5	0.791
Minimum	−44.00 (−57.00, −34.00)	−44.00 (−58.25, −29.00)	Mann–Whitney U	10,068.5	0.75
Range	579.50 (415.50, 814.25)	571.50 (445.75, 847.75)	Mann–Whitney U	9,730.5	0.433
Robust MAD	9.29 (8.43, 10.13)	9.40 (8.92, 10.94)	Mann–Whitney U	8,671	0.025
Root mean squared	47.21 (41.66, 51.93)	47.15 (41.68, 55.19)	Mann–Whitney U	10,241	0.936
Skewness	0.32 (0.01, 3.77)	1.01 (0.19, 4.42)	Mann–Whitney U	8,375	0.008
Total energy	2847674195.50 (1720109062.25, 4205653583.50)	2575449699.50 (1406155363.00, 4910820334.25)	Mann–Whitney U	10,661	0.619
Uniformity	0.54 (0.50, 0.63)	0.54 (0.50, 0.63)	Mann–Whitney U	10,773	0.515
Variance	286.91 (244.39, 468.43)	307.96 (256.65, 654.50)	Mann–Whitney U	9,045	0.084
Elongation	0.68	0.68	*t*-Test	1.31513754	0.189
Flatness	0.32 (0.29, 0.37)	0.32 (0.29, 0.42)	Mann–Whitney U	10,156	0.843
Least axis length	108.65 (101.74, 120.38)	109.74 (102.19, 151.06)	Mann–Whitney U	9,466	0.251
Major axis length	346.65 (320.52, 374.80)	356.08 (327.14, 381.18)	Mann–Whitney U	8,989	0.071
Maximum 2D diameter column	331.83 (307.16, 355.28)	334.92 (309.67, 366.12)	Mann–Whitney U	9,127	0.106
Maximum 2D diameter row	263.36 (238.13, 296.13)	265.74 (231.65, 304.67)	Mann–Whitney U	10,007	0.687
Maximum 2D diameter slice	330.38 (303.91, 365.72)	342.84 (313.74, 382.86)	Mann–Whitney U	8,369.5	0.008
Maximum 3D diameter	343.67 (318.46, 379.83)	359.58 (329.37, 408.87)	Mann–Whitney U	8,284	0.005
Mesh volume	1211160.13 (843465.21, 1518283.12)	1119268.25 (755711.11, 1583688.32)	Mann–Whitney U	10,954	0.368
Minor axis length	234.46 (212.14, 258.54)	238.73 (205.35, 263.39)	Mann–Whitney U	9,996	0.676
Sphericity	0.15 (0.12, 0.18)	0.15	Mann–Whitney U	12,137	0.011
Surface area	324534.33 (273229.10, 450328.40)	338175.63 (268258.35, 528150.56)	Mann–Whitney U	9,860	0.545
Surface volume ratio	0.32 (0.25, 0.40)	0.35 (0.28, 0.49)	Mann–Whitney U	8,247	0.005
Voxel volume	1211242.46 (846322.89, 1519310.32)	1117691.58 (755909.58, 1586327.08)	Mann–Whitney U	10,966	0.359

IQR, interquartile range; MAD, mean absolute deviation.

greater than that for the negative group; the results of MAD and robust MAD are consistent with those of the abdominal muscles, indicating a more uneven signal intensity distribution in the IPN group; the median Sphericity is equal for both groups, but the confidence interval for the positive group is smaller than that for the negative group; the surface volume ratio for the positive group is greater than that for the negative group (0.35 vs. 0.32). These findings are consistent with the overall abdominal muscles.

The skewness feature of the sacrospinalis shows statistical significance, whereas there is no statistical difference in the abdominal muscles, with the median of the positive group being greater than that of the negative group (1.01 vs. 0.32), indicating that the intensity distribution of the sacrospinalis in IPN patients may be more uneven.

For the sacrospinalis, the maximum 2D diameter slice and maximum 3D diameter exhibit statistical significance, whereas there is no statistical difference for the abdominal muscles. These features have significant discriminative or representative value in assessing the morphology, structure, or function of the sacrospinalis. Both features are greater in the positive group than the negative group, indicating that the volume of the sacrospinalis in IPN patients is larger in NPF patients.

4 Discussion

IPN, as a severe complication of ANP, poses a significant clinical challenge due to its high mortality rate and incidence (Li et al., 2022). However, prolonged antibiotic use in the absence of infection may lead to multidrug-resistant bacterial infections, further increasing mortality (Lu et al., 2022). Therefore, early and accurate diagnosis of IPN followed by timely and effective treatment measures is crucial. This study delineated muscles and fat tissues using deep learning techniques based on non-contrast CT images. Subsequently, we extracted 18 first-order features and 14 shape features using radiomic techniques and conducted a correlation analysis. Notably, we provided detailed medical interpretations for the statistically significant features, revealing the potential physiological and pathological significance behind these features. This study highlights the important role of body composition, including muscles and fat tissues, in acute pancreatitis. It provides a quantitative evaluation tool based on highly standardized and readily available non-contrast CT data using artificial intelligence.

The application of deep learning enables precise segmentation of muscle and fat (Graffy et al., 2019; Shen et al., 2023), which is crucial for delving into the complexity of infectious diseases. It not only enhances our comprehension of the disease's pathophysiological mechanisms but also significantly enriches the means of assessing patients' nutritional status and inflammatory responses. For instance, the study by Zhang et al. explored radiomic features from CT images of 1,245 adrenal glands and surrounding fat tissues, strongly demonstrating their correlation with disease progression in COVID-19 patients (Zhang M. et al., 2023). Similarly, Yoo et al. utilized deep learning techniques to quantify liver and spleen volumes, as well as SAT and VAT tissues and skeletal muscle indices, providing valuable prognostic information for patients with chronic hepatitis B (CHB) (Yoo et al., 2023). Compared to traditional assessment methods (such as modified CT severity index [MCTSI], Ranson, bedside index for severity in acute pancreatitis [BISAP], etc.), these methods, while concise and easy to understand, have limitations due to the limited factors they consider, making it difficult to fully reflect the complexity and dynamic changes of diseases. Previous studies have primarily focused on the pancreas itself (Zhang C. et al., 2024; Xue et al., 2023; Lin et al., 2020; Chen et al., 2023). In our study, the detailed evaluation of VAT, SAT, abdominal muscles, and sacrospinalis not only deepened our understanding of the disease's pathophysiological mechanisms but also significantly improved our ability to assess patients' nutritional status and inflammatory responses. Through in-depth analysis of radiomic data, we found that some parameters showed a good correlation with disease states, further validating the reliability and effectiveness of these body composition indicators as disease assessment tools. This discovery not only emphasizes the importance of muscle and fat segmentation in the evaluation of infectious diseases but also lays a solid foundation for the future development of more precise and comprehensive disease assessment systems.

The World Health Organization defines obesity as a pathological condition characterized by excessive accumulation of body fat, with a body mass index (BMI) of ≥ 30 kg/m^2 (Whitlock et al., 2009), which exerts certain effects on inflammation (Ponce-de-Leon et al., 2022). Numerous epidemiological studies and meta-analyses have demonstrated that obesity is a prognostic factor affecting the severity of acute pancreatitis (AP) (Chen et al., 2012; Martínez et al., 2006; Cruz-Monserrate et al., 2016; Wang et al., 2011). This study found significant correlations between certain radiomic features of SAT and VAT and the occurrence of IPN: our results suggest that for patients developing IPN, less localized accumulation of VAT and SAT, uniform density of VAT and SAT, and a larger volume of SAT are risk factors. These findings may be related to age and sex (Zhou et al., 2022; Lizcano and Guzmán, 2014; Tchernof and Després, 2013; Frank et al., 2019; Palmer and Kirkland, 2016; Pascot et al., 1999): under the influence of sex hormones, men tend to accumulate fat tissue predominantly in visceral regions, while women accumulate it more subcutaneously, leading to symmetrical differences in fat distribution between genders; with advancing age, the distribution of fat in both genders also changes, which may affect the localized accumulation and uniformity of VAT density, thereby explaining why adjusted body composition parameters based on age and sex in previous studies could more accurately predict the severity of AP and avoid the fat paradox (Horibe et al., 2022). Furthermore, we hypothesize that the statistical significance of skewness features may also be associated with the distribution of fat necrosis; if so, our study's results suggest that the inflammatory or necrotic process is more diffused in the adipose tissue of IPN patients, or there may be more synchronized pathological changes. Our conclusions indicate that the volume of SAT has a positive effect on the occurrence of IPN, and we speculate that the increased volume of subcutaneous adipose tissue may be related to enhanced or systemic inflammatory responses, which could exacerbate the infection of necrotic pancreatic tissue.

Skeletal muscle, as a pivotal component of the human body, not only supports fundamental motor functions but is also closely associated with individual metabolic status, inflammatory responses, and long-term disease prognosis (Akturk et al., 2021; Picca and Calvani, 2021; Modesto et al., 2020). In this study, we conducted an in-depth analysis of the sacrospinalis muscle and all abdominal muscles, including the sacrospinalis. Patients with IPN exhibited scattered and heterogeneous muscle density distributions in the overall abdominal muscles and the sacrospinalis, a significant difference compared to the NPN population. This alteration was manifested in an increased surface area-to-volume ratio and a morphological deviation from the ideal spherical structure. These findings suggest that such changes in muscle tissue may represent an adaptive response to the pathological state of IPN, and we have posited several hypotheses: First, the dispersion and heterogeneity of muscle density might be related to the level of inflammation within IPN patients. Inflammatory responses could lead to alterations in the intramuscular environment, affecting muscle cell growth, metabolism, and extracellular matrix remodeling (Tu and Li, 2023); second, the increased volume of the sacrospinalis, particularly in IPN patients, may reflect compensatory changes in muscle tissue during disease progression. The increase in muscle volume might be an adaptation to the additional load imposed by inflammation and metabolic disturbances, and it may also represent an attempt by the body to maintain essential physiological functions. The significant differences in skewness features indicate that the muscle density distribution in IPN patients deviates from the normal range, and the increased asymmetry may be related to the damage and repair processes in muscle tissue. During the continuous self-repair of muscle tissue, structural and functional asymmetries may arise due to the influence of an inflammatory environment.

While this study has made certain progress in body composition assessment and IPN prediction based on non-contrast CT, there are still some limitations. First, this is a single-center study with a relatively limited sample size, which may limit the generalizability of the findings. Second, this study did not quantitatively assess the degree of pancreatic infection, which may affect the in-depth understanding of the mechanism of IPN occurrence. Additionally, this study did not provide a detailed analysis of patient prognosis and length of hospital stay; future studies can further explore the relationship between body composition and the prognosis of pancreatitis patients. Finally, due to the unclear boundaries of pancreatitis, this study did not segment pancreatic lesions, which may limit the in-depth exploration of the pathophysiological mechanisms of pancreatitis itself.

5 Conclusion

This study, utilizing deep learning techniques in conjunction with unenhanced CT imaging, has elucidated the close association between muscle and fat tissue and the progression of ANP to IPN, providing a novel tool for early warning and personalized treatment of IPN. The research identified that first-order features of fat (such as skewness, uniformity, etc.), first-order features of muscle, and shape features (such as interquartile range, sphericity, etc.) are all significantly correlated with IPN. In-depth analysis revealed that less localized accumulation of VAT and SAT, uniform density of VAT and SAT, larger volume of SAT, and the dispersion and heterogeneity of abdominal muscle density distribution are all risk factors for IPN. This not only confirms the pivotal role of body composition in the progression of ANP but also provides a scientific basis for the implementation of early preventive treatment in clinical practice for high-risk patients. The findings of this study offer important references and guidance for improving the overall prognosis of pancreatitis patients and optimizing clinical management strategies.

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 Shengjing Hospital of China Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

BH: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization. YG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, 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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

Akturk, Y., Ozbal Gunes, S., and Hekimoglu, B. (2021). The effects of the fat distribution of body, skeletal muscle mass and muscle quality on acute pancreatitis severity: a retrospective cross-sectional study. *J. Comput. Assist. Tomogr.* 45, 500–506. doi: 10.1097/RCT.0000000000001175

Balthazar, E. J., Robinson, D. L., Megibow, A. J., and Ranson, J. H. (1990). Acute pancreatitis: value of CT in establishing prognosis. *Radiology* 174, 331–336. doi: 10.1148/radiology.174.2.2296641

Cao, K., Xia, Y., Yao, J., Han, X., Lambert, L., Zhang, T., et al. (2023). Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat. Med.* 29, 3033–3043. doi: 10.1038/s41591-023-02640-w

Chen, Z., Wang, Y., Zhang, H., Yin, H., Hu, C., Huang, Z., et al. (2023). Deep learning models for severity prediction of acute pancreatitis in the early phase from abdominal nonenhanced computed tomography images. *Pancreas* 52, e45–e53. doi: 10.1097/MPA.0000000000002216

Chen, S. M., Xiong, G. S., and Wu, S. M. (2012). Is obesity an indicator of complications and mortality in acute pancreatitis? An updated meta-analysis. *J. Dig. Dis.* 13, 244–251. doi: 10.1111/j.1751-2980.2012.00587.x

Cruz-Monserrate, Z., Conwell, D. L., and Krishna, S. G. (2016). The impact of obesity on gallstone disease, acute pancreatitis, and pancreatic Cancer. *Gastroenterol. Clin. N. Am.* 45, 625–637. doi: 10.1016/j.gtc.2016.07.010

Dawra, S., Gupta, P., Yadav, N., Singh, A. K., Samanta, J., Sinha, S. K., et al. (2023). Association between the distribution of adipose tissue and outcomes in acute pancreatitis: a comparison of methods of fat estimation. *Indian J. Radiol. Imaging* 33, 012–018. doi: 10.1055/s-0042-1758201

Frank, A. P., de Souza Santos, R., Palmer, B. F., and Clegg, D. J. (2019). Determinants of body fat distribution in humans may provide insight about obesity-related health risks. *J. Lipid Res.* 60, 1710–1719. doi: 10.1194/jlr.R086975

Fu, H., Li, P., Xing, Q., Jiang, H., and Sui, H. (2023). Cutoff value of psoas muscle area as reduced muscle mass and its association with acute pancreatitis in China. *Int. J. Gen. Med.* 16, 2733–2751. doi: 10.2147/IJGM.S413308

Graffy, P. M., Liu, J., Pickhardt, P. J., Burns, J. E., Yao, J., and Summers, R. M. (2019). Deep learning-based muscle segmentation and quantification at abdominal CT: application to a longitudinal adult screening cohort for sarcopenia assessment. *Br. J. Radiol.* 92:20190327. doi: 10.1259/bjr.20190327

Horibe, M., Takahashi, N., Weston, A. D., Philbrick, K., Yamamoto, S., Takahashi, H., et al. (2022). Association between computerized tomography (CT) study of body composition and severity of acute pancreatitis: use of a novel Z-score supports obesity paradox. *Clin. Nutr.* 41, 1676–1679. doi: 10.1016/j.clnu.2022.06.010

Hou, B., Mathai, T. S., Liu, J., Parnell, C., and Summers, R. M. (2024). Enhanced muscle and fat segmentation for CT-based body composition analysis: a comparative study. *Int. J. Comput. Assist. Radiol. Surg.* 19, 1589–1596. doi: 10.1007/s11548-024-03167-2

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18, 203–211. doi: 10.1038/s41592-020-01008-z

Janisch, M., Adelsmayr, G., Müller, H., Holzinger, A., Janek, E., Talakic, E., et al. (2022). Non-contrast-enhanced CT texture analysis of primary and metastatic pancreatic ductal adenocarcinomas: value in assessment of histopathological grade and differences between primary and metastatic lesions. *Abdom. Radiol.* 47, 4151–4159. doi: 10.1007/s00261-022-03646-7

Koç, U., and Taydaş, O. (2020). Evaluation of pancreatic steatosis prevalence and anthropometric measurements using non-contrast computed tomography. *Turk. J. Gastroenterol.* 31, 640–648. doi: 10.5152/tjg.2020.19434

Li, W., Ou, L., Fu, Y., Chen, Y., Yin, Q., and Song, H. (2022). Risk factors for concomitant infectious pancreatic necrosis in patients with severe acute pancreatitis: a systematic review and meta-analysis. *Clin. Res. Hepatol. Gastroenterol.* 46:101901. doi: 10.1016/j.clinre.2022.101901

Lin, Q., Ji, Y.-F., Chen, Y., Sun, H., Yang, D.-D., Chen, A.-L., et al. (2020). Radiomics model of contrast-enhanced MRI for early prediction of acute pancreatitis severity. *J. Magn. Reson. Imaging* 51, 397–406. doi: 10.1002/jmri.26798

Lizcano, F., and Guzmán, G. (2014). Estrogen deficiency and the origin of obesity during menopause. *Biomed. Res. Int.* 2014:757461, 1–11. doi: 10.1155/2014/757461

Lu, J., Ding, Y., Qu, Y., Mei, W., Guo, Y., Fang, Z., et al. (2022). Risk factors and outcomes of multidrug-resistant Bacteria infection in infected pancreatic necrosis patients. *Infect. Drug Resist.* 15, 7095–7106. doi: 10.2147/IDR.S387384

Martínez, J., Johnson, C. D., Sánchez-Payá, J., de Madaria, E., Robles-Díaz, G., and Pérez-Mateo, M. (2006). Obesity is a definitive risk factor of severity and mortality in acute pancreatitis: an updated meta-analysis. *Pancreatology* 6, 206–209. doi: 10.1159/000092104

Modesto, A. E., Ko, J., Stuart, C. E., Bharmal, S. H., Cho, J., and Petrov, M. S. (2020). Reduced skeletal muscle volume and increased skeletal muscle fat deposition characterize diabetes in individuals after pancreatitis: a magnetic resonance imaging study. *Diseases* 8:25. doi: 10.3390/diseases8030025

Palmer, A. K., and Kirkland, J. L. (2016). Aging and adipose tissue: potential interventions for diabetes and regenerative medicine. *Exp. Gerontol.* 86, 97–105. doi: 10.1016/j.exger.2016.02.013

Pascot, A., Lemieux, S., Lemieux, I., Prud'homme, D., Tremblay, A., Bouchard, C., et al. (1999). Age-related increase in visceral adipose tissue and body fat and the metabolic risk profile of premenopausal women. *Diabetes Care* 22, 1471–1478. doi: 10.2337/diacare.22.9.1471

Picca, A., and Calvani, R. (2021). Molecular mechanism and pathogenesis of sarcopenia: an overview. *Int. J. Mol. Sci.* 22:3032. doi: 10.3390/ijms22063032

Ponce-de-Leon, M., Hannemann, A., Linseisen, J., Nauck, M., Lerch, M. M., Bülow, R., et al. (2022). Links between ectopic and abdominal fat and systemic inflammation: new insights from the SHIP-trend study. *Dig. Liver Dis.* 54, 1030–1037. doi: 10.1016/j.dld.2022.02.003

Shen, H., He, P., Ren, Y., Huang, Z., Li, S., Wang, G., et al. (2023). A deep learning model based on the attention mechanism for automatic segmentation of abdominal muscle and fat for body composition assessment. *Quant. Imaging Med. Surg.* 13, 1384–1398. doi: 10.21037/qims-22-330

Tchernof, A., and Després, J.-P. (2013). Pathophysiology of human visceral obesity: an update. *Physiol. Rev.* 93, 359–404. doi: 10.1152/physrev.00033.2011

Tu, H., and Li, Y.-L. (2023). Inflammation balance in skeletal muscle damage and repair. *Front. Immunol.* 14:1133355. doi: 10.3389/fimmu.2023.1133355

Vogele, D., Otto, S., Sollmann, N., Haggenmüller, B., Wolf, D., Beer, M., et al. (2023). Sarcopenia - definition, radiological diagnosis, clinical significance. *Rofo* 195, 393–405. doi: 10.1055/a-1990-0201

Wang, S., Li, S., Feng, Q., Feng, X., Xu, L., and Zhao, Q. (2011). Overweight is an additional prognostic factor in acute pancreatitis: a meta-analysis. *Pancreatology* 11, 92–98. doi: 10.1159/000327688

Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., et al. (2009). Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* 373, 1083–1096. doi: 10.1016/S0140-6736(09)60318-4

Xue, M., Lin, S., Xie, D., Wang, H., Gao, Q., Zou, L., et al. (2023). The value of CT-based radiomics in predicting the prognosis of acute pancreatitis. *Front. Med.* 10:1289295. doi: 10.3389/fmed.2023.1289295

Yee, E. J., Maatman, T. K., Sood, A. J., Gunderman, L., and Zyromski, N. J. (2021). Dynamic frailty: objective physiological assessment to guide management in necrotizing pancreatitis. *Surgery* 169, 1078–1085. doi: 10.1016/j.surg.2020.10.017

Yin, M., Lin, J., Wang, Y., Liu, Y., Zhang, R., Duan, W., et al. (2024). Development and validation of a multimodal model in predicting severe acute pancreatitis based on radiomics and deep learning. *Int. J. Med. Inform.* 184:105341. doi: 10.1016/j.ijmedinf.2024.105341

Yoo, J., Cho, H., Lee, D. H., Cho, E. J., Joo, I., and Jeon, S. K. (2023). Prognostic role of computed tomography analysis using deep learning algorithm in patients with chronic hepatitis B viral infection. *Clin. Mol. Hepatol.* 29, 1029–1042. doi: 10.3350/cmh.2023.0190

Zhang, R., He, A., Xia, W., Su, Y., Jian, J., Liu, Y., et al. (2023). Deep learning-based fully automated segmentation of regional muscle volume and spatial intermuscular fat using CT. *Acad. Radiol.* 30, 2280–2289. doi: 10.1016/j.acra.2023.06.009

Zhang, X., Mao, W., Ke, L., Wu, T., Xu, M., Tan, Y., et al. (2024). Visceral adipose tissue area predicts major adverse kidney events in patients with acute necrotizing pancreatitis. *Am J Med Sci* S0002-9629, 1394–1396. doi: 10.1016/j.amjms.2024.08.003

Zhang, C., Peng, J., Wang, L., Wang, Y., Chen, W., Sun, M.-W., et al. (2024). A deep learning-powered diagnostic model for acute pancreatitis. *BMC Med. Imaging* 24:154. doi: 10.1186/s12880-024-01339-9

Zhang, M., Yin, X., Li, W., Zha, Y., Zeng, X., Zhang, X., et al. (2023). A radiomics based approach using adrenal gland and periadrenal fat CT images to allocate COVID-19 health care resources fairly. *BMC Med. Imaging* 23:181. doi: 10.1186/s12880-023-01145-9

Zhou, Y.-H., Guo, Y., Wang, F., Zhou, C.-L., Tang, C.-Y., Tang, H.-N., et al. (2022). Association of sex Hormones and fat Distribution in men with different obese and metabolic statuses. *Int. J. Gen. Med.* 15, 1225–1238. doi: 10.2147/IJGM.S351282

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 17 December 2024
DOI 10.3389/fmicb.2024.1512091

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Bo Jin,
Dalian University of Technology, China
Ying Fan,
Fudan University, China

*CORRESPONDENCE
Yalian Yu
✉ yyl_hope@sina.cn
Guangjun Fan
✉ 17709875899@163.com

†These authors have contributed equally to this work

RECEIVED 16 October 2024
ACCEPTED 03 December 2024
PUBLISHED 17 December 2024

CITATION
Xiong Y, Liu G, Tang X, Xia B, Yu Y and Fan G (2024) Prediction and analysis of toxic and side effects of tigecycline based on deep learning.
Front. Microbiol. 15:1512091.
doi: 10.3389/fmicb.2024.1512091

Prediction and analysis of toxic and side effects of tigecycline based on deep learning

Yin Xiong[2†], Guoxin Liu[1†], Xin Tang[3], Boyang Xia[4], Yalian Yu[5]* and Guangjun Fan[1]*

[1]Department of Pharmacy, The Second Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, China, [2]Department of Intervention, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [3]Department of Orthopedics, The Second Hospital of Dalian Medical University, Dalian, Liaoning, China, [4]School of Government and Public Affairs, Communication University of China, Beijing, China, [5]Department of Otorhinolaryngology, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning, China

Background: In recent years, with the increase of antibiotic resistance, tigecycline has attracted much attention as a new broad-spectrum glycylcycline antibiotic. It is widely used in the treatment of complex skin and soft tissue infections, complex abdominal infections and hospital-acquired pneumonia by inhibiting bacterial protein synthesis. Tigecycline can exhibit significant time-dependent bactericidal activity, and its efficacy is closely related to pharmacokinetics. It can be evaluated by the ratio of AUC0-24 to the minimum inhibitory concentration (MIC) of pathogens. However, tigecycline may cause nausea, vomiting, diarrhea and a few patients have elevated serum aminotransferase, especially in critically ill patients. The safety of patients still needs further study.

Methods: In this study, the clinical data of 263 patients with pulmonary infection in Shengjing Hospital of China Medical University and the Second Affiliated Hospital of Dalian Medical University were collected retrospectively, and the hepatotoxicity prediction model was established. The potential correlation between the toxic and side effects of tigecycline and the number of hospitalization days was preliminarily discussed, and the correlation analysis between the number of hospitalization days and continuous variables was established. Finally, the deep learning model was used to predict the hospitalization days of patients through simulated blood drug concentration and clinical laboratory indicators.

Results: The degree of abnormal liver function was significantly correlated with AST, GGT, MCHC and hospitalization days. Secondly, the correlation between hospitalization time and clinical test indexes and simulated drug concentration was analyzed. It was found that multiple clinical laboratory parameters of patients (such as EO #, HCT, HGB, MCHC, PCT, PLT, WBC, AST, ALT, Urea), first dose (Dose), age and APACHE II score were significantly correlated with hospitalization days. The simulated blood drug concentration was correlated with the length of hospital stay from 12 h after administration, and reached the strongest between 24 and 48 h. The AUC of the liver function prediction model can reach 0.90. Further analysis showed that there was a potential correlation between hepatotoxicity and hospitalization days. The median hospitalization days of patients in the non-hepatotoxicity group, liver function injury group and hepatotoxicity group were 20, 23, and 30 days, respectively. Based on these results, the length of hospital stay was predicted by the deep learning prediction model with an error within 1 day.

Conclusion: In this study, the hospitalization days of infected patients were predicted by deep learning model with low error. It was found that it was related to clinical test parameters, hepatotoxicity and dosage after administration. The results provided an important reference for the clinical application of tigecycline, and emphasized the need to pay attention to its toxic and side effects in use.

KEYWORDS

deep learning, tigecycline, pharmacokins, hospital days, hepatotoxicity

1 Introduction

In recent years, with the increasing resistance of antibiotics, the treatment of infection has become more complicated, which further highlights the necessity of new antibiotics in clinical application (Gupta et al., 2017).

Tigecycline, as a new broad-spectrum glycylcycline antibiotic, plays a role by inhibiting bacterial protein synthesis. It is widely used in the treatment of complex skin and soft tissue infections, complex abdominal infections, and hospital-acquired pneumonia and other infections (Bradford et al., 2005; Xie et al., 2017). First, tigecycline has obvious time-dependent bactericidal activity, and the efficacy is closely related to the relationship between pharmacokinetics (PK) and pharmacodynamics (PD). The ratio of AUC0-24 to the minimum inhibitory concentration (MIC) of the pathogen can better predict the therapeutic effect of the drug (Van Wart et al., 2006; Koomanachai et al., 2009; Bhavnani et al., 2012). Tigecycline is mainly excreted through bile, and its excretion in the kidney is low, only about 20% of the prototype drug, which provides more options for the use of patients with renal insufficiency (Ap, 2008; Yamashita et al., 2014).

The most common adverse reactions of tigecycline in clinical application are nausea, vomiting and diarrhea, but in phase 2 and phase 3 clinical trials, it was found that about 2–5% of patients had elevated serum aminotransferase (Babinchak et al., 2005; Ellis-Grosse et al., 2005; Sacchidanand et al., 2005). Elevated serum aminotransferase often suggests abnormal changes in liver function. However, so far (Geng et al., 2018), in the field of related research, there are few studies on the abnormal liver function caused by the use of tigecycline in critically ill patients. This may hinder the comprehensive understanding of the safety characteristics of tigecycline, and the use of tigecycline may cause serious damage to the liver function of critically ill patients, thus affecting the therapeutic effect of patients.

Population pharmacokinetic model (PPK) is a mathematical model that can describe the typical pharmacokinetic characteristics and variability of the population by integrating the plasma concentration and individual information of multiple individuals and considering the variability between individuals and within individuals. The model can effectively capture the influence of covariates such as patient's age, weight, and disease status on pharmacokinetic parameters. Through the combination of PPK model and Bayesian method, compared with the traditional analysis method, the advantage of Bayesian theorem is that it can make full use of prior information and improve the accuracy of estimation. Through the dynamic feedback mechanism, the continuous optimization of model parameters is realized, which is effectively applied to complex and changeable situations, so as to accurately simulate the blood concentration of individual patients.

In recent years, artificial intelligence technology (Iezzi et al., 2019) has gradually shown broad application prospects in pharmacokinetic studies. Based on Deep Learning (LeCun et al., 2015), it can not only process large-scale biomedical data, but also identify complex nonlinear relationships. This ability makes AI a powerful tool in pharmacokinetic studies, especially in the fields of drug concentration prediction, drug interaction analysis, and risk assessment of adverse reactions, thereby providing support for personalized medication and clinical decision-making. A 2023 study explored the significant development of therapeutic drug monitoring (TDM) and model-guided precision drug delivery (MIPD) driven by advances in computing and mathematical technology (Poweleit et al., 2023). A 2022 study that combines a physiologically based pharmacokinetic (PBPK) model with machine learning (ML) or artificial intelligence (AI) techniques to predict ADME parameters using ML/AI, and integrates these prediction models into the PBPK model to predict pharmacokinetic (PK) statistical results (Chou and Lin, 2023). In another study, neural-ODE is applied to PK modeling for the first time. The final results show that it has a wide range of applicability and may have an important impact on future research (Lu et al., 2021). All of these indicate the potential application of artificial intelligence in pharmacokinetic analysis.

Therefore, this study aims to explore the potential risk factors for hepatotoxicity in patients treated with tigecycline by means of artificial intelligence-based technology. By analyzing the correlation between abnormal liver function and laboratory parameters, hospitalization days, etc., the mechanism of hepatotoxicity of tigecycline and its potential relationship with prolonged hospitalization days were revealed.

In addition, by exploring the effect of tigecycline hepatotoxicity on hospitalization days, a prediction model of hospitalization days was established to provide scientific reference for clinical practice, and then provide an important reference for optimizing the clinical application of tigecycline.

2 Method

2.1 Study population

This study retrospectively collected the clinical data of two patients with cardiopulmonary infection. The study was approved by the Ethics Committee, Ethics No. (2019 no. 049). Inclusion criteria: Patients included in this study should meet the following criteria: (1) patients with clinical intravenous use of tigecycline for more than three days; (2) Tigecycline is for therapeutic use; (3) Pulmonary infection caused by Gram-positive or Gram-negative bacteria, such as pneumonia or bronchitis, is diagnosed or highly suspected by clinicians. Exclusion criteria: (1) patients with cirrhosis or liver failure; (2) Patients died within 24 h after the use of tigecycline; (3) Pregnancy; and (4) Other medications that may affect the liver during tigecycline treatment. Inclusion and exclusion as shown in Figure 1.

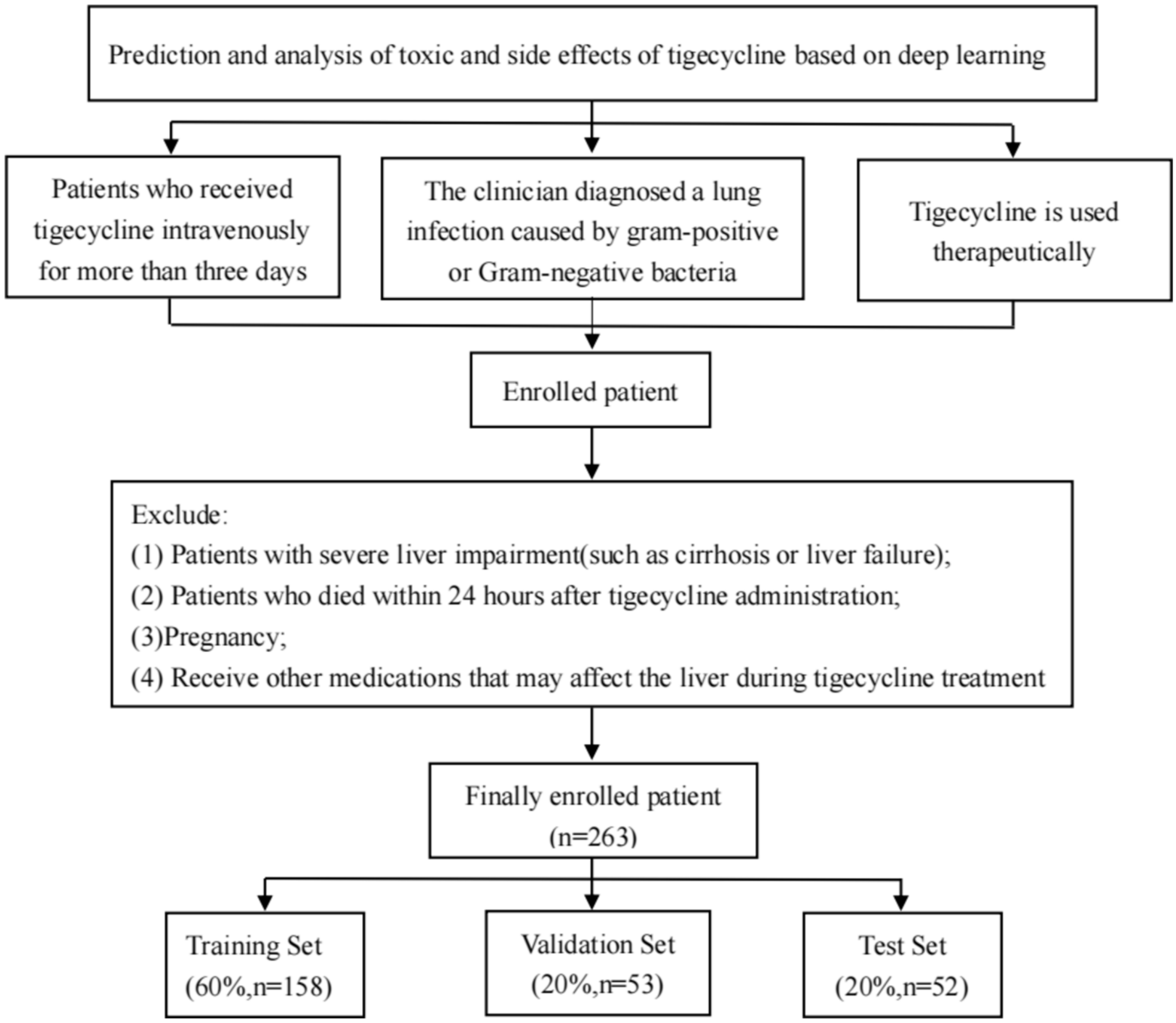


FIGURE 1
Inclusion and exclusion process of patients in this study. Finally, 263 patients using tigecycline in two centers were retrospectively collected through the inclusion and exclusion criteria, and the data were trained in a deep learning model according to the corresponding proportion.

2.2 Data collection

All enrolled patients were treated with tigecycline, and laboratory indicators and basic information were recorded in detail. This study collected patient data through the hospital's electronic medical record system and nursing system. The collected laboratory indicators included: (1) patient age, initial dose of tigecycline (Dose), and length of hospital stay; (2) laboratory examination results of patients during medication. Such as EO # (absolute number of eosinophils), EO % (percentage of eosinophils), HCT (hematocrit), HGB (hemoglobin), MCH (mean corpuscular hemoglobin content), MCHC (mean corpuscular hemoglobin concentration), MCV (mean corpuscular volume), MPV (mean platelet volume), PCT (platelet hematocrit), PLT (platelet count), WBC (white blood cell count), RH (hemorheology), A/G ratio (albumin/globulin ratio), ALT (alanine aminotransferase), AST (aspartate aminotransferase), GGT (γ-glutamyl transferase), TP (total protein), Urea (urea), and APACHE II score. (3) Drug-induced adverse reactions, such as hepatotoxicity.

The abnormal liver function of tigecycline was defined as the ALT value measured twice in a row was between the upper limit of the normal value (5–40 U/L) and the upper limit of the normal value by 3 times, and its hepatotoxicity was defined as the ALT value measured twice in a row >3 times the upper limit of the normal value (or blood bilirubin >1.5 times the upper limit of the normal value), or greater than 1.5 times the baseline value (if the baseline value is abnormal) (Fan et al., 2020).

2.3 Blood concentration simulation

In this study, we constructed a population pharmacokinetic model based on previous research results (Luo et al., 2023) (Equations 1, 2). Bayesian feedback method (Aggelopoulos, 2015) was used to simulate the blood concentration of patients at different time points after administration. We determined the mathematical form of the model, including key kinetic processes such as drug absorption, distribution, metabolism, and excretion, taking into account individual differences, residual variation, and drug characteristics. The parameter values of the model were based on previous research reports, and the Bayesian feedback method was used to simulate the corresponding blood concentration values at different time points. The final model formula:

$$CL\left(L/h\right)=\left(11.30-0.14*APACHE\ II\right)*e^{0.065} \tag{1}$$

$$V\left(L\right)=\left[105.00*\left(1-0.0059*AGE\right)\right]*e^{0.160} \tag{2}$$

2.4 Deep learning model

In order to solve the complex problem of hepatotoxicity prediction during treatment, we used the latest KAN network (Liu et al., 2024) to establish a tigecycline hepatotoxicity prediction model. KAN is derived from Kolmogorov-Arnold representation theorem, whose core idea is that any continuous function can be represented by a combination of one-dimensional functions. We map the blood drug concentration and the patient's clinical laboratory indicators to the input layer of the KAN model. Through its specific multi-layer structure, KAN nests these multi-dimensional input features into a one-dimensional function combination, and then approximates the complex nonlinear relationship between drugs and liver function indicators. Based on these input characteristics, we established a predictive model that can not only predict the number of days of hospitalization, but also further predict the possible liver toxicity of patients during treatment, and provide support for personalized treatment and clinical decision-making. The research flow chart of this study is shown in Figure 2.

2.5 Statistical analysis

Statistical analysis was performed using IBM SPSS 26.0 software. First, we test the normality of all data using the Kolmogorov–Smirnov test. For data that conform to the normal distribution, we use the mean ± standard deviation to describe; for data that do not conform to the normal distribution, the median (quartile) is used to describe. For categorical data, the t test was used for data that conformed to the normal distribution, and the Mann–Whitney U test was used for data that did not conform to the normal distribution. For continuous variables, Pearson correlation analysis was used for data that met the normal distribution, and Spearman correlation analysis was used for data that did not meet the normal distribution. Differences were considered statistically significant at $p < 0.05$. The statistics chart will be drawn using the Python-based matplotlib library.

3 Results

3.1 Baseline

A total of 263 patients from two centers were included in this study: 200 from Institution I (Shengjing Hospital Affiliated to China Medical University) and 63 from Institution II (Second Hospital Affiliated to Dalian Medical University). All patients were treated with tigecycline. In terms of data, the training, validation and test sets are divided according to the proportion of 6: 2: 2 patients, which ensures the rationality of model training, tuning and evaluation. This division method helps to improve the generalization ability of the model and effectively avoid over-fitting. Table 1 lists the baseline information of the included patients.

3.2 Abnormal liver function, toxicity and laboratory relevance

Different laboratory indicators are helpful to evaluate the degree of liver injury. The overall ALT group, abnormal liver function group and hepatotoxicity group were established to explore the relationship between these three groups and laboratory indicators. Based on this grouping, the effects of different types and degrees of liver injury on various test indicators can be accurately analyzed. As shown in Table 2, it was found that there was a significant correlation between the number of hospital stays and the ALT group and the abnormal liver function group, suggesting that liver function damage may affect the length of hospital stay. At the same time, MCHC, AST and GGT showed significant correlation with the three groups, indicating that poor liver function can indirectly affect the formation and function of red blood cells, thus affecting the MCHC value. In addition, liver cells are damaged, and AST is released from the liver cells into the blood, resulting in an increase in AST, and if the biliary system is damaged, GGT levels will increase accordingly.

3.3 Prediction of hepatotoxicity based on deep learning

Due to the small sample size of patients with hepatotoxicity in this study, there are some challenges in directly predicting hepatotoxicity. Therefore, we indirectly reflect the potential risk of tigecycline-induced hepatotoxicity by predicting whether the patient has abnormal liver function. The training results are shown in Figure 3. The model showed high performance in the testing set, with a sensitivity of 0.94, a specificity of 0.87, and an AUC value of 0.9, as shown in Figure 4. This shows that by predicting the abnormal liver function of hospitalized patients, we can indirectly capture the potential risk of hepatotoxicity and provide an early warning model for early clinical identification and intervention of possible liver injury.

3.4 Potential association between liver function and length of hospital stay

Although we cannot directly analyze the relationship between drug-induced hepatotoxicity and length of hospital stay, we can preliminarily explore whether there is a potential correlation between the degree of abnormal liver function after drug treatment and the length of hospital stay. Therefore, we divided the patients into three groups according to the definition of hepatotoxicity as shown in Table 3: hepatotoxic group, liver function injury group (defined as two consecutive ALT values between 40 U/L and 120 U/L), and non-hepatotoxic group. Through Figure 5, the box diagram visually shows the distribution characteristics and abnormal values of hospitalization days in different patient groups. The results showed that the length of hospital stay in the non-hepatotoxic

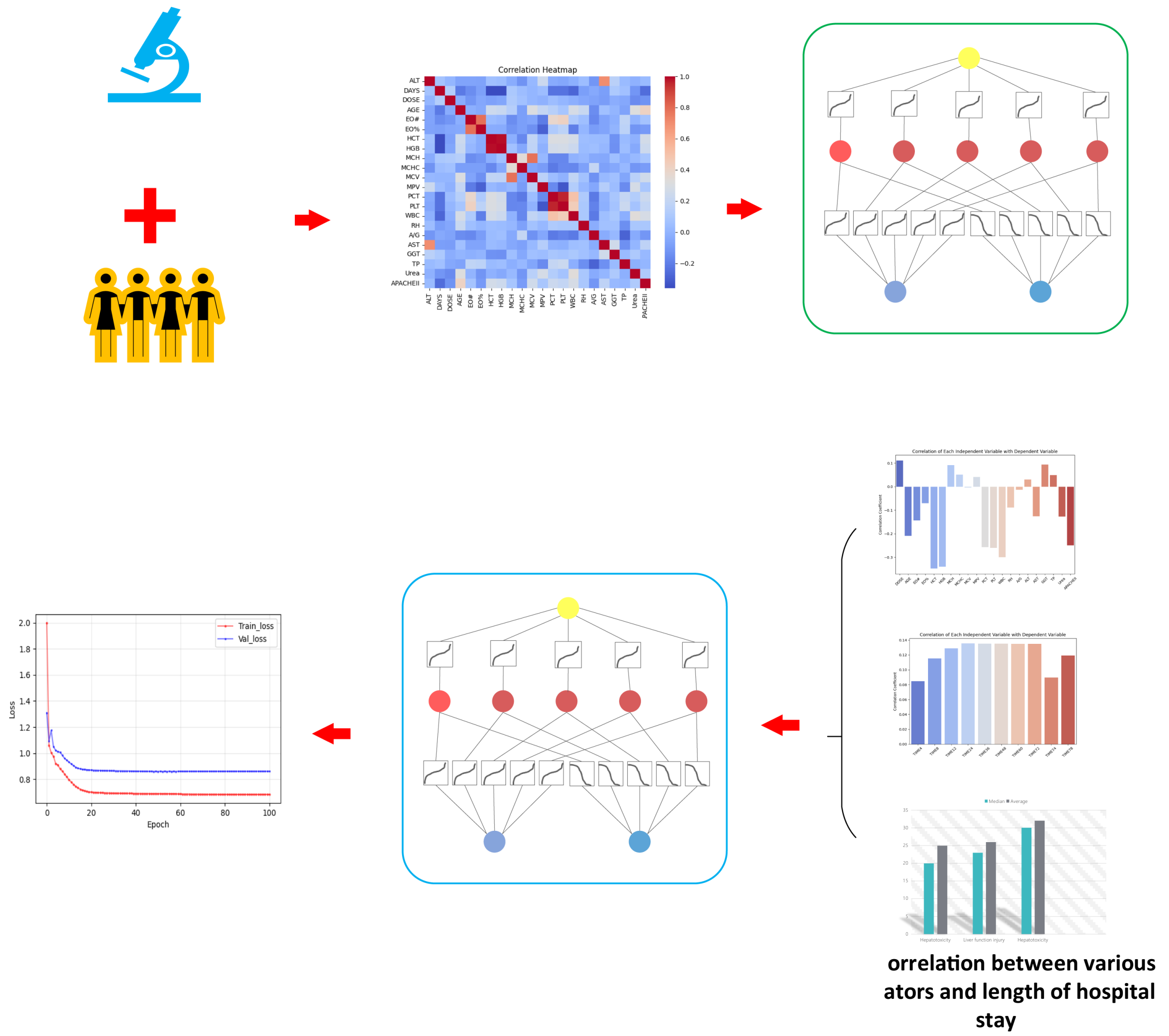


FIGURE 2
Overall flow chart of this study. **(A)** This study retrospectively collected the clinical laboratory indicators of the two centers, and the blood concentration simulated by Bayesian feedback method, established the correlation between liver function and clinical test, and based on this, a hepatotoxicity prediction model was constructed. **(B)** Establish the correlation between hospitalization days and clinical experimental indicators and simulated blood concentration, and explore and confirm the potential relationship between hospitalization days and hepatotoxicity. By inputting test indicators and blood concentration to the KAN network, the hospitalization days were predicted and the hepatotoxicity was indirectly predicted, and finally the prediction results with lower errors were obtained.

TABLE 1 **Patient baseline information.**

Hospital name	Total number of patients	Number of participants in the training set (60%)	Number of people in the validation set (20%)	Number of people in the test set (20%)	Average age (year)	Age range (year)
Institution I	200	120	40	40	66	14–98
Institution II	63	38	13	12	58	18–89
Total	263	158	53	52	62	14–98

group was shorter, with a median of 20 days, but the distribution of length of hospital stay was wider, suggesting that there may be extreme values of individual length of hospital stay in this group. For the liver function injury group, the distribution of hospitalization days was relatively concentrated and the variability was small. The median (23 days) was located above the box, indicating that the hospitalization days of this group of patients were generally longer. Finally, the box of the hepatotoxicity group was wider and the overall was higher, indicating that the median length of hospital stay in this group was higher (30 days), and the overall length of hospital stay was the longest among the three groups. In addition, there was no significant extreme value or outlier in the length of hospital stay in this group. The overall

TABLE 2 Abnormal liver function, toxicity and laboratory relevance.

Characteristic	DAYS	AGE	EO#	EO%	HCT	HGB	MCH	MCHC	MCV	MPV	PCT	PLT	WBC	RH	A/G	AST	GGT	TP	Urea	APACHE II
ALT	0.027*	0.950	0.687	0.702	0.265	0.254	0.650	0.027*	0.435	0.210	0.836	0.719	0.826	0.035*	0.169	0.001*	0.001*	0.389	0.643	0.690
Abnormal liver function	0.043*	0.661	0.037*	0.111	0.874	0.918	0.889	0.016*	0.216	0.610	0.740	0.975	0.950	0.437	0.041*	0.001*	0.001*	0.724	0.164	0.843
Hepatotoxicity	0.073	0.813	0.649	0.075	0.417	0.557	0.054	0.005*	0.601	0.001*	0.074	0.851	0.243	0.768	0.118	0.001*	0.048*	0.777	0.106	0.212

*The difference was statistically significant ($p < 0.05$).

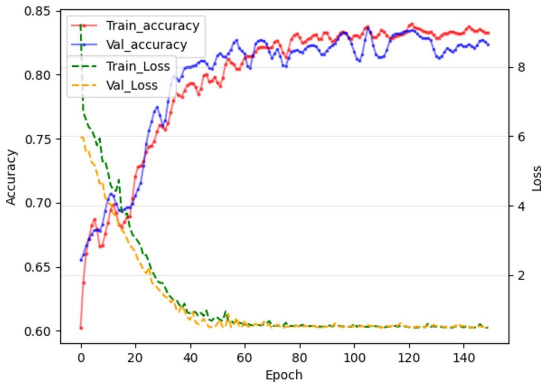


FIGURE 3
In the 121st round, the training and validation phases achieved the best results, and the accuracy of the liver function prediction model reached 84.71 and 83.76, respectively.

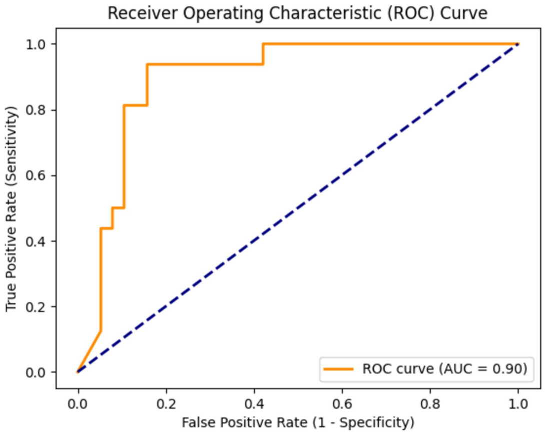


FIGURE 4
The area under the curve (AUC) was 0.90, indicating that the model had a high ability to distinguish whether there was abnormal liver function.

results reflect the significant differences in the number of hospitalization days among different patient groups, suggesting that hepatotoxicity may have a greater impact on the length of hospitalization, while the liver function injury group showed a more uniform hospitalization demand.

3.5 Correlation between hospitalization days and clinical laboratory indexes

Table 4 shows the correlation between the length of hospital stay and clinical information of the included patients. Among the laboratory parameters, EO#, HCT, HGB, MCHC, PCT, PLT, WBC, AST, ALT, urea $p < 0.05$, that is, these indicators may be potential factors reflecting hepatotoxicity. The p value of dose was <0.05, indicating that higher doses may lead to greater toxic and side effects, which in turn leads to longer hospital stay. Similarly, age and APACHEII p values <0.05 indicated that tigecycline had a higher risk of toxicity in elderly or more severe patients.

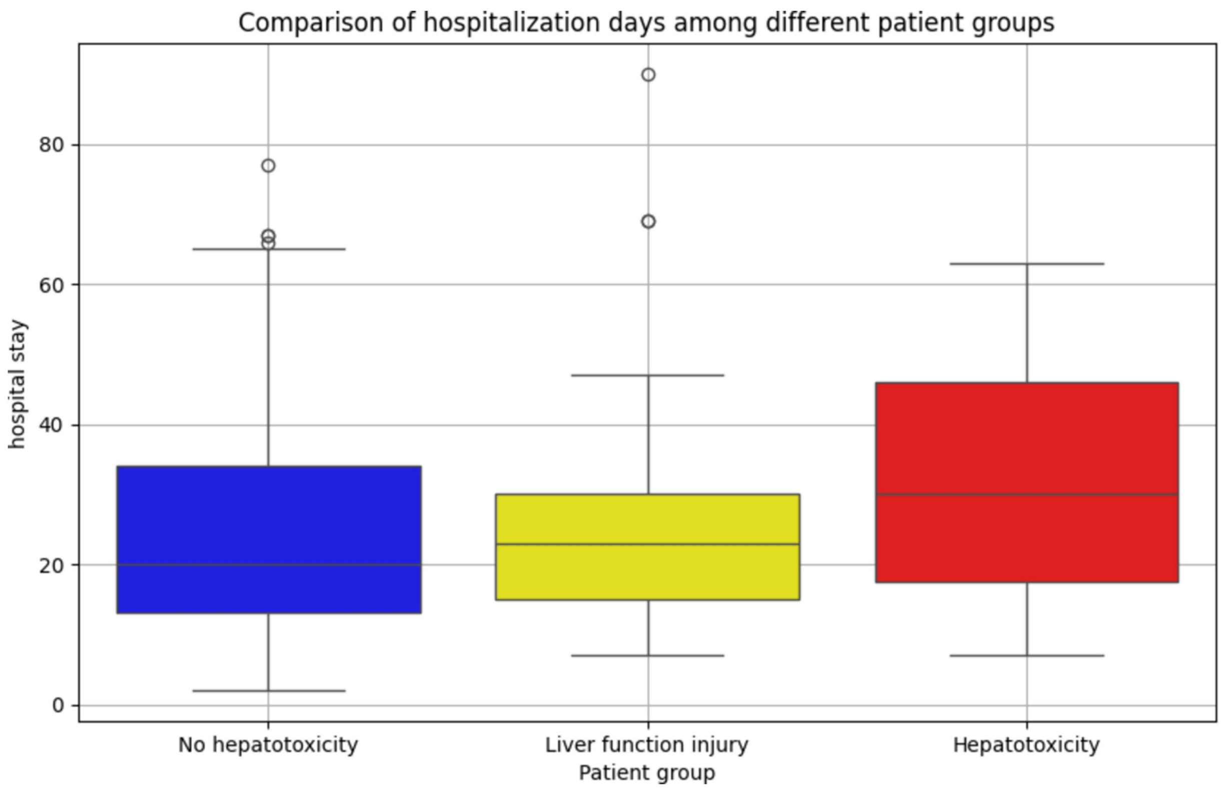


FIGURE 5
Distribution and difference of hospitalization days in patients with different liver function status. The number of hospitalization days in the non-hepatotoxicity group was the most widely distributed, and there were more extreme hospitalization days. The distribution of hospitalization days in the liver function injury group was more concentrated, and the hospitalization time was longer as a whole; the distribution of hospitalization days in the hepatotoxic group was more balanced, and no abnormal value was found. It reflects the difference in hospitalization days among different patient groups.

TABLE 3 Descriptive statistics of the number of hospitalization days in the patient group.

Patient group	Number of people	Average length of hospital stay	Median length of hospital stay	Standard deviation
No hepatotoxicity	202	25	20	16.37
Liver function injury	50	26	23	16.59
Hepatotoxicity	11	32	30	19.97

The difference was statistically significant ($p < 0.05$).

3.6 The correlation between hospitalization days and blood drug concentration

The plasma concentration (ng/ml) of patients at 10 time points (4, 8, 12, 24, 36, 48, 60, 72, 74, and 78 h) after the first administration was simulated by the constructed pharmacokinetic model. The results are shown in Table 5. In the correlation analysis of hospitalization days and blood drug concentration, as shown in Table 6, the correlation began to appear 12 h after administration, and reached the strongest between 24 and 48 h. To a certain extent, this shows that the residual concentration of the drug is the most significant toxic and side effects for patients in the 24–48 h period. Therefore, it is recommended to use this time period as an important node for clinical testing so that clinicians can make corresponding intervention and treatment decisions in a timely manner.

3.7 Prediction results of hospitalization days based on deep learning

The simulated blood drug concentration and laboratory parameters were integrated into the constructed deep learning model. The training results are shown in Figure 6. In the training and verification stages, we observed that the model loss was maintained below 1. This result means that the error between the predicted hospitalization days and the actual hospitalization days can be effectively controlled within one day, showing the high accuracy

TABLE 4 The correlation between continuous variables and hospitalization days.

Indicator	Statistic	*p*-value	Indicator	Statistic	*p*-value
Age	263	0.001*	WBC	263	0.001*
EO#	263	0.004*	Dose	263	0.048*
EO%	263	0.065	RH	263	0.053
HCT	263	0.001*	A/G	263	0.345
HGB	263	0.001*	ALT	263	0.027*
MCH	263	0.092	AST	263	0.001*
MCHC	263	0.008*	GGT	263	0.367
MCV	263	0.161	TP	263	0.151
MPV	263	0.323	Urea	263	0.048*
PCT	263	0.001*	APACHEII	263	0.001*
PLT	263	0.001*			

*The difference was statistically significant ($p < 0.05$).

TABLE 5 Through the constructed pharmacokinetic model, the blood concentration at different time points after the first administration was simulated.

	Training set/158	Validation set/53	Testing set/52
4.0 h	1568.23 (728.41, 3119.10)	1605.29 (780.61, 2266.90)	1480.72 (762.71， 2380.40)
8.0 h	900.72 (432.17, 1881.60)	870.55 (426.39, 1413.80)	886.90 (449.74, 1582.30)
12.0 h	526.26 (216.02, 1425.90)	480.67 (194.17, 881.76)	522.87 (224.83, 1051.80)
24.0 h	371.16 (134.19, 1016.30)	325.21 (115.42, 654.79)	371.71 (140.44, 834.83)
36.0 h	340.65 (124.87, 929.15)	298.69 (107.98, 599.72)	340.84 (363.75, 771.1)
48.0 h	334.30 (123.61, 910.61)	293.87 (107.28, 586.37)	334.27 (128.61, 752.39)
60.0 h	332.91 (123.45, 906.67)	292.95 (107.21, 583.13)	332.81 (128.45, 746.89)
72.0 h	332.58 (123.43, 905.83)	292.76 (107.21, 582.34)	332.48 (128.42, 745.27)
74.0 h	1283.29 (633.02, 2562.00)	1298.10 (641.67, 1895.20)	1245.04 (654.06, 2067.50)
78.0 h	742.19 (349.25, 1578.20)	710.59 (348.88, 1182.00)	731.19 (363.74, 1374.30)

and reliability of the model in predicting hospitalization days (Table 6).

4 Discussion

This study first analyzed the effect of liver function injury on test indicators. The results showed that the length of hospital stay was significantly correlated with ALT group and abnormal liver function group, suggesting that liver function injury may prolong the length of hospital stay. At the same time, MCHC, AST and GGT were significantly correlated with the three groups, reflecting the impact of poor liver function on red blood cell function, liver cell damage and biliary system damage, and had potential clinical evaluation value.

Secondly, the liver function early warning model established in this study has good performance and indirectly captures the potential risks of hepatotoxicity. By exploring the potential relationship between hospitalization days and tigecycline hepatotoxicity, the results showed that the degree of liver function damage was associated with

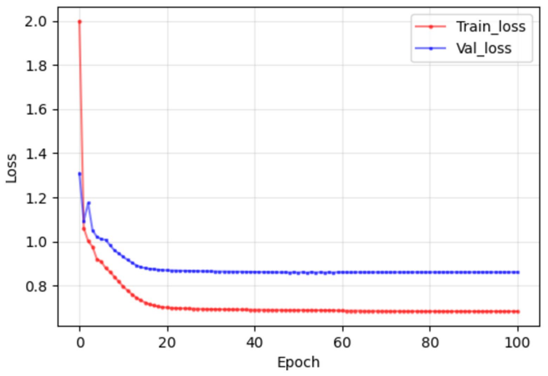


FIGURE 6
In the training and validation stages, the model maintained a good consistency in the prediction of hospitalization days, and the error could eventually be maintained within 1 day, which confirmed that the model established in this study had good prediction performance.

TABLE 6 Correlation between blood drug concentration and hospitalization days.

		4 h	8 h	12 h	24 h	36 h	48 h	60 h	72 h	74 h	78 h
Hospital days	Statistic	263	263	263	263	263	263	263	263	263	263
	P-value	0.193	0.111	0.029*	0.010*	0.010*	0.010*	0.012*	0.011*	0.201	0.110

*The difference was statistically significant ($p < 0.05$).

hospitalization days. The hospitalization time of the patients in the hepatotoxicity group was longer as a whole, and the hospitalization days of the patients in this group showed a large change, which may be related to the severity of hepatotoxicity and the increase of treatment needs.

Then, by analyzing the correlation between the actual hospitalization days and the patient information and clinical indicators, it was found that there was a significant correlation between a number of clinical indicators and the actual hospitalization days. The actual hospitalization days obtained by the analysis were strongly related to the patient's age, WBC, EO #, Dose, HCT, HGB, AST, ALT, MCHC, Urea, PCT, PLT, APACHEII.

Finally, this experiment uses simple and easy-to-obtain hospitalization days to replace the clinical index values that need to be measured multiple times, and uses the constructed deep learning model to predict hospitalization days. The error between the predicted results and the actual days is controlled at about one day. However, this cannot completely replace the clinical judgment of the liver side effects of tigecycline. There is a certain error. As the patient ages, the body function and various indicators will gradually decline. According to the study of Huang et al., the in-hospital mortality and length of stay gradually increased with age (Huang et al., 2023), which was consistent with the results of this experiment. However, if we want to further explore the relationship between the liver side effects of tigecycline and age in this experiment, we need to further collect clinical index data for correlation test.

At the same time, it was also found that the liver side effects of tigecycline were negatively correlated with the WBC level of the patients. As the side effects increased, the WBC and PCT levels of the patients decreased. Li et al. found that cefoperazone combined with tigecycline in the treatment of ICU infection can effectively improve the therapeutic effect of the disease, and significantly enhance the bacterial clearance, while reducing serum WBC and PCT levels (Li and Zhang, 2022). This suggests that when tigecycline is used alone, although there is a certain effect on the infection of patients, the toxic and side effects also increase. Whether there is the possibility of cefoperazone alleviating the toxic and side effects of tigecycline, which also provides a direction for future research and provides evidence for this experiment; the toxic and side effects were negatively correlated with the levels of HCT, HGB, AST, MCHC, PCT and PLT. With the increase of toxic and side effects, the levels of HCT, HGB, AST, MCHC, PCT and PLT decreased, which represented the decrease of liver function and coagulation function.

In a number of case reports and experiments (Sabanis et al., 2015; McMahan and Moenster, 2017; Cui et al., 2019), it is mentioned that the use of tigecycline will affect the coagulation system of patients, cause coagulation disorders, and produce adverse clinical outcomes. Therefore, in this experiment, the use of simple and easy-to-obtain hospitalization days can be a good warning for the occurrence of adverse coagulation events in patients. At the same time, in the study of Zhang et al., it was also found that in patients using tigecycline, renal dysfunction also caused tigecycline-induced coagulation-related adverse events (Zhang et al., 2020). In this experiment, the toxic and side effects of patients were negatively correlated with urea levels. Urea showed the liver function of patients to a certain extent, which also provided more evidence sources for the clinical manifestations of tigecycline toxicity.

We can find that most clinical indicators point to the decline of liver function. In recent studies, the deep learning of pharmacokinetics is combined with clinical imaging to empower the metabolic changes of pharmacokinetics in various organs, and accurately segment and refine the role of organs (Arledge et al., 2022; Ota and Yamashita, 2022; Dhaliwal et al., 2024). Although this study replaced the toxic and side effects of tigecycline and analyzed the correlation with a number of laboratory indicators, the experiment can further carry out more accurate analysis in the metabolic imaging of the patient's liver, match the number of days of hospitalization, and refer to the number of days of hospitalization. More accurate, and in the future, we can try to build a liver model to achieve a simulated drug metabolism process, provide clinicians with more predictive medication recommendations, and avoid excessive or excessive medication.

At the same time, there are still corresponding limitations in this experiment. In view of the lack of sample size of patients with hepatotoxicity in this study, although the liver function prediction model has alleviated the problem of insufficient number of patients with hepatotoxicity to a certain extent, it also limits the specific interpretation of drug hepatotoxicity prediction. At the same time, further research on the correlation between hepatotoxicity and hospitalization days needs to verify this conclusion in a larger patient group and other external centers in the future to enhance the reliability and generalization of the results. As well as the defects that the study did not include the total number of patients' medications and the total amount of medications into the experimental data, the patient stratification was not diversified enough, resulting in the inability to present a more detailed correlation analysis, which will be improved in future research.

5 Conclusion

Tigecycline, as a new type of antibiotic, has shown good clinical application potential in the treatment of complex infections. This study preliminarily explored the correlation between drug liver dysfunction, toxicity and laboratory parameters. By indirectly reflecting the potential possibility of hepatotoxicity, an early warning model of liver injury that can be clinically identified and intervened early was established. The significant correlation between liver function status and hospitalization days, as well as hospitalization days and clinical laboratory parameters and simulated dose after the first administration was determined. Finally, through the prediction and analysis of drug side effects on the simple and easy-to-obtain hospitalization days, it provides an important reference for the clinical application of tigecycline, and suggests that the side effects of drugs should be paid attention to in clinical use, especially in critically ill patients.

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 Shengjing Hospital Affiliated to China Medical University Ethics Committee of the Second Affiliated Hospital of Dalian Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because this is a retrospective study, so there is no need for it.

Author contributions

YX: Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. GL: Data curation, Investigation, Writing – review & editing. XT: Supervision, Writing – review & editing. BX: Supervision, Writing – review & editing. YY: Supervision, Writing – review & editing. GF: Funding acquisition, 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 approved by the Ethics Committee of the Second Affiliated Hospital of Dalian Medical University (2019 no. 049).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Aggelopoulos, N. C. (2015). Perceptual inference. *Neurosci. Biobehav. Rev.* 55, 375–392. doi: 10.1016/j.neubiorev.2015.05.001

Ap, M. (2008). Tigecycline pharmacokinetic/pharmacodynamic update. *J. Antimicrob. Chemother.* 62, i11–i16. doi: 10.1093/jac/dkn242

Arledge, C.A., Sankepalle, D. M., Crowe, W. N., Liu, Y., Wang, L., Zhao, D., et al. (2022). Deep learning quantification of vascular pharmacokinetic parameters in mouse brain tumor models. *Front. Biosci.* 27:99. doi: 10.31083/j.fbl2703099

Babinchak, T., Ellis-Grosse, E., Dartois, N., Rose, G. M., Loh, E., Tigecycline 301 Study Groupet al. (2005). The efficacy and safety of tigecycline for the treatment of complicated intra-abdominal infections: analysis of pooled clinical trial data. *Clin. Infect. Dis.* 41, S354–S367. doi: 10.1086/431676

Bhavnani, S. M., Rubino, C. M., Hammel, J. P., Forrest, A., Dartois, N., Cooper, C. A., et al. (2012). Pharmacological and patient-specific response determinants in patients with hospital-acquired pneumonia treated with tigecycline. *Antimicrob. Agents Chemother.* 56, 1065–1072. doi: 10.1128/AAC.01615-10

Bradford, P. A., Weaver-Sands, D. T., and Petersen, P. J. (2005). In vitro activity of tigecycline against isolates from patients enrolled in phase 3 clinical trials of treatment for complicated skin and skin-structure infections and complicated intra-abdominal infections. *Clin. Infect. Dis.* 41, S315–S332. doi: 10.1086/431673

Chou, W.-C., and Lin, Z. (2023). Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling.

Cui, N., Cai, H., Li, Z., Lu, Y., Wang, G., and Lu, A. (2019). Tigecycline-induced coagulopathy: a literature review. *Int. J. Clin. Pharm.* 41, 1408–1413. doi: 10.1007/s11096-019-00912-5

Dhaliwal, A., Ma, J., Zheng, M., Lyu, Q., Rajora, M. A., Ma, S., et al. (2024). Deep learning for automatic organ and tumor segmentation in nanomedicine pharmacokinetics. *Theranostics* 14, 973–987. doi: 10.7150/thno.90246

Ellis-Grosse, E. J., Babinchak, T., Dartois, N., Rose, G., Loh, E., Tigecycline 300 cSSSI Study Groupet al. (2005). The efficacy and safety of tigecycline in the treatment of skin and skin-structure infections: results of 2 double-blind phase 3 comparison studies with vancomycin-aztreonam. *Clin. Infect. Dis.* 41, S341–S353. doi: 10.1086/431675

Fan, G., Jin, L., Bai, H., Jiang, K., Xie, J., and Dong, Y. (2020). Safety and efficacy of Tigecycline in intensive care unit patients based on therapeutic drug monitoring. *Ther. Drug Monit.* 42, 835–840. doi: 10.1097/FTD.0000000000000784

Geng, T.-T., Xu, X., and Huang, M. (2018). High-dose tigecycline for the treatment of nosocomial carbapenem-resistant *Klebsiella pneumoniae* bloodstream infections: a retrospective cohort study. *Medicine* 97:e9961. doi: 10.1097/MD.0000000000009961

Gupta, R., Malik, A., Rizvi, M., Ahmed, M., and Singh, A. (2017). Epidemiology of multidrug-resistant gram-negative pathogens isolated from ventilator-associated pneumonia in ICU patients. *J. Glob. Antimicrob. Resistan.* 9, 47–50. doi: 10.1016/j.jgar.2016.12.016

Huang, L., Shi, Y., Wang, L., Rong, L., Ren, Y., Xu, C., et al. (2023). Characteristics and in-hospital outcomes of elderly patients with cancer in a top-ranked hospital in China, 2016-2020: real-world study. *Cancer Med.* 12, 2885–2905. doi: 10.1002/cam4.5203

Iezzi, R., Goldberg, S. N., Merlino, B., Posa, A., Valentini, V., and Manfredi, R. (2019). Artificial intelligence in interventional radiology: a literature review and future perspectives. *J. Oncol.* 2019, 6153041–6153045. doi: 10.1155/2019/6153041

Koomanachai, P., Kim, A., and Nicolau, D. P. (2009). Pharmacodynamic evaluation of tigecycline against *Acinetobacter baumannii* in a murine pneumonia model. *J. Antimicrob. Chemother.* 63, 982–987. doi: 10.1093/jac/dkp056

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, N., and Zhang, H. (2022). Effects of Tigecycline combined with Cefoperazone on bacterial clearance and serum biochemical indexes in patients with pulmonary infections in ICU. *Pak. J. Med. Sci.* 38, 1622–1626. doi: 10.12669/pjms.38.6.5872

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljacic, M., et al. (2024). KAN: Kolmogorov-Arnold networks. arXiv [Preprint]. Available at: http://arxiv.org/abs/2404.19756 (Accessed: 15 October 2024).

Lu, J., Deng, K., Zhang, X., Liu, G., and Guan, Y. (2021). Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience* 24:102804. doi: 10.1016/j.isci.2021.102804

Luo, X., Wang, S., Li, D., Wen, J., Sun, N., and Fan, G. (2023). Population pharmacokinetics of tigecycline in critically ill patients. *Front. Pharmacol.* 14:1083464. doi: 10.3389/fphar.2023.1083464

McMahan, J., and Moenster, R. P. (2017). Tigecycline-induced coagulopathy. *Am. J. Health Syst. Pharm.* 74, 130–134. doi: 10.2146/ajhp150894

Ota, R., and Yamashita, F. (2022). Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics. *J. Control. Release* 352, 961–969. doi: 10.1016/j.jconrel.2022.11.014

Poweleit, E. A., Vinks, A. A., and Mizuno, T. (2023). Artificial intelligence and machine learning approaches to facilitate therapeutic drug management and model-informed precision dosing. *Ther. Drug Monit.* 45, 143–150. doi: 10.1097/FTD.0000000000001078

Sabanis, N., Paschou, E., Gavriilaki, E., Kalaitzoglou, A., and Vasileiou, S. (2015). Hypofibrinogenemia induced by tigecycline: a potentially life-threatening coagulation disorder. *Infect. Dis.* 47, 743–746. doi: 10.3109/23744235.2015.1043942

Sacchidanand, S., Penn, R. L., Embil, J. M., Campos, M. E., Curcio, D., Ellis-Grosse, E., et al. (2005). Efficacy and safety of tigecycline monotherapy compared with vancomycin plus aztreonam in patients with complicated skin and skin structure infections: results from a phase 3, randomized, double-blind trial. *Int. J. Infect. Dis.* 9, 251–261. doi: 10.1016/j.ijid.2005.05.003

Van Wart, S. A., Owen, J. S., Ludwig, E. A., Meagher, A. K., Korth-Bradley, J. M., Cirincione, B. B., et al. (2006). Population pharmacokinetics of tigecycline in patients with complicated intra-abdominal or skin and skin structure infections. *Antimicrob. Agents Chemother.* 50, 3701–3707. doi: 10.1128/AAC.01636-05

Xie, J., Roberts, J. A., Alobaid, A. S., Roger, C., Wang, Y., Yang, Q., et al. (2017). Population pharmacokinetics of Tigecycline in critically ill patients with severe infections. *Antimicrob. Agents Chemother.* 61:e00345-17. doi: 10.1128/AAC.00345-17

Yamashita, N., Matschke, K., Gandhi, A., and Korth-Bradley, J. (2014). Tigecycline pharmacokinetics, tolerability, safety, and effect on intestinal microflora in healthy Japanese male subjects. *J. Clin. Pharmacol.* 54, 513–519. doi: 10.1002/jcph.236

Zhang, Q., Wang, J., Liu, H., Ma, W., Zhou, S., and Zhou, J. (2020). Risk factors for tigecycline-induced hypofibrinogenaemia. *J. Clin. Pharm. Ther.* 45, 1434–1441. doi: 10.1111/jcpt.13250

frontiers | Frontiers in Microbiology

TYPE Systematic Review
PUBLISHED 08 January 2025
DOI 10.3389/fmicb.2024.1510026

Revolutionizing diagnosis of pulmonary *Mycobacterium tuberculosis* based on CT: a systematic review of imaging analysis through deep learning

Fei Zhang[1†], Hui Han[2†], Minglin Li[1†], Tian Tian[1], Guilei Zhang[1], Zhenrong Yang[3], Feng Guo[4], Maomao Li[5], Yuting Wang[6*‡], Jiahe Wang[1*‡] and Ying Liu[7*‡]

[1]Department of Family Medicine, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [2]Science and Technology Research Center of China Customs, Beijing, China, [3]Department of Pulmonary and Critical Care Medicine, Anshan Central Hospital, Anshan, Liaoning, China, [4]Department of Emergency Medicine, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [5]Department of General Practice, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [6]Department of Cardiology, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [7]Department of Nephrology, Shengjing Hospital of China Medical University, Shenyang, China

Introduction: The mortality rate associated with *Mycobacterium tuberculosis* (MTB) has seen a significant rise in regions heavily affected by the disease over the past few decades. The traditional methods for diagnosing and differentiating tuberculosis (TB) remain thorny issues, particularly in areas with a high TB epidemic and inadequate resources. Processing numerous images can be time-consuming and tedious. Therefore, there is a need for automatic segmentation and classification technologies based on lung computed tomography (CT) scans to expedite and enhance the diagnosis of TB, enabling the rapid and secure identification of the condition. Deep learning (DL) offers a promising solution for automatically segmenting and classifying lung CT scans, expediting and enhancing TB diagnosis.

Methods: This review evaluates the diagnostic accuracy of DL modalities for diagnosing pulmonary tuberculosis (PTB) after searching the PubMed and Web of Science databases using the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines.

Results: Seven articles were found and included in the review. While DL has been widely used and achieved great success in CT-based PTB diagnosis, there are still challenges to be addressed and opportunities to be explored, including data scarcity, model generalization, interpretability, and ethical concerns. Addressing these challenges requires data augmentation, interpretable models, moral frameworks, and clinical validation.

Conclusion: Further research should focus on developing robust and generalizable DL models, enhancing model interpretability, establishing ethical guidelines, and conducting clinical validation studies. DL holds great promise for transforming PTB diagnosis and improving patient outcomes.

KEYWORDS

deep learning, pneumonia, tuberculosis, diagnosis, review

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
David Couvin,
Institut Pasteur de la Guadeloupe,
Guadeloupe
Dechao Tang,
University of Münster, Germany
Jindong Li,
University of Erlangen-Nuremberg, Germany

*CORRESPONDENCE
Ying Liu
✉ yaoduo68@foxmail.com
Jiahe Wang
✉ wangjh1@sj-hospital.org
Yuting Wang
✉ 897259567@qq.com

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

‡These authors have contributed equally to this work

RECEIVED 12 October 2024
ACCEPTED 23 December 2024
PUBLISHED 08 January 2025

CITATION
Zhang F, Han H, Li M, Tian T, Zhang G, Yang Z, Guo F, Li M, Wang Y, Wang J and Liu Y (2025) Revolutionizing diagnosis of pulmonary *Mycobacterium tuberculosis* based on CT: a systematic review of imaging analysis through deep learning. *Front. Microbiol.* 15:1510026. doi: 10.3389/fmicb.2024.1510026

1 Introduction

Tuberculosis (TB) is caused by the *Mycobacterium tuberculosis* (MTB), which predominantly targets the lungs, resulting in pulmonary tuberculosis (PTB). TB has coexisted with humans for a thousand years and approximately 1.3 million people died from TB in 2022, according to a recent report by the World Health Organization (WHO) in 2023 (Kaufmann, 2016; Ruiz-Tagle et al., 2024). There have been reports of TB across all age categories and in all nations, making it the second leading infectious killer globally, behind corona virus disease 2019 (COVID-19) (Qi et al., 2024). Furthermore, the WHO has estimated that 10.6 million individuals, including 5.8 million men, 3.5 million women, and 1.3 million children, have been diagnosed with confirmed TB infections globally (WHO, 2022). The progression of TB infection can be divided into four stages: innate immune response, immune balance, TB reactivation, and transmission. The development of effective preventative and treatment methods for this disease depends on the capacity to understand its underlying mechanisms and patterns of progression (Ernst, 2012).

PTB involves multiple pathological processes, including inflammatory exudation, granuloma formation, necrosis absorption, fibrosis, and calcification (Xu et al., 2015). TB's pathogenesis involves transmitting MTB by releasing aerosols containing the bacteria, which occurs when an infected individual coughs or sneezes (Figure 1). These aerosols can be inhaled by another individual, leading to infection. Upon entering the pulmonary system, MTB is phagocytosed by macrophages. Recognition of MTB components by pattern recognition receptors on macrophages, such as Toll-like receptors, initiates an immune response. Macrophages then secrete cytokines that activate T cells and promote the activation and proliferation of macrophages. The aggregation of MTB and immune cells results in granulomas forming, which confine bacterial dissemination. However, compromised immunity can lead to the breakdown of these granulomas, facilitating the recurrence and transmission of TB (Borah et al., 2021).

Artificial intelligence (AI) offers a faster and more convenient technology for predicting the effectiveness of TB treatments, especially in the detection of PTB. AI can improve the precision of medical evaluations by screening, diagnosing, and predicting outcomes with the help of imaging or clinical data (Zhan et al., 2022). A crucial part of AI is machine learning (ML), which works by training models with current data so that they can accurately predict future outcomes given past knowledge (Hussain and Junejo, 2019). Many fields of medical study have used ML, including cancer research, pharmaceutical development, illness detection, and structure of proteins prediction (Yang et al., 2022; Zhu et al., 2024). ML and deep learning (DL) have substantially contributed to computer-aided detection. Yet, DL, particularly convolutional neural networks (CNNs), has risen to prominence for detecting various pulmonary conditions, with a significant focus on diagnosing PTB. In recent years, DL, a subset of ML, has been widely adopted to develop automatic and semi-automatic systems.

DL methods are a type of hierarchical learning of representations. They are better than regular ML because they use many layers of computations to learn patterns that are not linear and have a lot of dimensions. CNNs are a type of DL design that is translation invariant. This means that once a pattern is learned, it can find it anywhere in an image, no matter where it is or how it is oriented. CNNs are often used in DL, a well-known field. These networks have many layers: input, convolutional, pooling, fully connected, and output. They can make specific predictions from digital inputs like images, sounds, genetic sequences, and clinical data (LeCun et al., 2015). Identifying PTB through imaging and ML faces several challenges, including the diverse and sometimes subtle presentations of PTB in imaging, making it difficult for ML algorithms to distinguish it from other lung conditions. ML models may overfit the training data, performing well on the data they were trained on but poorly on unseen data. As new data and knowledge about PTB become available, ML models must be updated and retrained to maintain their accuracy and relevance. Several strategies can address those challenges: employ regularization techniques to mitigate overfitting and enhance model efficacy on novel data. Utilize transfer learning methodologies in which a model initially trained on an extensive dataset is refined on a smaller, more specialized dataset to enhance performance. Extensive testing must be conducted to ensure that ML models are robust against various types of input data and can handle different imaging conditions.

This review examines seven studies about applying DL techniques in identifying PTB (Ma et al., 2020; Zhang et al., 2020, 2024; Li X. K. et al., 2021; Haq et al., 2022; Lu et al., 2022; Huang et al., 2023). The primary focus of the study is to assess the performance of various DL algorithms in diagnosing PTB. Utilize metrics such as precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) to evaluate and compare the diagnostic accuracy of different DL models. Explore strategies for integrating DL tools into clinical practice and identify the critical areas for future enhancement of the methodologies above. The outline of the review is as follows: section 2 describes the main methods of this review. Section 3 describes the traditional detection methods for PTB. Section 4 presents the DL diagnosis process for TB. Section 5 depicts the applications of DL in CT-based PTB detection. Section 6 discusses the future directions. Finally, section 7 summarizes the conclusions drawn from the review. Prior studies have not yet offered an integrated, comprehensive analysis of detecting PTB using DL alongside imaging modalities

---

Abbreviations: TB, tuberculosis; MTB, *Mycobacterium tuberculosis*; PTB, pulmonary tuberculosis; WHO, World Health Organization; COVID-19, corona virus disease 2019; CT, computed tomography; AI, artificial intelligence; ML, machine learning; DL, deep learning; CNN, convolutional neural network; AUC, area under the receiver operating characteristic curve; PRISMA, preferred reporting items for systematic reviews and meta-analyses; CAP, community-acquired pneumonia; TST, tuberculin skin test; IGRA, interferon-gamma release assay; QFT, QuantiFERON-TB Gold; LAM, lipoglycan lipoarabinomannan; SPECT, single-photon emission computed tomography; PCR, polymerase chain reaction; NAA, nucleic acid amplification; LAMP, loop-mediated isothermal amplification; LPA, line probe assay; CXR, chest X-ray; US, ultrasound; MLR, monocyte to lymphocyte ratio; NGS, next-generation sequencing; LSTM, long short-term memory; RNN, recurrent neural network; GAN, generative adversarial network; ROI, region of interest; GMM, Gaussian mixture model; JI, Jaccard index; DSC, Dice similarity coefficient; ROC, receiver operating characteristic; ANN, artificial neural network; MLP, multi-layer perceptron; MAResNet, multi-scale attention residual network; CBAM, Convolutional Block Attention Module; NTM-LD, non-tuberculous mycobacterial lung disease; mSv, millisievert; ULDCT, ultra-low-dose computed tomography; MDR-PTB, multidrug-resistant PTB; Grad-CAM, gradient-weighted class activation mapping.

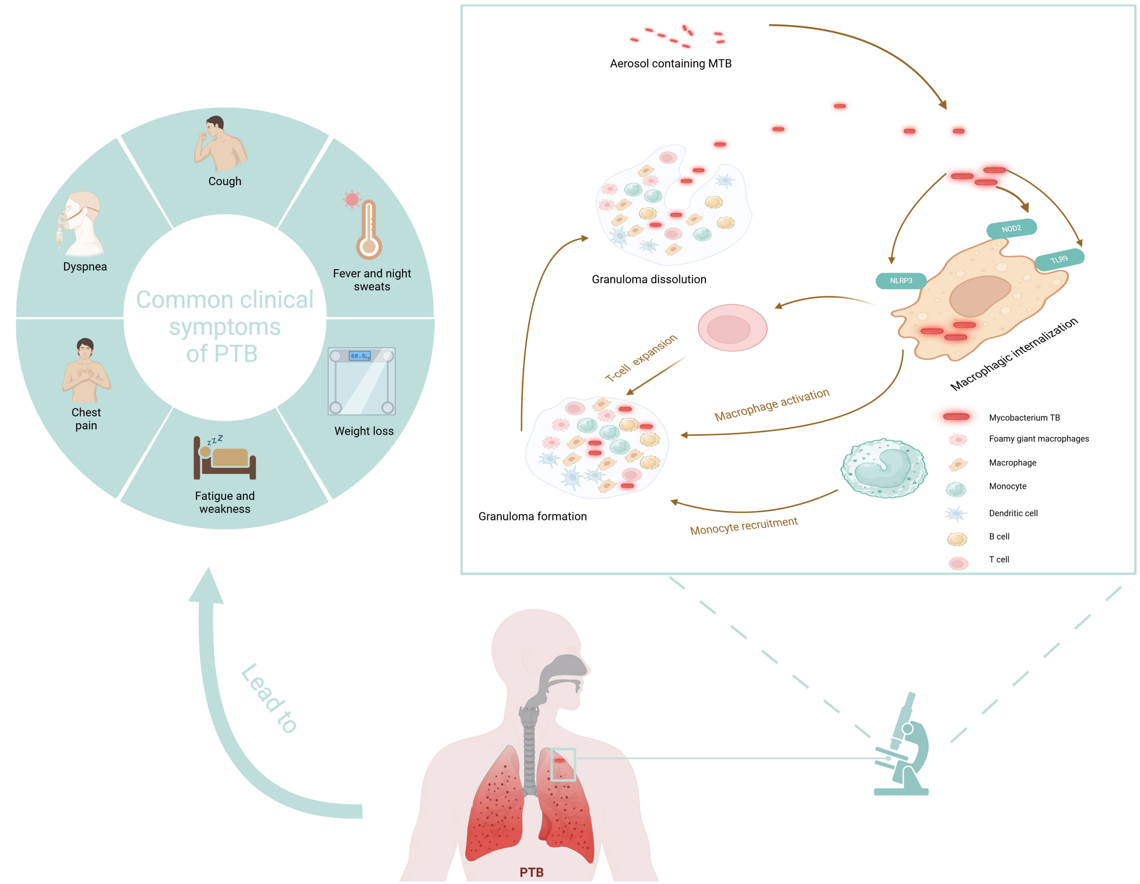


FIGURE 1
Pathogenesis and typical symptoms of TB. Created in BioRender. Zhang (2024a).

datasets. This study examines the methodologies, procedures, and techniques of DL and imaging modalities.

2 Materials and methods

This systematic review aimed to assess the role of DL in diagnosing PTB based on CT imaging. The preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines were used in conducting this research (Page et al., 2021). These guidelines were instrumental in structuring the study selection, offering a standardized framework to streamline and document articles' identification, screening, eligibility, and inclusion. A thorough search was conducted among PubMed and Web of Science databases. The search query was formulated using the PICO strategy and included the terms "tuberculosis," "artificial intelligence," "machine learning," "deep learning," "neural network," and "natural language processing" (Cacciamani et al., 2023). In this study, the inclusion criteria were developed by the research question: in individuals undergoing diagnosis for PTB using CT scans, how effective are diagnostic models utilizing DL techniques for image analysis? This question was formulated using the PICO strategy. The population (P) comprised individuals undergoing diagnosis for PTB using CT scans, with the intervention (I) being the application of DL techniques for image analysis, and while there was no direct comparison group (C), comparisons were made with traditional diagnostic methods; the outcomes (O) focused on quantitative data regarding the performance of DL models, including primary outcomes such as accuracy, specificity, and sensitivity in diagnosing PTB. The exclusion criteria were non-English articles, letters, and reviews. Studies not related to PTB or not using CT scans for diagnosis. Studies lack a clear description of the DL methodology (studies without original data or not providing performance metrics).

Fei Zhang and Maomao Li screened the titles and abstracts of the identified studies, and full texts were retrieved for further analysis. Any discrepancies were resolved through consensus and by consulting Jiahe Wang. Data were extracted using a predefined form to collect details on the author's name, paper publication year, journal, country of the dataset, number of patients (with male/female distribution), study purpose, type of DL algorithm, dataset source, validation methods, reference standard, and the reported performance. The research search yielded 1,643 records; following the title and abstract screening, 120 records underwent full-text evaluation, including six publications in the review. Moreover, following citation analysis of pertinent documents, an additional article was incorporated, increasing the total number of included articles to seven. A PRISMA flow diagram was created to depict the article selection process, specifying the number of records obtained from all sources and the explanation for exclusion

(Figure 2). The QUADAS-2 instrument was utilized to appraise the potential for bias within the studies under consideration (Table A1). This framework assesses bias across four principal areas (Whiting et al., 2011). Figure 3 illustrates the yearly number of publications on AI in TB from 2005 to 2024, there was a rapid increase in the publication of studies.

3 Traditional detection methods for PTB

3.1 Clinical symptoms and physical examination diagnosis

Acute PTB often presents with acute respiratory symptoms, including dry cough, fever, and chest pain (Moreira et al., 2011). Symptom duration generally surpasses 2 weeks before hospitalization. The clinical appearance is comparable to non-tuberculous community-acquired pneumonia (CAP), but patients with PTB experience less intense pleural unease, toxemia, and fatigue than those with non-tuberculous bacterial pneumonia (Sharma and Mohan, 2004). Acute PTB is more likely to cause weight loss than non-tuberculous CAP, while hemoptysis is relatively rare (Morena et al., 2023). The typical symptoms of TB are vividly shown in Figure 1. Generally speaking, the diagnosis of PTB is made using a combination of conventional and contemporary techniques. This diagnosis depends primarily on the patient's medical symptoms and findings from the physical examination, complemented by various diagnostic test results. These include bacteriological tests, the tuberculin skin test (TST), imaging studies such as X-ray or CT scans, histopathological evaluations, T-SPOT test, and the reaction to the therapy plan of antituberculosis medications (Kang et al., 2021). The distribution of mycobacterium TB in the human body is shown in Figure 4.

3.2 Etiologic diagnosis

The WHO regards TB culture as the "gold standard" for diagnosing TB; nonetheless, conventional solid and liquid culture media exhibit several drawbacks, such as impracticality, bacterial cross-contamination, and extended culture durations. Sputum smear microscopy is a diagnostic technique for TB that is widely acknowledged as effective (Kessel et al., 2023, pp. 2013–2017). Due to its affordability and relative simplicity compared to other advanced diagnostic methods, sputum smear microscopy remains a critical diagnostic instrument for PTB, particularly in nations with low incomes. In this process, sputum samples are coughed up by patients exhibiting symptoms and are treated with chemicals and applied to plain glass microscope slides. Yet the inconsistent clinical performance of this method, combined with the challenges in sputum collection from patients and accessibility to healthcare services, constitutes one of the main reasons for TB being undiagnosed. Subsequently, these slides are subjected to laboratory analysis to detect the presence of TB. The resulting images from a sputum smear test are typically viewed using fluorescence or light microscopy. The resolution and size of these images are determined by the level of magnification employed. For laryngeal swabs, the pooled sensitivity was 57.8% (95% CI: 50.5–65.0), and the specificity was 93.8% (88.4–96.8). For nasopharyngeal aspirates, the sensitivity was 65.2% (95% CI: 52.0–76.4), and the specificity was 97.9% (95% CI: 96.0–99.0). For oral swabs, the sensitivity was 56.7% (95% CI: 44.3–68.2), and the specificity was 91.3% (95% CI: 81.0–96.3) (Savage et al., 2023). Fluorescence microscopy offers advantages such as labor reduction and enhanced productivity. Nevertheless, the potential drawback of this technique is the danger of false-positive results, which is a result of the fluorochrome dyes' non-specific binding (Steingart et al., 2006). The non-invasive collection process and its association with TB transmission have long been why breath has been regarded as an appealing diagnostic specimen for TB (Ghosh et al., 2021; Pham and Beauchamp, 2021). Electronic nose tests reportedly have an estimated sensitivity of 92% (95% CI: 82–97%) (Saktiawati et al., 2019).

3.3 Immunological diagnosis

Serological assays frequently demonstrate inadequate sensitivity and specificity, relying on a humoral immune reaction to detect antibodies against TB antigens (Steingart et al., 2011). A strategy attracting increased attention is the identification of host responses indicative of TB infection. In this regard, interferon-gamma release assays (IGRAs), such as T-Spot (Oxford Immunotec) and QuantiFERON (Qiagen), have limited use in detecting acute infections but are effective in detecting latent TB. However, IGRAs are affected by diseases such as diabetes, and the high cost also limits its application in underdeveloped regions (Takasaki et al., 2018). Currently, the T-SPOT test is extensively utilized for diagnosing infections caused by TB (Sollai et al., 2014). Despite its importance in identifying Mtb infections, a significant drawback of the T-SPOT test is that it cannot differentiate between active TB and latent TB infection (Ling et al., 2013).

The benefits and constraints of the QuantiFERON-TB Gold (QFT) test parallel those of the T-SPOT test. The QFT test offers a more straightforward operational process than T-SPOT, eliminating the need to separate peripheral blood mononuclear cells and opting to use whole blood instead (Lalvani and Pareek, 2010). Only a handful of antigenic biomarkers have been identified for TB. Lipoglycan lipoarabinomannan (LAM) is the most extensively investigated, most promising, and accessible from a simple sample such as urine (MacLean et al., 2019). The FujiLAM is a lateral flow urine test that finds LAM antigens. In adults with TB, it has a sensitivity of 70% and a specificity of 93% (Li Z. et al., 2021). The TST is a traditional diagnostic technique that uses a pure protein derivative of tuberculin to identify delayed-type hypersensitivity responses. Individuals infected with TB can generate sensitized T cells that recognize MTB antigens. Upon re-stimulation by MTB antigens, these sensitized T cells secrete different soluble lymphokines that enhance permeability, local erythema, and induration (Kowalewicz-Kulbat et al., 2018). The TST method finds the average diameter of the induration 72 h after an injection of a pure protein derivative tuberculin. An induration diameter <5 mm or no reaction is negative; ≥ 5 mm is positive (Abubakar et al., 2018).

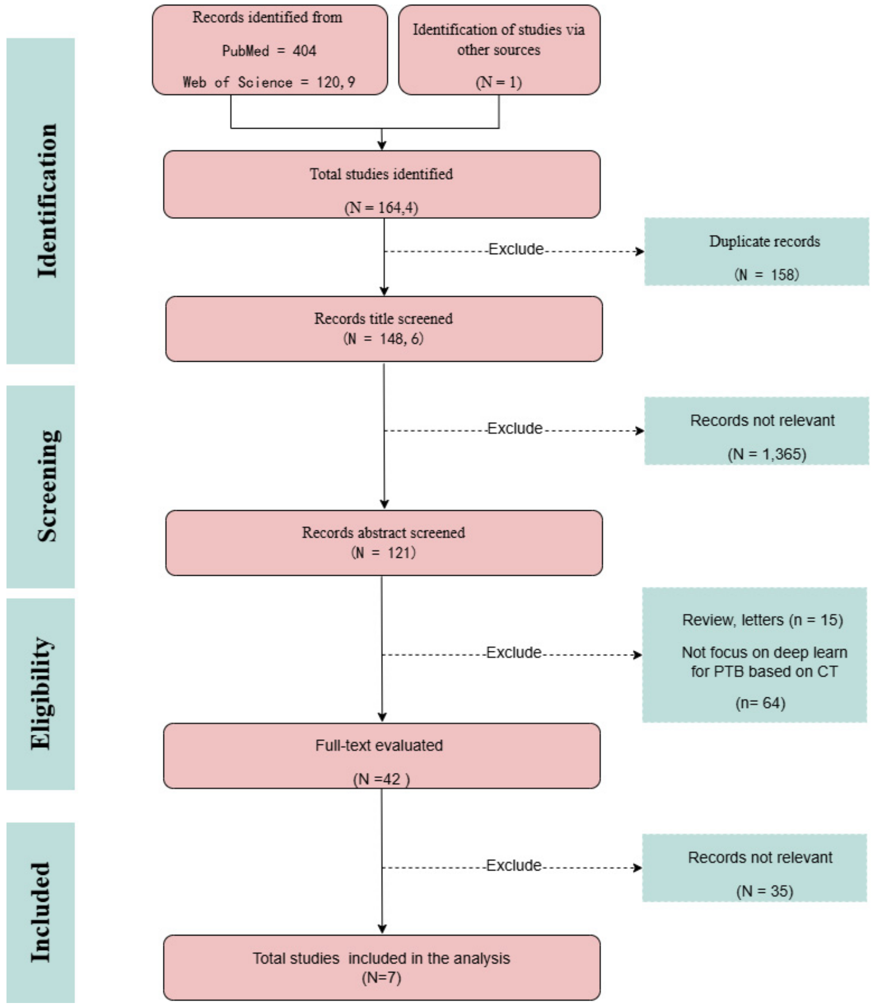
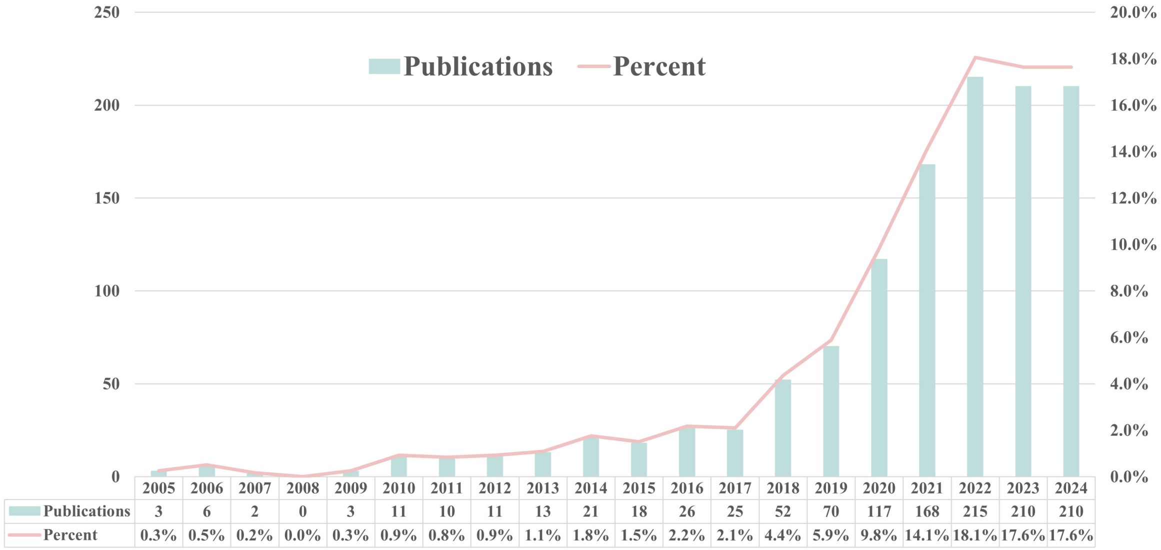


FIGURE 2
PRISMA study flow diagram.



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Publications	3	6	2	0	3	11	10	11	13	21	18	26	25	52	70	117	168	215	210	210
Percent	0.3%	0.5%	0.2%	0.0%	0.3%	0.9%	0.8%	0.9%	1.1%	1.8%	1.5%	2.2%	2.1%	4.4%	5.9%	9.8%	14.1%	18.1%	17.6%	17.6%

FIGURE 3
Annual number of publications of AI in TB from 2005 to 2024.

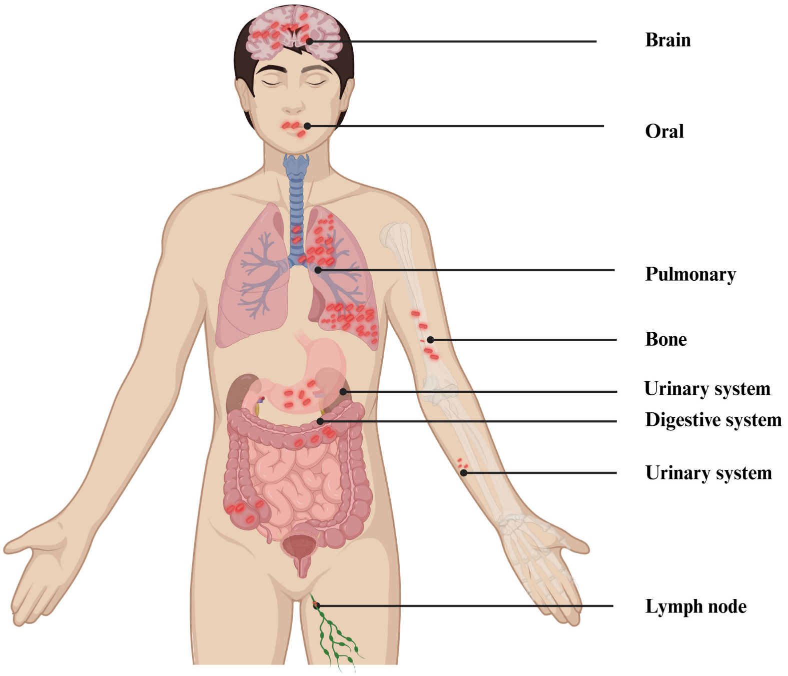


FIGURE 4
Distribution of MTB in the human body. Created in BioRender. Zhang (2024b).

3.4 Molecular techniques

Molecular imaging, which integrates molecular biology with medical imaging, is increasingly being explored to enhance our understanding of PTB. This high-tech imaging method allows biological processes at the molecular and cellular levels inside live things to be witnessed and measured. Techniques such as single-photon emission computed tomography (SPECT) are being investigated for their potential to offer detailed molecular-level data. SPECT's high sensitivity and resolution make it a promising tool for identifying and tracking disease processes associated with PTB (Dimastromatteo et al., 2018). Along with smear microscopy, the Centers for Disease Control and Prevention suggests that each individual who might have PTB should also have at least one nucleic acid amplification (NAA) test, like polymerase chain reaction (PCR) (Parrish and Carroll, 2011). The high guanine and cytosine in the TB genome make it harder to conduct PCR studies. As a result, when dealing with MTB, it is essential to meticulously consider the methods for sample collection, bacterial cell disruption, nucleic acid isolation, and PCR assay design. The "Xpert MTB/RIF assay" is a rapid NAA test capable of detecting TB and determining resistance to rifampicin (Naidoo et al., 2023). The sensitivity of Xpert MTB/RIF Ultra when applied to oral swabs varied between 45% and 77.8%, in contrast to an approximate sensitivity of 90% for sputum samples (Mesman et al., 2019; Lima et al., 2020; Andama et al., 2022). Identifying bacterial RNA allows for the pinpointing of active TB. The quantification of bacterial quantity in sputum samples can be achieved using the detection of 16S ribosomal RNA, providing a sensitivity comparable to that of solid culture techniques (Honeyborne et al., 2014). Loop-mediated isothermal amplification (LAMP) is a PCR method that functions at a uniform temperature. The TB-LAMP method exhibits a sensitivity marginally lower than the Xpert MTB/RIF assay, yet both tests maintain similar specificity (Phetsuksiri et al., 2020). The WHO advocates TB-LAMP as a superior alternative to smear microscopy because of its improved diagnostic efficacy (Huang et al., 2022). A line probe test (LPA) is a quick, accurate, and flexible way to determine if someone has TB. It can detect TB in various clinical specimens, including sputum, pleural, and cerebrospinal fluid. The LPA can identify resistance to first-line TB medications such as isoniazid. Different commercial LPA kits are available, including the GenoType MTBDRplus 1.0 assay from Hain Lifescience and the INNO-LiPA Rif TB kit offered by Innogenetics (Rossau et al., 1997; Crudu et al., 2012). LPA is crucial for the management of multidrug-resistant TB. Truenat MTB, Truenat MTB Plus, and Truenat MTB-Rif Dx assays are quick molecular real-time PCR tests that can find TB. Results are usually ready in an hour (Nikam et al., 2013; Georghiou et al., 2021).

3.5 Imaging techniques

Most individuals with PTB exhibit abnormal chest X-ray (CXR) findings, which indicate a PTB diagnosis (Nachiappan et al., 2017). Given the relative accessibility of CXR and its utility in identifying these signs, the WHO advises using chest radiography for TB screening in populations at high risk for the disease (Liang et al., 2022). Different TB lesions appear differently on X-rays; exudative lesions appear as cloud-like or patchy shadows, proliferative lesions as nodular shadows, and caseous lesions as high-density and uneven shadows. The advantages of CT scanning include its relatively low cost, enhanced capability to differentiate between tissue types, rapid image acquisition, and broader accessibility. Compared to X-ray chest films, CT scans provide more explicit sectional images of the lungs, avoiding the issue of overlapping pictures and displaying the delicate structures and lesion details of lung tissue. Various pathological changes of TB, such as exudation, proliferation, caseation, fibrosis, and calcification, can be well displayed on CT scans. CT can detect early small lesions, bronchial dissemination foci, and mediastinal lymph node enlargement, aiding in a definitive diagnosis. It is also useful in diagnosing suspected TB cases that are atypical or negative on X-ray chest films. Additionally, CT has an extra role in assisting with fluid aspiration, biopsy confirmation, and guiding therapeutic interventional procedures, such as fluid drainage. PET/CT, which utilizes the cellular uptake of ^{18}F-fluorodeoxyglucose to assess pulmonary inflammation, is highly sensitive for the early detection of TB (Ankrah et al., 2018). X-ray chests are the initial step in investigating PTB, followed by ultrasound (US), CT, and magnetic MRI for further evaluation. Additional imaging modalities, such as intravenous urography and barium studies, may also be helpful.

3.6 Other techniques

Innovative techniques are being developed to evaluate, track, and quantify PTB conditions at the point of care. Techniques such as lung ultrasonography and electrical impedance tomography are becoming more popular as they complement traditional diagnostic methods. These methods are being extensively researched for their potential to supplement standard procedures and, in certain respiratory conditions, to serve as an alternative due to the absence of ionizing radiation and their simplicity (Ball et al., 2017). A meta-analysis has revealed that chest US, when used for diagnosing pediatrics PTB, has a sensitivity of 84% and a specificity of 38% (Muljadi et al., 2024). Many molecular biomarkers can also be used to diagnose PTB. The combined sensitivity and specificity of monocyte to lymphocyte ratio (MLR) in detecting TB are 79.5% (95% CI: 68.5–87.3) and 80.2% (95% CI: 67.3–88.9), respectively (Adane et al., 2022).

Non-sputum biomarker tests for TB could have a market value of between \$56 million and \$84 million in countries with a high incidence of TB, like South Africa, Brazil, China, and India. It is thought that 14 million tests will be done (Tb Diagnostics Market Analysis Consortium, 2014, 2015; Maheshwari et al., 2016; Zhao et al., 2016). Dai et al. (2019) made a model for predicting TB by testing three iron-related biomarkers in blood serum, which can recognize TB. There is optimism that this approach could be broadened to enhance the diagnostic techniques for PTB. Next-generation sequencing is considered a revolutionary method for medication susceptibility testing of TB, providing data much faster than conventional clinical culture-based methods (Walker et al., 2015). Mass spectrometry technology can accurately detect biomarkers, which helps with the early diagnosis of TB. Chen et al. (2020) utilized Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry and Liquid Chromatography-Tandem Mass Spectrometry to identify a TB-specific serum peptide signature, thereby creating diagnostic models for swift and accurate TB detection. More details can be found in Table 1, which compares various TB diagnostic methods.

4 DL diagnosis process for TB

4.1 Overview of DL architectures

Deep learning neural networks, a category of computational models, can learn complex feature hierarchies by deriving advanced features from simpler ones. Fukushima introduced this concept in 1980, inspired by the mechanisms of human vision based on biological principles (Yoon and Kim, 2020). DL emulates the human brain's process of information filtering to facilitate accurate decision-making. DL instructs a computational model to handle inputs through a series of layers, analogous to the human brain's approach, to bolster data prediction and categorization. Each layer feeds its output to the subsequent layer, akin to the progressive filtering mechanisms employed by neural networks within the brain. The iterative feedback process continues until the production remains consistent with the previous iteration. Weights are first allocated to each layer to produce the desired output, and these weights are further refined during the training process to attain the exact result (Lalmuanawma et al., 2020).

In the fields of CT-based disease diagnosis, multiple DL models are frequently utilized for tasks. CNNs like VGGNet, Google Net, ResNet, and DenseNet are essential for image categorization and feature extraction (Pavithra et al., 2023). U-Net, V-Net, SegNet, and DeepLab are commonly employed for segmentation jobs because of their proficiency in accurately delineating regions of interest (ROIs) (Yuan et al., 2022). Detection and classification jobs frequently employ models such as YOLO, SSD, and the R-CNN series, which are proficient in recognizing and categorizing objects in photos (Vilcapoma et al., 2024). GANs and their derivatives, such as DCGAN, cGAN, CycleGAN, and Pix2Pix, are utilized for data augmentation, picture reconstruction, and style transfer, improving training data's diversity and quality (Simion et al., 2024). Three-dimensional (3D) CNNs and their variants, including 3D U-Net, are explicitly engineered for volumetric data, which is essential for examining 3D structures in CT scans (Wen et al., 2023). Furthermore, sophisticated models such as Dual-Path Networks and those developed by neural architecture search are reviewed to enhance performance in CT image processing, guaranteeing that the models are efficient and precise (Min et al., 2024).

Deep learning comprises three fundamental methodologies: supervised, semi-supervised, and unsupervised (Liporaci et al., 2024). For medical image analysis tasks, including disease detection and classification, CNNs have grown into the predominant DL architecture (Odimayo et al., 2024). A schematic representation of a typical CNN architecture is shown in Figure 5. Widely utilized during the diagnosis of numerous illnesses, CNNs specialize in extracting pertinent information from medical pictures, including X-rays, CT scans, and MRI scans. While CNNs excel at processing spatial information in medical images, recurrent neural networks (RNNs) and their variants, such as long short-term memory, have proven effective in handling sequential data, which is often encountered in healthcare settings (Chae et al., 2024). Generative adversarial networks (GANs) have emerged as a powerful tool to address the limited availability of labeled data by generating synthetic medical data that can be used to augment the training datasets (Su et al., 2023). Among these techniques, GANs stand out for their unique capabilities in data augmentation and image

TABLE 1 Comparison of TB methods.

Detection method	Methodology	Interpretation	Shortcomings
X-ray	A chest X-ray is taken to visualize the lungs and chest cavity.	Detects lung shadows, cavities, and calcifications.	Difficulty in distinguishing between active and latent TB, limited value for diagnosing extrapulmonary TB.
CT	A CT scan of the chest is performed to obtain detailed images of the lungs.	Detects lung inflammation, nodules, and cavities with precise localization.	Higher cost, radiation exposure, misinterpretations.
MRI	MRI scan of the chest is performed to visualize the lungs and chest cavity.	Detects lung inflammation, nodules, cavities, and soft tissue lesions.	Higher cost, less effective in detecting calcifications and bony lesions compared to X-ray.
Ultrasound	Ultrasound scan of the chest is performed to visualize the lungs and chest cavity.	Detects lung effusions, consolidations, and pleural effusions.	Requires skillful operator, lower specificity for diagnosing TB.
Bronchoscopy	A bronchoscope is used to visualize the bronchi and lung tissue, and samples are collected.	Direct visualization of lung lesions and pathological examination.	Invasive procedure with associated risks.
Tissue biopsy	Tissue samples are collected from the lungs or lymph nodes for pathological examination.	Confirms TB diagnosis and performs drug susceptibility testing.	Invasive procedure with associated risks.
Tuberculin skin test	Tuberculin is injected under the skin, and the skin reaction is observed.	Detects the immune response to *Mycobacterium tuberculosis* to determine past infection.	Cannot distinguish between active and latent TB, limited value in vaccinated populations.
Sputum smear test	Sputum is stained and examined under a microscope for the presence of *Mycobacterium tuberculosis*.	Rapid detection of active TB, low cost.	Lower sensitivity, requires multiple tests.
Fluorescent microscopy	Sputum is stained with a fluorescent dye and examined under a fluorescent microscope for the presence of *Mycobacterium tuberculosis*.	Improves the sensitivity of sputum smear test.	Higher cost, requires specialized equipment.
Culturing bacteria	Sputum or tissue samples are cultured on selective media to grow *Mycobacterium tuberculosis*.	Confirms TB diagnosis and performs drug susceptibility testing.	Higher cost, requires longer time.
Xpert MTB/RIF	Molecular diagnostic technique that detects *Mycobacterium tuberculosis* DNA in sputum and determines drug resistance.	Rapid, accurate, simultaneous detection of drug resistance.	Higher cost, requires specialized equipment.
LAMP	Rapid diagnostic technique based on nucleic acid amplification that amplifies *Mycobacterium tuberculosis* DNA within a short time.	Rapid, easy to operate.	Slightly lower specificity than Xpert MTB/RIF.
LPA	Molecular diagnostic technique based on DNA probes that detects drug resistance genes of *Mycobacterium tuberculosis*.	Rapid, accurate detection of drug resistance.	Requires specialized equipment, more complex operation.
Micro real-time PCR	Rapid diagnostic method based on real-time fluorescent quantitative PCR that amplifies and detects *Mycobacterium tuberculosis* DNA within a brief time.	Rapid, accurate, detection of drug resistance.	Requires specialized equipment, higher cost.
Next-generation sequencing	Molecular diagnostic method based on high-throughput sequencing that comprehensively analyzes the genetic information of *Mycobacterium tuberculosis*.	Detection of drug resistance and identification of new resistance genes.	Very high cost, requires specialized bioinformatics analysis.
Mass spectrometry	Molecular diagnostic method based on mass spectrometry that detects proteins or metabolites of *Mycobacterium tuberculosis*.	Detection of drug resistance, identification of new resistance mechanisms.	Very high cost, requires specialized mass spectrometry analysis.
IGRAs	Detects the level of interferon-γ in the blood to determine *Mycobacterium tuberculosis* infection.	Detection of latent TB infection.	Requires laboratory conditions, cannot distinguish between active and latent TB.
Antibody detection	Detects *Mycobacterium tuberculosis*-specific antibodies in the blood.	Assists in the diagnosis of TB.	Lower specificity, easily affected by other factors.

LAMP, loop-mediated isothermal amplification; LPA, line probe assay; IGRAs, interferon-gamma release assays.

quality enhancement. GANs consist of two neural networks, a generator and a discriminator, that engage in a competitive training process. While the discriminator assesses their veracity against actual data samples, the generator generates new data instances. The generator learns to create more realistic data because of this adversarial process, which is especially useful in medical imaging where class imbalance and data scarcity are frequent problems. GANs can produce synthetic CT images of PTB that resemble actual TB lesions, increasing the data available for DL model training. This improves the model's generalization across various patient populations and imaging settings and reduces overfitting. Semi-supervised learning methods, which leverage labeled and unlabeled data, have shown promising results in tasks like medical image segmentation and disease classification (Tseng et al., 2024). These techniques can help improve model performance when the labeled data is limited. Furthermore, unsupervised learning approaches, such as clustering and anomaly detection, have been utilized to identify novel disease subtypes, detect rare diseases, and uncover hidden patterns in medical data (Thabtah et al., 2022). These methodologies can yield essential insights for physicians and scientists, potentially facilitating the identification of novel disease biomarkers and enhanced patient screening.

4.2 Overview of the pipeline for PTB detection based on CT

4.2.1 Data acquisition and pre-processing

The lack of CT data might lead to data overfitting and affect the efficiency of the imaging model. Studies by Li X. K. et al. (2021) increased the number of training samples through random cropping and left-right flipping (Wu et al., 2019). A workflow for PTB diagnosis using DL based on a CT pipeline is shown in Figure 6. The initial phase of image preparation entails transforming the unprocessed images into a suitable format for subsequent analysis. Medical imaging data from various equipment exhibit dimensions, layer thickness, and scan count differences. Collectively, these variables result in a diverse array of imaging datasets, causing inconsistencies among the data sets. For accurate classification of medical images, the preprocessing phase should significantly minimize noise without compromising the integrity of vital image elements. Consequently, the preprocessing phase consists of resizing, normalizing, and occasionally converting color images from RGB to grayscale. Additionally, images are enhanced using techniques such as Gaussian blurring, median filtering, morphological smoothing, and various other methods for image adjustment.

4.2.2 Feature extraction and classification

The process of transforming images into features that reflect various image attributes is called feature extraction. The success of DL models in CT-based diagnosis is heavily dependent on the quality of annotated data used for training and validation. In this systematic review, CT scans were annotated by experienced radiologists or pulmonary disease specialist following standardized clinical guidelines (Ma et al., 2020; Zhang et al., 2020, 2024; Li X. K. et al., 2021; Haq et al., 2022; Lu et al., 2022; Huang et al., 2023). Image segmentation is indeed a pivotal step in image processing. It involves partitioning images into distinct sections or ROIs to isolate and analyze specific features or objects within the image (Shahzad et al., 2024). Several types of features are used in image analysis. These include texture, shape, contrast, and brightness (Kaifi, 2023). Slices exhibiting PTB lesions and regular slices devoid of pathological findings were individually marked manually and employed as the benchmark dataset to train the DL model. Some open-source tools, such as ITK-SNAP, delineate bounding boxes around CT imaging lesions (Yan et al., 2021b). This process entails identifying and segmenting the lesions from the scans. Nevertheless, manual segmentation of the lung region is a laborious, monotonous, and time-consuming endeavor that significantly depends on the proficiency and experience of radiologists. Feature extraction is a cornerstone of diagnostic imaging, particularly when utilizing DL to analyze CT scans.

Gordaliza et al.'s (2019) study offers a glimpse into the potential of unsupervised learning for lung image segmentation. According to Wang et al. (2023), they created a GAN-based design that can separate different lung lesions. GAN is an ML model made up of a generator and a discriminator. It is often used to create images and split them into groups. This model can identify and segment multiple lesion areas present in CT scans. In a GAN, the discriminator may experience "forgetting," which means losing the ability to recognize certain features during training. They implement a method to mitigate this forgetting phenomenon. It introduces a self-supervised rotation loss to help address the issue of discriminator forgetting. Self-supervised learning is free of data to be labeled by hand, and rotation loss might include flipping pictures to aid the model in learning better. The recommended approach achieved Dice coefficients of 68.5% on test datasets for multi-center PTB. The architecture consists of a dual attention module and a cascaded context-aware pyramid feature extraction method, making it possible to understand the semantic dependencies linked to lung lesion characteristics in space and time. This unified method makes the model's training more effective. The study by Gordaliza et al. (2019) develops a methodology for the automated extraction of a radiological biomarker from CT scans to assess the disease burden of TB, which may also be modified for pneumonia identification. The pipeline involves lung segmentation, tissue type classification, and applying a Gaussian mixture model (GMM) to differentiate between healthy and diseased tissue. The process consists of using an adaptive thresholding method to identify air-like organs in chest CT scans, such as healthy lungs, the airway tree, and the stomach, by utilizing the topological properties of the organs. Geodesic Active Contours are pivotal in refining the lung boundaries by including lesions attached to the pleura and discarding motion artifacts. Furthermore, the GMM is employed to model the probability distribution of voxel intensities within the segmented images. By assuming that the tissue intensities follow a Gaussian mixture, the GMM, in conjunction with the Expectation-Maximization algorithm, allows for the automatic computation of thresholds that distinguish different tissue types. This statistical approach provides a robust framework for classifying lung tissue based on its intensity values.

4.2.3 Performance evaluation

The effectiveness of the entire pipeline is measured using evaluation metrics such as precision, accuracy, recall, specificity, F1-score, and AUC, among others. The training subset is utilized to generate a specific model. In contrast, the suitability of the training process and the model is evaluated by simultaneously observing overfitting or underfitting on the validation subset. Ultimately, the unseen testing subset is used to judge the performance to which the created model works. Sensitivity is the ratio of accurate positive results to the actual positive cases. Specificity refers to the proportion of true negative cases that are accurately recognized as such. The Jaccard index (JI) is a percentage that shows how

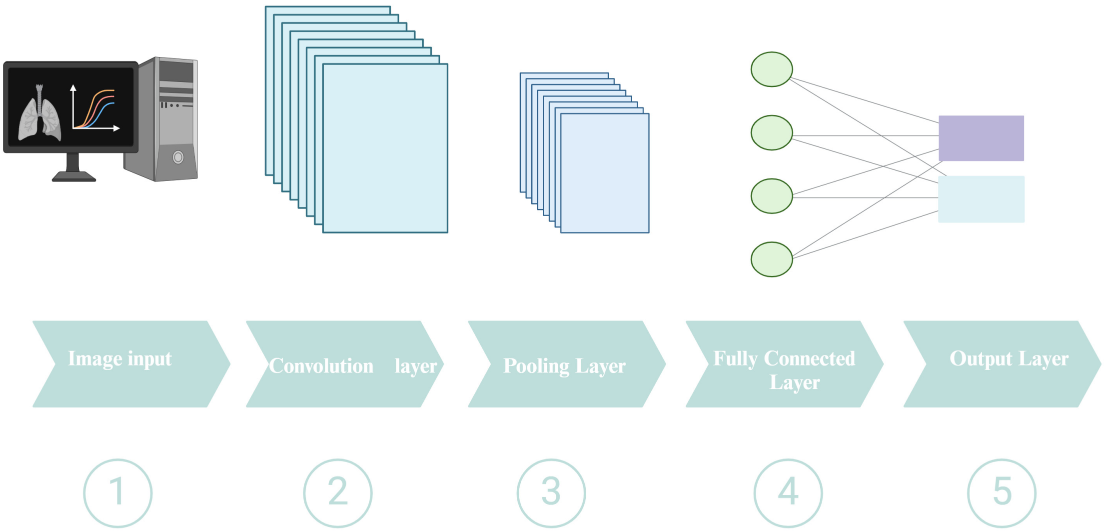


FIGURE 5
Schematic representation of a typical CNN architecture. Created in BioRender. Zhang (2024c).

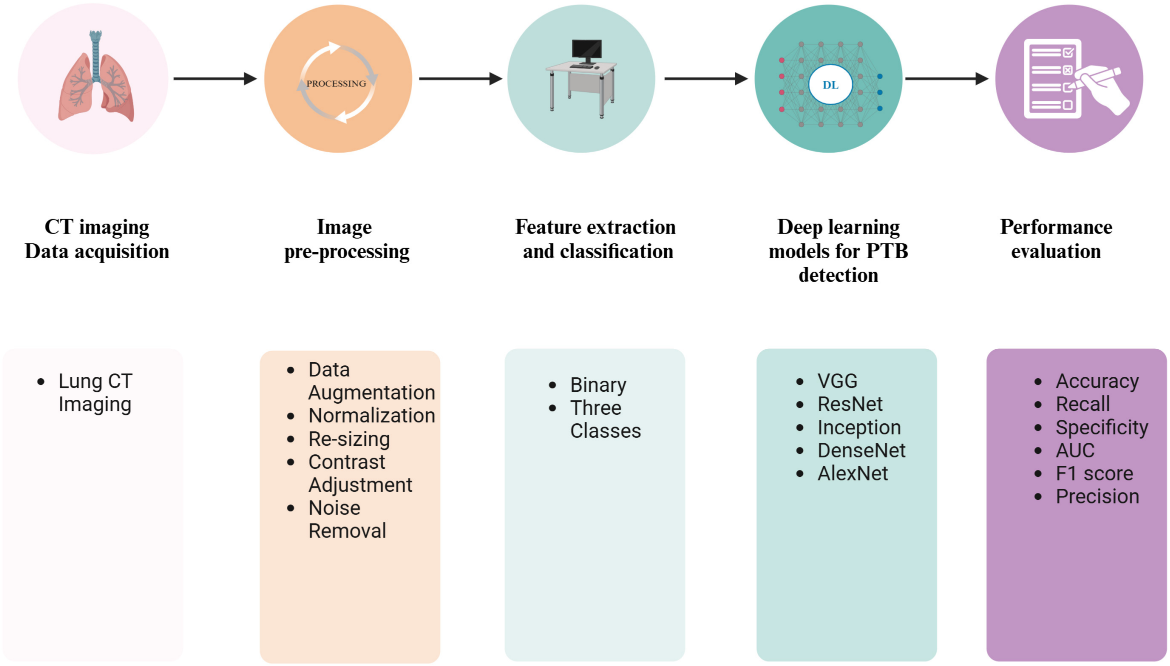


FIGURE 6
A workflow for PTB diagnosis using DL based on a CT pipeline. Created in BioRender. Zhang (2024d).

much the model's predicted output and the accurate annotation ground-truth mask match. The similarity index measures the unity between the segmentation generated by the model and the expert-annotated ground truth. It evaluates the extent to which the model's delineation of the PTB region aligns with the input image's actual PTB area. A Dice similarity coefficient (DSC) of zero indicates no spatial overlap between the model's annotations and the actual PTB location, while a DSC of one signifies perfect spatial overlap. The AUC summarizes the receiver operating characteristic (ROC) curve. The ROC curve compares the sensitivity to the false positive rate to see how well a classifier can tell the difference between classes. Additional details can be found in Table 2, which provides some standard performance metrics for DL models.

4.2.4 DL-related concepts

Deep learning is a subset of ML that focuses on deep artificial neural networks (ANNs). Common types of DL algorithms encompass multi-layer perceptrons (MLPs), CNNs, RNNs, graph neural networks, Transformers, and more. Overfitting is a

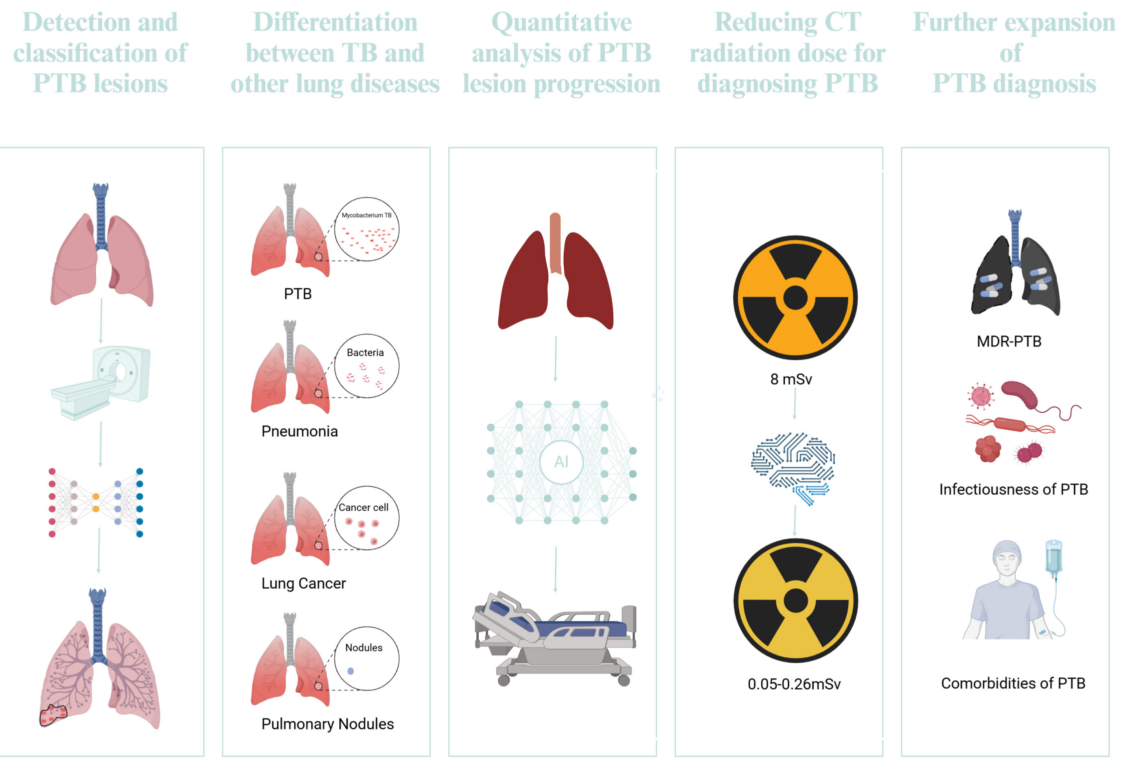


FIGURE 7
The application of DL in TB management. Created in BioRender. Zhang (2024e)

TABLE 2 Common performance metrics for DL models.

Metric	Description	Formula
Accuracy	The ratio of correctly predicted instances to the total instances	Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision	The ratio of true positive predictions to the total positive predictions	Precision = TP / (TP + FP)
Recall	The ratio of true positive predictions to the total actual positives	Recall = TP / (TP + FN)
Specificity	The ratio of true negative predictions to the total actual negatives	Specificity = TN / (TN + FP)
Kappa	A statistic that measures inter-rater agreement for categorical items	Kappa = $(P_o - P_e) / (1 - P_e)$[1]
F1-score	The harmonic mean of precision and recall, balancing both metrics	F1-score = 2 × precision × recall / (precision + recall)
Dice similarity coefficient	A measure of overlap between two sets, often used in image segmentation tasks	Dice = (2 × TP) / (2 × TP + FP + FN)

[1] P_o is observed agreement, and P_e is expected agreement.

modeling error that arises when a model learns the random noise and fluctuations in the training data to the extent that it negatively impacts the model's performance on new, unseen data. Essentially, the model becomes too tailored to the training set and fails to generalize well to independent data sets, such as those used for testing. Cross-validation is a way to see if the outcomes of a statistical test can be applied to a different data set. It is mainly used when the goal is to make a prediction and figure out how well a prediction model will work in real life. In *k*-fold cross-validation, the original sample is split into *k* subsamples of the same size. Only one *k* subsamples are kept as confirmation data to test the model, and the other $k - 1$ subsamples are used as training data. After that, this process is done *k* times, and each *k* subsample is used only once as confirmation data. In leave-one-out cross-validation, one observation from the sample is used as the validation set, and the rest are used as the training set. This is a type of *k*-fold cross-validation, where *k* is the total number of data points. Because this is done repeatedly, each measurement in the dataset is used as the validation set a single time. The same dataset is used for training and validation in cross-validation, so it is an internal validation method. External validation, on the other hand, uses a different set of data that was not used to train and test the model in the first place. This could involve data from a different time, location, or group of subjects. In the bootstrap validation technique, for every iteration, a subset of the original dataset is selected randomly with replacement to serve as the training dataset for the model. The data points not included in the training subset, known as the out-of-sample points, constitute the validation set. This procedure is conducted n times consecutively, and the average error rate from these n iterations is calculated to assess the model's predictive accuracy.

5 Applications of DL in CT-based PTB detection

Multiple DL algorithms have been widely applied to CT-based PTB diagnosis, including 3D CNN, MLP, U-Net, DTE-SVM, ICNN, and GNN. These algorithms have unique feature extraction and classification capability characteristics suitable for different datasets and diagnostic tasks. U-Net is often used for medical image segmentation tasks and performs well in segmenting and diagnosing PTB lesions. Ma et al. (2020) utilized U-Net to process CT data from 337 ATB cases, 110 pneumonia cases, and 120 healthy individuals. Utilizing an independent dataset for testing, they achieved exceptional results, demonstrating a positive predictive value of 0.971 and an AUC score of 0.980.

Three-dimensional CNNs can process 3D chest CT data, fully exploiting spatial information for diagnosis. Li X. K. et al. (2021) developed a 3D CNN to assist in diagnosing PTB, achieving 93.7% precision and 98.7% recall by learning from 501 PTB patients and an equal number of standard samples. Most studies employ cross-validation methods like 5-fold or 10-fold cross-validation. For example, Zhang et al. (2024) used MAResNet with 905 chest CT samples provided by Beijing Chest Hospital, adopting fivefold cross-validation, achieving an accuracy of 94% with sensitivity and specificity reaching 93.80% and 94.20%, respectively. Haq et al. (2022) used an ANN-based classifier, MLP, with 10-fold cross-validation, achieving an accuracy of 99% and a very high Kappa coefficient (0.98). Some studies use independent test data for model performance evaluation, such as in the study by Li X. K. et al. (2021). Open data sharing can accelerate the development and validation of CT-based PTB algorithms, improving research validation capabilities. The study by Huang et al. (2023) illustrates this value well; their research is based on the data released by Zhang et al. (2020). They employed a DTE-SVM algorithm, showing satisfactory results in terms of accuracy and sensitivity. In conclusion, these research findings demonstrate that as DL algorithms mature, their application in CT-based PTB diagnosis is becoming increasingly widespread and practical. Characteristics of the included studies are shown in Table 3.

5.1 Detection and classification of TB lesions

Conventional techniques for PTB detection frequently depend on radiologists' expertise, which may be subjective and protracted. The integration of DL into this process has shown promising advancements. The application of DL in TB management is shown in Figure 7. Zhang et al. (2024) introduces a 3D multi-scale attention residual network (MAResNet) to recognize PTB utilizing CT images. MAResNet is the integration of the Convolutional Block Attention Module (CBAM) alongside residual modules. This dual mechanism enhances the distinguishability of image features and allows for the efficient reuse of shallow features. The accuracy of MAResNet in classifying PTB reaches 94%, which is essential for differential diagnosis and treatment planning.

Another study by Yoon et al. (2023) a 3D neural network model, nnU-Net, will be created to investigate the clinical significance of CT cavity volume and evaluate the model's efficacy in cavity detection. The research retrospectively analyzed 392 patients with mycobacterial pulmonary disease, including TB and non-tuberculous pulmonary disease. The nnU-Net model demonstrated high sensitivity in detecting cavities, with a mean DSC of 78.9. One notable application is the cascading deep supervision U-Net model, as highlighted in the study by Hu et al. (2022), which concentrates on the diagnosis of pneumoconiosis complicated by PTB. The CSNet model leverages the strengths of HRCT to provide high-resolution imaging coupled with the DL framework to enhance the segmentation and diagnosis of affected lung tissues. This approach has shown superior performance over traditional U-Net models, with an AUC value of 0.947.

5.2 Differentiation between TB and other lung diseases

5.2.1 Distinguishing PTB and non-tuberculous mycobacteria lung disease

Wang et al. (2021) retrospectively amassed chest CT images from 301 patients with non-tuberculous mycobacterial lung disease (NTM-LD) and 804 patients with PTB. The definitive diagnostic criterion was pathogenic microbiological analysis. They utilized a 3D ResNet model, attaining AUC scores of 0.90, 0.88, and 0.86 for the training, validation, and testing datasets. Additionally, when assessed on an external dataset consisting of 40 cases of NTM-LD and 40 cases of MTB-LD, the AUC was 0.78. The 3D-ResNet model had a markedly enhanced capacity to distinguish between the two circumstances relative to radiologists with 10 years of expertise, and its diagnosis speed surpassed that of the radiologists by more than 1,000 times.

5.2.2 Distinguishing PTB and pneumonia

The advent of DL has revolutionized the field of medical imaging, particularly in the differentiation between PTB and CAP. Han et al. (2023) have used the power of 3D-CNNs to discern PTB from CAP using chest CT images. Their model was trained and validated using a dataset comprising 493 patients from two imaging centers. The model achieved an accuracy of 0.989 in the internal and 0.934 in the external test set, showcasing its robustness in differentiating the two conditions. The ability of the 3D-CNN to directly extract abstract features from images without the need for manual segmentation aligns with the growing trend in radiomics, which relies on high-throughput feature analysis. This method accelerates the diagnosis process and reduces the impact of subjective interpretation prevalent in conventional radiological evaluations.

5.2.3 Distinguishing PTB and lung cancer

Lung cancer can be categorized into three main groups. Distinguishing PTB from lung cancer is challenging due to their overlapping clinical and radiological features. Feng et al. (2020) used a CNN method to extract features from CT images, creating a DL signature to predict the likelihood of PTB or lung adenocarcinoma. They also developed a DL nomogram combining

TABLE 3 Characteristics of the included studies.

References	Journal	Country	Number of patients (male/female)	Purpose	Deep learn algorithm type	Dataset source	Dataset	Validation	Reference standard	Performances
Zhang et al., 2024	Medical & Biological Engineering & Computing	China	N/A	Diagnosis of PTB with 3D neural network	MAResNet	Beijing Chest Hospital	905 chest CT scans (500 PTB vs. 405 normal)	Fivefold cross-validation	Radiologist annotation	Accuracy: 94%; sensitivity: 93.80%; specificity: 94.20%; AUC: 0.97
Haq et al., 2022	Symmetry	Pakistan	N/A	Diagnosis of PTB	ANN based classifier MLP	Bahawal Victoria Hospital	200 chest CT scans (100 PTB vs. 100 normal)	10-fold cross-validation	Pulmonary disease specialist label	Accuracy of 99%; kappa: 0.98
Ma et al., 2020	Journal of X-Ray Science and Technology	China	518/328	Diagnosis of ATB	U-Net	Hebei University Affiliated Hospital	337 ATB, 110 pneumonia, and 120 normal cases	Independent test data containing 139 ATB, 40 pneumonia, and 100 normal cases	Sputum smear for ATB patients; CT report result for normal and pneumonia patients.	Accuracy: 0.968; sensitivity: 0.964; specificity: 0.971; positive predictive value: 0.971; negative predictive value: 0.964; AUC: 0.980
Huang et al., 2023	IEEE/ACM Transactions on Computational Biology and Bioinformatics	N/A	88/46	Diagnosis of PTB	DTE-SVM	Hospital database	288 CT images (144 PTB, 144 normal)/68 PTB and 66 normal	10-fold cross-validation	Radiologist	Accuracy: 94.62% ± 1.00; F1-score: 94.62% ± 1.00; precision: 95.30% ± 1.24; sensitivity: 93.89% ± 1.96; specificity: 95.35% ± 1.31; AUC: 0.9579
Zhang et al., 2020	Journal of Ambient Intelligence and Humanized Computing	N/A	88/46	Diagnosis of secondary PTB	ICNN	Hospital database	144 CT imaging datasets from 68 secondary PTB and 144 CT image datasets from 66 normal people	Independent test data containing 29 secondary PTB and 29 normal images	Radiologists	Accuracy: 93.95%; sensitivity: 94.19%; specificity: 93.72%
Li X. K. et al., 2021	Applied Intelligence	China	N/A	Diagnosis of PTB	3D CNN	Affiliated Hospital of Zhejiang University	501 CT imaging datasets from 223 PTB and 501 CT image datasets from normal people	Five-folder cross-validation and independent test data (containing 75 PTB and 75 normal cases)	Radiologist label	Precision = 93.7%, recall = 98.7%
Lu et al., 2022	Computer Methods and Programs in Biomedicine	China	N/A	Diagnosis of PTB	Graph neural network	Fourth Hospital of Huai'an	840 chest CT scans (420 PTB vs. 420 normal)	Fivefold cross-validation	Radiologist	Accuracy: 98.93%; sensitivity: 100%; specificity: 97.94%; precision: 97.86%; F1 score: 98.91%

PTB, pulmonary tuberculosis; HC, healthy controls; MAResNet, multi-scale attention ResNet; ANN, artificial neural network; MLP, multi-layer perceptron; ATB, active tuberculosis; DTE-SVM, deep transferred efficientNet with SVM; CNN, convolutional neural network; ICNN, improved convolutional neural network.

the DL signature with clinical factors and CT-based findings. The DL nomogram showed impressive AUCs of 0.889 in the training set, 0.879 in the internal validation set, and 0.809 in the external validation set. Tan et al. (2022) utilized a customized VGG16 model trained with transfer learning and achieved an accuracy of 90.4% in distinguishing between TB lung nodules and lung cancer. The accurate differentiation between TB and pulmonary nodules in CT images is crucial for effective diagnosis and treatment planning. Conventional techniques frequently depend on the proficiency of radiologists, which may be subjective and labor-intensive. The advent of DL has introduced a paradigm shift in this domain, offering automated and efficient solutions.

5.3 Quantitative analysis of PTB lesion progression

In addition to identifying PTB, DP is adept at monitoring changes in patients' conditions following treatment and assessing the severity of the disease.

The accurate quantification of lesion progression in PTB from CT images is pivotal for disease monitoring and treatment response evaluation. The integration of depth information in the ResNet model allowed for capturing the 3D characteristics of pulmonary lesions, providing a more comprehensive analysis than traditional 2D image assessments. This approach underscores the potential of DL in identifying the presence of disease and quantifying its extent and severity. Gao et al. (2020) developed a 3D ResNet incorporating depth information at each layer, the suggested depth-ResNet model demonstrated remarkable performance, with an average classification accuracy of 92.7% in predicting severity scores. Wu et al. (2019) utilize DL to create a diagnostic framework that detects PTB lesions and classifies them into specific types, such as military and tuberculoma. Applying a Noisy-Or Bayesian function to calculate an overall infection probability enhances the diagnostic report with quantitative analysis, offering clinicians a more thorough comprehension of the infection's scope and characteristics. The approach utilized advanced 3D CNNs to examine CT imaging datasets from 233 patients with active PTB and 501 healthy controls. The recall and precision for identifying PTB patients were 98.7% and 93.7%, respectively. The classification accuracy of PTB was 90.9%.

5.4 Reducing CT radiation dose for diagnosing PTB

Studies have shown that the effective radiation dose from a single CT scan can range from a few millisieverts (mSv) to over 8 mSv, which is significantly higher than the typical annual background radiation exposure of around 3 mSv (Sharma and Surani, 2020; McKenna and McMonagle, 2024). Using ionizing radiation in CT scans has raised concerns about the potential health risks. The mean radiation exposure for ultra-low-dose computed tomography (ULDCT) ranges from 0.05 to 0.26 mSv, representing a significant reduction when compared to the radiation levels associated with standard-dose CT scans (Heltborg et al., 2024). Yan et al. (2021a) demonstrates the application of a CycleGAN model for denoising ultralow-dose CT images in evaluating PTB. The optimized CycleGAN model improved the peak signal-to-noise ratio by 2.0 dB and the structural similarity index by 0.21, providing satisfactory image quality with lower noise levels than hybrid and model-based iterative reconstruction techniques. The optimized CycleGAN technology might enable chest ULDCT to generate diagnostically acceptable images for TB evaluation.

5.5 Further expansion of PTB diagnosis

5.5.1 Diagnosis of multidrug-resistant TB

Postprimary TB manifests in five distinct forms: infiltrative, focal, tuberculoma, miliary, and fibrocavernous. Multidrug-resistant PTB (MDR-PTB) often exhibits similar characteristics to those of drug-susceptible TB. DL techniques have demonstrated the potential to improve multidrug-resistant TB's diagnostic precision and efficacy (MDR-TB). Gao and Qian (2018) using CT lung image data from a public dataset, a patch-based DL approach was proposed to classify multidrug-resistant TB and drug-sensitive TB. The CNN allied to the SVM classifier achieved an accuracy of 91.11% with the patch-based DL technique. This study overcame the challenge of a limited dataset of only 230 samples by using patches instead of full images, effectively expanding the dataset from hundreds to thousands. Duwairi and Melhem (2023) employed multi-channel models that incorporated image frames, mask frames, and gender/age data as inputs, utilizing transfer learning based on VGG19 and ResNet neural networks for feature extraction from CT scans. Their study's best-performing model for MDR classification achieved an accuracy of 74.13% and an AUC of 64.2%. DeepTB is a DL system created using CNN-ResNet to learn transfer learning. It can quickly diagnose DR-TB and divide it into three main types: rifampicin-RTB, MDR-TB, and extensively drug-RTB. Utilizing complex network structures, DeepTB transforms input data into target predictions, achieving high performance for DR-TB diagnosis (AUC: 0.943). The model also attained an AUC of 0.880 for RR-TB, 0.928 for MDR-TB, and 0.918 for XDR-TB. Integrating class activation maps (CAMs) offers a visual explanation of the decision-making process, addressing the "black-box" issue of CNNs and boosting clinical trust in the system's outputs (Liang et al., 2024).

5.5.2 Diagnosing the infectiousness of PTB

The utilization of DL models in the analysis of CT images has demonstrated considerable potential in differentiating the infectivity of PTB patients. Gao et al. (2023) created a DL model called TBINet, which employs a 2D projection-based CNNs to assess the infectivity of PTB patients using CT images. The algorithm was trained on a dataset of 925 individuals from four sites, with infectivity classified according to several sputum samples conducted within a month. The TBiNet model exhibited enhanced performance, achieving an AUC of 0.753 on the external test set, surpassing current DL methodologies. Gradient-weighted class activation mapping (Grad-CAM) technology indicated that CT scans exhibiting increased consolidation, voids, upper lobe involvement, and larger lymph nodes were more frequently associated with patients suffering from highly infectious types of PTB.

5.5.3 Diagnosing comorbidities of PTB

The diagnosis of comorbidities in patients with PTB is crucial for effective management and treatment. Several studies have highlighted the significance of identifying comorbidities such as diabetes mellitus and HIV infection in recently diagnosed PTB individuals (Sama et al., 2023). When diagnosing PTB, it is crucial to take into account a variety of factors, including age, sex, previous TB treatment history, and other comorbidities (Jiang et al., 2024). Additional research and guidelines are required to improve the diagnosis and management of comorbidities in PTB patients.

5.6 Integration with clinical decision support systems

While the initial validation results are impressive, further prospective validation studies are necessary within actual clinical settings. Once these commercial AI systems have been thoroughly tested, they could offer physicians globally convenient, efficient, and precise diagnostic tools, thereby aiding clinical decision-making in the foreseeable future. Integrating clinical decision support systems is a key challenge in applying DL to CT-based PTB diagnosis. This requires seamlessly integrating AI models with existing healthcare systems to ensure that diagnostic results are effectively communicated to physicians and influence clinical decision-making. Specifically, it requires addressing several challenges: First, integrating the output of AI models with data from existing systems like electronic medical records and PACS to ensure diagnostic results are associated with the patient's other clinical information. This requires addressing issues around data formats and security. Second, ensuring AI-assisted diagnosis can seamlessly integrate into the physician's clinical workflow without adding extra steps or disrupting the normal diagnostic process. This requires optimizing and redesigning existing workflows. Finally, physicians must understand the rationale and logic behind the AI model's diagnoses to evaluate the results and perform secondary confirmations. This requires improving the interpretability of the AI model so that physicians can gain insights into its inner workings. Interpretability remains a critical challenge in the domain of CT-based pneumonia and PTB diagnosis, as DL models used for image analysis are often regarded as "black boxes" due to their high-dimensional and non-linear nature. Providing clear and actionable explanations can help physicians and DL models jointly improve diagnostic accuracy, reducing the risk of misdiagnosis of PTB. One effective approach involves using Grad-CAM to visualize the regions of the lungs that the model focuses on during the diagnosis of PTB, such as lesions or areas of consolidation (Jegatheeswaran et al., 2024). Additionally, Shapley additive explanations or local interpretable model-agnostic explanations can be applied to identify the most significant features contributing to the model's predictions, such as pixel intensities or specific ROIs (Chung et al., 2024; Peng et al., 2024). Interactive tools like heatmaps can further enhance interpretability by allowing physicians to explore the model's behavior and examine specific predictions in detail (Liu et al., 2020). It is also important to evaluate interpretability methods with physicians to ensure that the explanations are comprehensible and clinically relevant. Finally, ensuring compliance with regulatory standards and incorporating feedback from medical professionals will further support the safe and effective deployment of these systems in actual clinical settings.

6 Discussion

Data scarcity is a significant challenge in medical imaging, particularly for diseases like PTB, where annotated CT scans are limited. One potential solution is to employ a GAN framework to create synthetic CT images like real-world data features. Current research indicates that models trained on synthetic data can perform comparable to those trained on real data alone (Ali et al., 2024). Various geometric transformations, such as rotations, scaling, and flipping, help the model learn invariant features crucial for accurate diagnosis across different patient presentations. These transformations enhance the model's ability to recognize disease patterns despite patient positioning and imaging technique variations (Mastouri et al., 2024). Transfer learning represents another promising strategy to address data scarcity (Wajgi et al., 2024). Researchers can initialize the network weights and fine-tune the model on the PTB CT data by leveraging pre-trained models on large-scale medical imaging datasets or even non-medical image datasets. For instance, models pre-trained in general lung disease detection tasks may have learned useful low-level and mid-level features such as lung structure identification and texture analysis. These pre-trained features can be transferred and further adapted to the specific task of TB diagnosis, thereby enhancing the model's performance on limited datasets. To prevent overfitting and reduce model complexity, regularization techniques such as dropout and weight decay improve the model's ability to generalize to new data by discouraging the model from relying too heavily on any training example. Future studies could focus on developing more advanced data augmentation techniques that mimic real-world variations in imaging data. Additionally, exploring hybrid transfer learning methods that combine multiple pre-trained models could optimize generalization for PTB diagnosis. Such approaches may lead to more robust and accurate diagnostic models, even in limited training data.

The application of DL in clinical settings has revolutionized healthcare, offering promising advancements in diagnostics. However, this technological leap also presents many ethical concerns that require careful consideration and resolution.

First, using DL in healthcare often involves processing sensitive patient data, raising concerns about privacy and security. To address this, robust encryption, anonymization, and secure data-sharing protocols are proposed to protect patient data (Mirzaei et al., 2024). Additionally, federated learning techniques are being explored to train models on decentralized data, which can help preserve privacy while allowing for practical model training (Mukund et al., 2024). Second, DL models can be affected by biases in the training data, which can cause doctors to make bad decisions. To fix this problem, researchers use data augmentation, data balance, and fairness-aware training to ensure that models accurately represent diverse groups of people and do not make differences worse. This method is essential for ensuring that DL systems are trustworthy in clinical settings. Third, the intricacy of DL models, frequently called "black boxes," makes it challenging to understand and interpret how they make decisions fully. In clinical settings, the lack of transparency in the predictions made by algorithms can lead to skepticism among medical professionals regarding the model's reliability (Rajpurkar et al., 2022). Scientists are creating explainable AI methods like saliency maps and attention processes to improve explainability (Cerekci et al., 2024). These offer insights into model predictions and support the development of transparency and confidence in clinical decision-making. Finally, the algorithms fail to complete accuracy, and the accountability for any detrimental outcomes resulting from erroneous predictions remains ambiguous. This engenders uncertainty among physicians and patients. The imprecise accuracy of AI systems presents significant responsibility concerns related to harm (Morin-Martel, 2023). This requires careful evaluation of DL's diagnostic accuracy and its impact on clinical workflows, ensuring the technology is practical and ethically integrated into healthcare practices.

The accuracy of DL models in TB diagnosis is essential, as misdiagnoses can have severe consequences. Rigorous testing and validation of DL models against gold standards are required to guarantee accuracy. Consistently enhance the models to accommodate the evolving clinical landscape and emerging scientific findings. Innovative models must be evaluated in real-world medical settings and integrated smoothly into the standard operational procedures, particularly in nations with a high TB burden and limited access to sophisticated medical technology and specialized medical personnel to guide clinical practice effectively. Additionally, improving patient comprehension of the diagnostic procedure and clinician trust are benefits of developing explainable AI approaches for DL models used in TB diagnosis. These factors are critical for properly implementing these technologies in clinical settings.

Two frameworks were recommended to ensure the ethical application of DL in diagnosing PTB. The Principles of Biomedical Ethics serve as a foundational guide, emphasizing four central bioethical principles: autonomy, beneficence, non-maleficence, and justice (Mirzaei et al., 2024). These principles are crucial for evaluating the ethical implications of DL applications in the diagnosis of PTB, ensuring that they benefit patients without causing harm, respecting patient autonomy, and promoting equitable access to care. Furthermore, the Trustworthy AI Framework provides a comprehensive set of criteria for AI systems (Schwabe et al., 2024). It highlights the importance of human agency and oversight, diversity, non-discrimination, and fairness, underscoring the necessity for DL systems to be designed and deployed trustworthy to respect human rights.

7 Conclusion

Artificial intelligence-based techniques, such as DL and other traditional ML algorithms, when applied to PTB, offer an autonomous, convenient, and efficient approach to enhance diagnostic precision and speed, often surpassing the capabilities of radiologists. This study underscores the complexity involved in diagnosing PTB. It emphasizes the significant role of sophisticated DL and imaging diagnostic methods. The main goal of medical image processing is to use algorithms to get accurate and valuable information out of images with as little mistake as possible. The segmenting, classifying, and diagnosing PTB utilizing CT data generally comprises four essential stages: data acquisition and preprocessing, feature extraction, and classification. Furthermore, it is imperative to prioritize the interpretability of DL models when they are implemented in clinical decision-making processes. The research may be improved by examining the integration of multi-modal datasets and deploying real-time DL solutions in healthcare environments. Therefore, DL tools can be considered a promising diagnostic resource for PTB and various other life-threatening diseases.

Data availability statement

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

Author contributions

FZ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MML: Methodology, Software, Writing – original draft. MLL: Conceptualization, Investigation, Validation, Writing – review & editing. TT: Formal analysis, Investigation, Methodology, Writing – review & editing. GZ: Conceptualization, Methodology, Visualization, Writing – review & editing. ZY: Investigation, Methodology, Validation, Visualization, Writing – review & editing. FG: Investigation, Supervision, Validation, Writing – original draft. HH: Visualization, Writing – review & editing. YW: Software, Supervision, Validation, Writing – original draft, Writing – review & editing. JW: Formal analysis, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. YL: Supervision, Validation, Visualization, Writing – review & editing, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Key Research and Development Program of Liaoning Province (Grant Number 2019JH8/10300021).

Acknowledgments

We are grateful for Biorender's support in generating our images (www.biorender.com). Thank you for Chat GPT's contribution to proofreading this 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.

Generative AI statement

The author(s) verify and take full responsibility for the use of generative AI in the preparation of this manuscript. Generative AI was used. Fei Zhang utilized ChatGPT to proofread this manuscript.

Publisher's note

References

Abubakar, I., Drobniewski, F., Southern, J., Sitch, A. J., Jackson, C., Lipman, M., et al. (2018). Prognostic value of interferon-γ release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet Infect. Dis.* 18, 1077–1087. doi: 10.1016/S1473-3099(18)30355-4

Adane, T., Melku, M., Ayalew, G., Bewket, G., Aynalem, M., and Getawa, S. (2022). Accuracy of monocyte to lymphocyte ratio for tuberculosis diagnosis and its role in monitoring anti-tuberculosis treatment: systematic review and meta-analysis. *Medicine (Baltimore)* 101:e31539. doi: 10.1097/MD.0000000000031539

Ali, M., Ali, M., Hussain, M., and Koundal, D. (2024). Generative Adversarial Networks (GANs) for medical image processing: recent advancements. *Arch. Comput. Methods Eng.* doi: 10.1007/s11831-024-10174-8

Andama, A., Whitman, G. R., Crowder, R., Reza, T. F., Jaganath, D., Mulondo, J., et al. (2022). Accuracy of tongue swab testing using Xpert MTB-RIF ultra for tuberculosis diagnosis. *J. Clin. Microbiol.* 60:e0042122. doi: 10.1128/jcm.00421-22

Ankrah, A. O., Glaudemans, A. W. J. M., Maes, A., Van de Wiele, C., Dierckx, R. A. J. O., Vorster, M., et al. (2018). Tuberculosis. *Semin. Nucl. Med.* 48, 108–130. doi: 10.1053/j.semnuclmed.2017.10.005

Ball, L., Vercesi, V., Costantino, F., Chandrapatham, K., and Pelosi, P. (2017). Lung imaging: how to get better look inside the lung. *Ann. Transl. Med.* 5:294. doi: 10.21037/atm.2017.07.20

Borah, P., Deb, P. K., Venugopala, K. N., Al-Shar'i, N. A., Singh, V., Deka, S., et al. (2021). Tuberculosis: an update on pathophysiology, molecular mechanisms of drug resistance, newer anti-TB drugs, treatment regimens and host- directed therapies. *Curr. Top. Med. Chem.* 21, 547–570. doi: 10.2174/1568026621999201211200447

Cacciamani, G. E., Chu, T. N., Sanford, D. I., Abreu, A., Duddalwar, V., Oberai, A., et al. (2023). PRISMA AI reporting guidelines for systematic reviews and meta-analyses on AI in healthcare. *Nat. Med.* 29, 14–15. doi: 10.1038/s41591-022-02139-w

Cerekci, E., Alis, D., Denizoglu, N., Camurdan, O., Ege Seker, M., Ozer, C., et al. (2024). Quantitative evaluation of Saliency-Based Explainable artificial intelligence (XAI) methods in Deep Learning-Based mammogram analysis. *Eur. J. Radiol.* 173:111356. doi: 10.1016/j.ejrad.2024.111356

Chae, S., Street, W. N., Ramaraju, N., and Gilbertson-White, S. (2024). Prediction of cancer symptom trajectory using longitudinal electronic health record data and long short-term memory neural network. *JCO Clin. Cancer Inform.* 8:e2300039. doi: 10.1200/CCI.23.00039

Chen, X., Jia, X., Lei, H., Wen, X., Hao, Y., Ma, Y., et al. (2020). Screening and identification of serum biomarkers of osteoarticular tuberculosis based on mass spectrometry. *J. Clin. Lab. Anal.* 34:e23297. doi: 10.1002/jcla.23297

Chung, W. Y., Yoon, J., Yoon, D., Kim, S., Kim, Y., Park, J. E., et al. (2024). Development and validation of deep learning-based infectivity prediction in pulmonary tuberculosis through chest radiography: retrospective study. *J. Med. Internet Res.* 26:e58413. doi: 10.2196/58413

Crudu, V., Stratan, E., Romancenco, E., Allerheiligen, V., Hillemann, A., and Moraru, N. (2012). First evaluation of an improved assay for molecular genetic detection of tuberculosis as well as rifampin and isoniazid resistances. *J. Clin. Microbiol.* 50, 1264–1269. doi: 10.1128/JCM.05903-11

Dai, Y., Shan, W., Yang, Q., Guo, J., Zhai, R., Tang, X., et al. (2019). Biomarkers of iron metabolism facilitate clinical diagnosis in Mycobacterium tuberculosis infection. *Thorax* 74, 1161–1167. doi: 10.1136/thoraxjnl-2018-212557

Dimastromatteo, J., Charles, E. J., and Laubach, V. E. (2018). Molecular imaging of pulmonary diseases. *Respir. Res.* 19:17. doi: 10.1186/s12931-018-0716-0

Duwairi, R., and Melhem, A. (2023). A deep learning-based framework for automatic detection of drug resistance in tuberculosis patients. *Egypt. Inform. J.* 24, 139–148. doi: 10.1016/j.eij.2023.01.002

Ernst, J. D. (2012). The immunological life cycle of tuberculosis. *Nat. Rev. Immunol.* 12, 581–591. doi: 10.1038/nri3259

Feng, B., Chen, X., Chen, Y., Lu, S., Liu, K., Li, K., et al. (2020). Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. *Eur. Radiol.* 30, 6497–6507. doi: 10.1007/s00330-020-07024-z

Gao, X. W., and Qian, Y. (2018). Prediction of multidrug-resistant TB from CT pulmonary images based on deep learning techniques. *Mol. Pharm.* 15, 4326–4335. doi: 10.1021/acs.molpharmaceut.7b00875

Gao, X. W., James-Reynolds, C., and Currie, E. (2020). Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture. *Neurocomputing* 392, 233–244. doi: 10.1016/J.NEUCOM.2018.12.086

Gao, Y., Zhang, Y., Hu, C., He, P., Fu, J., Lin, F., et al. (2023). Distinguishing infectivity in patients with pulmonary tuberculosis using deep learning. *Front. Public Health* 11:1247141. doi: 10.3389/fpubh.2023.1247141

Ge, Y., Xu, L., Wang, X., Que, Y., and Piran, M. J. (2024). A novel framework for multimodal brain tumor detection with scarce labels. *IEEE J. Biomed. Health Inform.* 13, 1–14. doi: 10.1109/JBHI.2024.3467343

Georghiou, S. B., Gomathi, N. S., Rajendran, P., Nagalakshmi, V., Prabakaran, L., Prem Kumar, M. M., et al. (2021). Accuracy of the truenat MTB-RIF Dx assay for detection of rifampicin resistance-associated mutations. *Tuberc. Edinb. Scotl.* 127:102064. doi: 10.1016/j.tube.2021.102064

Ghosh, C., Leon, A., Koshy, S., Aloum, O., Al-Jabawi, Y., Ismail, N., et al. (2021). Breath-based diagnosis of infectious diseases: a review of the current landscape. *Clin. Lab. Med.* 41, 185–202. doi: 10.1016/j.cll.2021.03.002

Gordaliza, P. M., Muñoz-Barrutia, A., Via, L. E., Sharpe, S., Desco, M., and Vaquero, J. J. (2019). Computed tomography-based biomarker for longitudinal assessment of disease burden in pulmonary tuberculosis. *Mol. Imaging Biol.* 21, 19–24. doi: 10.1007/s11307-018-1215-x

Han, D., Chen, Y., Li, X., Li, W., Zhang, X., He, T., et al. (2023). Development and validation of a 3D-convolutional neural network model based on chest CT for differentiating active pulmonary tuberculosis from community-acquired pneumonia. *Radiol. Med. (Torino)* 128, 68–80. doi: 10.1007/s11547-022-01580-8

Haq, I., Mazhar, T., Nasir, Q., Razzaq, S., Mohsan, S. A. H., Alsharif, M. H., et al. (2022). Machine vision approach for diagnosing Tuberculosis (TB) based on Computerized Tomography (CT) scan images. *Symmetry* 14:1997. doi: 10.3390/sym14101997

Heltborg, A., Mogensen, C. B., Skjøt-Arkil, H., Giebner, M., Al-Masri, A., Khatry, U. B., et al. (2024). Can clinicians identify community-acquired pneumonia on ultralow-dose CT? A diagnostic accuracy study. *Scand. J. Trauma Resusc. Emerg. Med.* 32:67. doi: 10.1186/s13049-024-01242-w

Honeyborne, I., Mtafya, B., Phillips, P. P. J., Hoelscher, M., Ntinginya, E. N., Kohlenberg, A., et al. (2014). The molecular bacterial load assay replaces solid culture for measuring early bactericidal response to antituberculosis treatment. *J. Clin. Microbiol.* 52, 3064–3067. doi: 10.1128/JCM.01128-14

Hu, M., Wang, Z., Hu, X., Wang, Y., Wang, G., Ding, H., et al. (2022). High-resolution computed tomography diagnosis of pneumoconiosis complicated with pulmonary tuberculosis based on cascading deep supervision U-Net. *Comput. Methods Programs Biomed.* 226:107151. doi: 10.1016/j.cmpb.2022.107151

Huang, C., Wang, W., Zhang, X., Wang, S. H., and Zhang, Y. D. (2023). Tuberculosis diagnosis using deep transferred EfficientNet. *IEEE ACM Trans. Comput. Biol. Bioinform.* 20, 2639–2646. doi: 10.1109/tcbb.2022.3199572

Huang, Y., Ai, L., Wang, X., Sun, Z., and Wang, F. (2022). Review and updates on the diagnosis of tuberculosis. *J. Clin. Med.* 11:5826. doi: 10.3390/jcm11195826

Hussain, O. A., and Junejo, K. N. (2019). Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. *Inform. Health Soc. Care* 44, 135–151. doi: 10.1080/17538157.2018.1433676

Jegatheeswaran, T., Kumaralingam, L., Selvarajah, S., and Ratnarajah, N. (2024). "Explainable hybrid CNN swin-transformer network for tuberculosis diagnosis in chest X-rays of Sri Lankan patients," in *Proceedings of the 2024 4th International Conference on Advanced Research in Computing (ICARC)*, Belihuloya, 85–90. doi: 10.1109/ICARC61713.2024.10499756

Jiang, L., Xin, J., Liang, L., Xia, M., Li, J., Tong, J., et al. (2024). Enhanced diagnosis of pulmonary tuberculosis through nucleotide MALDI-TOF MS analysis of BALF: a retrospective clinical study. *Sci. Rep.* 14:18416. doi: 10.1038/s41598-024-66178-8

Kaifi, R. (2023). A review of recent advances in brain tumor diagnosis based on AI-based classification. *Diagn. Basel Switz.* 13:3007. doi: 10.3390/diagnostics13183007

Kang, W., Du, J., Yang, S., Yu, J., Chen, H., Liu, J., et al. (2021). The prevalence and risks of major comorbidities among inpatients with pulmonary tuberculosis in China from a gender and age perspective: a large-scale multicenter observational study. *Eur. J. Clin. Microbiol. Infect. Dis.* 40, 787–800. doi: 10.1007/s10096-020-04077-2

Kaufmann, S. H. E. (2016). EFIS lecture. Immune response to tuberculosis: how to control the most successful pathogen on earth. *Immunol. Lett.* 175, 50–57. doi: 10.1016/j.imlet.2016.05.006

Kessel, J., Göymen, E., Wolf, T., Wetzstein, N., Küpper-Tetzel, C., Behrens, P., et al. (2023). Time discrepancy for tuberculosis-negative microscopy and culture - the diagnostic gap remains: systematic analysis from a large tertiary care tuberculosis-clinic, Germany 2013-2017. *Int. J. Infect. Dis.* 134, 269–272. doi: 10.1016/j.ijid.2023.07.009

Kowalewicz-Kulbat, M., Szpakowski, P., Locht, C., Biet, F., Kaplonek, P., Krawczyk, K. T., et al. (2018). Tuberculin skin test reaction is related to memory, but not naive CD4+ T cell responses to mycobacterial stimuli in BCG-vaccinated young adults. *Vaccine* 36, 4566–4577. doi: 10.1016/j.vaccine.2018.05.068

Lalmuanawma, S., Hussain, J., and Chhakchhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos Solitons Fractals* 139:110059. doi: 10.1016/j.chaos.2020.110059

Lalvani, A., and Pareek, M. (2010). Interferon gamma release assays: principles and practice. *Enferm. Infecc. Microbiol. Clin.* 28, 245–252. doi: 10.1016/j.eimc.2009.05.012

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, X. K., Zhou, Y. K., Du, P., Lang, G. J., Xu, M., and Wu, W. (2021). A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. *Appl. Intell.* 51, 4082–4093. doi: 10.1007/s10489-020-02051-1

Li, Z., Tong, X., Liu, S., Yue, J., and Fan, H. (2021). The value of FujiLAM in the diagnosis of tuberculosis: a systematic review and meta-analysis. *Front. Public Health* 9:757133. doi: 10.3389/fpubh.2021.757133

Liang, S., Ma, J., Wang, G., Shao, J., Li, J., Deng, H., et al. (2022). The application of artificial intelligence in the diagnosis and drug resistance prediction of pulmonary tuberculosis. *Front. Med.* 9:935080. doi: 10.3389/fmed.2022.935080

Liang, S., Xu, X., Yang, Z., Du, Q., Zhou, L., Shao, J., et al. (2024). Deep learning for precise diagnosis and subtype triage of drug-resistant tuberculosis on chest computed tomography. *MedComm* 5:e487. doi: 10.1002/mco2.487

Lima, F., Santos, A. S., Oliveira, R. D., Silva, C. C. R., Gonçalves, C. C. M., Andrews, J. R., et al. (2020). Oral swab testing by Xpert§MTB/RIF Ultra for mass tuberculosis screening in prisons. *J. Clin. Tuberc. Mycobact. Dis.* 19:100148. doi: 10.1016/j.jctube.2020.100148

Ling, D. I., Nicol, M. P., Pai, M., Pienaar, S., Dendukuri, N., and Zar, H. J. (2013). Incremental value of T-SPOT.TB for diagnosis of active pulmonary tuberculosis in children in a high-burden setting: a multivariable analysis. *Thorax* 68, 860–866. doi: 10.1136/thoraxjnl-2012-203086

Liporaci, F., Carlotti, D., and Carlotti, A. (2024). A machine learning model for the early diagnosis of bloodstream infection in patients admitted to the pediatric intensive care unit. *PLoS One* 19:e0299884. doi: 10.1371/journal.pone.0299884

Liu, C., Ma, C., Duan, J., Qiu, Q., Guo, Y., Zhang, Z., et al. (2020). Using CT texture analysis to differentiate between peripheral lung cancer and pulmonary inflammatory pseudotumor. *BMC Med. Imaging* 20:75. doi: 10.1186/s12880-020-00475-2

Lu, S. Y., Wang, S. H., Zhang, X., and Zhang, Y. D. (2022). TBNet: a context-aware graph network for tuberculosis diagnosis. *Comput. Methods Programs Biomed.* 214:9. doi: 10.1016/j.cmpb.2021.106587

Ma, L., Wang, Y., Guo, L., Zhang, Y., Wang, P., Pei, X., et al. (2020). Developing and verifying automatic detection of active pulmonary tuberculosis from multi-slice spiral CT images based on deep learning. *J. Xray Sci. Technol.* 28, 939–951. doi: 10.3233/XST-200662

MacLean, E., Broger, T., Yerlikaya, S., Fernandez-Carballo, B. L., Pai, M., and Denkinger, C. M. (2019). A systematic review of biomarkers to detect active tuberculosis. *Nat. Microbiol.* 4, 748–758. doi: 10.1038/s41564-019-0380-2

Maheshwari, P., Chauhan, K., Kadam, R., Pujani, A., Kaur, M., Chitalia, M., et al. (2016). Market assessment of tuberculosis diagnostics in India in 2013. *Int. J. Tuberc. Lung Dis.* 20, 304–313. doi: 10.5588/ijtld.15.0571

Mastouri, R., Khlifa, N., Neji, H., and Hantous-Zannad, S. (2024). Performance analysis of various deep learning models based on Max-Min CNN for lung nodule classification on CT images. *Mach. Vis. Appl.* 35:83. doi: 10.1007/s00138-024-01569-5

McKenna, D. P., and McMonagle, M. P. (2024). Plain film of the abdomen remains a low sensitivity test in A&E. *Ir. J. Med. Sci.* 193, 341–343. doi: 10.1007/s11845-023-03427-1

Mesman, A. W., Calderon, R., Soto, M., Coit, J., Aliaga, J., Mendoza, M., et al. (2019). Mycobacterium tuberculosis detection from oral swabs with Xpert MTB/RIF ULTRA: a pilot study. *BMC Res. Notes* 12:349. doi: 10.1186/s13104-019-4385-y

Min, C., Luo, X., Fan, W., Fang, H., Zeng, Z., Chen, X., et al. (2024). "Deformable dual-path networks for chronic obstructive pulmonary disease staging in CT images," in *Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, 1–8. doi: 10.1109/IJCNN60899.2024.10650451

Mirzaei, T., Amini, L., and Esmaeilzadeh, P. (2024). Clinician voices on ethics of LLM integration in healthcare: a thematic analysis of ethical concerns and implications. *BMC Med. Inform. Decis. Mak.* 24:250. doi: 10.1186/s12911-024-02656-3

Moreira, J., Fochesatto, J. B., Moreira, A. L., Pereira, M., Porto, N., and Hochhegger, B. (2011). Pneumonia tuberculosa: um estudo de 59 casos confirmados microbiologicamente. *J. Bras. Pneumol.* 37, 232–237. doi: 10.1590/s1806-37132011000200014

Morena, D., Campos, C., Castillo, M., Alonso, M., Benavent, M., and Izquierdo, J. L. (2023). Impact of the COVID-19 Pandemic on the Epidemiological Situation of Pulmonary Tuberculosis-Using Natural Language Processing. *J. Pers. Med.* 13:1629. doi: 10.3390/jpm13121629

Morin-Martel, A. (2023). Machine learning in bail decisions and judges' trustworthiness. *AI Soc.* 39, 1–12. doi: 10.1007/s00146-023-01673-6

Mukund, A., Afridi, M. A., Karolak, A., Park, M. A., Permuth, J. B., and Rasool, G. (2024). Pancreatic Ductal Adenocarcinoma (PDAC): a review of recent advancements enabled by artificial intelligence. *Cancers* 16:2240. doi: 10.3390/cancers16122240

Muljadi, R., Koesbandono, and Octavius, G. S. (2024). A systematic review and meta-analysis of diagnostic test accuracy of chest ultrasound in diagnosing pediatric pulmonary tuberculosis. *Pediatr. Pulmonol.* 59, 2381–2391. doi: 10.1002/ppul.27044

Nachiappan, A. C., Rahbar, K., Shi, X., Guy, E. S., Mortani Barbosa, E. J., Shroff, G. S., et al. (2017). Pulmonary tuberculosis: role of radiology in diagnosis and management. *Radiographics* 37, 52–72. doi: 10.1148/rg.2017160032

Naidoo, J., Shelmerdine, S. C., Charcape, C. F. U., and Sodhi, A. S. (2023). Artificial intelligence in paediatric tuberculosis. *Pediatr. Radiol.* 53, 1733–1745. doi: 10.1007/s00247-023-05606-9

Nikam, C., Jagannath, M., Narayanan, M. M., Ramanabhiraman, V., Kazi, M., Shetty, A., et al. (2013). Rapid diagnosis of Mycobacterium tuberculosis with Truenat MTB: a near-care approach. *PLoS One* 8:e51121. doi: 10.1371/journal.pone.0051121

Odimayo, S., Olisah, C. C., and Mohammed, K. (2024). Structure focused neurodegeneration convolutional neural network for modelling and classification of Alzheimer's disease. *Sci. Rep.* 14:15270. doi: 10.1038/s41598-024-60611-8

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: 10.1136/bmj.n71

Parrish, N. M., and Carroll, K. C. (2011). Role of the clinical mycobacteriology laboratory in diagnosis and management of tuberculosis in low-prevalence settings. *J. Clin. Microbiol.* 49, 772–776. doi: 10.1128/JCM.02451-10

Pavithra, K. C., Kumar, P., Geetha, M., and Bhandary, S. V. (2023). Comparative analysis of pre-trained ResNet and DenseNet models for the detection of diabetic macular edema. *J. Phys. Conf. Ser.* 2571:012006. doi: 10.1088/1742-6596/2571/1/012006

Peng, A.-Z., Kong, X.-H., Liu, S.-T., Zhang, H.-F., Xie, L.-L., Ma, L.-J., et al. (2024). Explainable machine learning for early predicting treatment failure risk among patients with TB-diabetes comorbidity. *Sci. Rep.* 14:6814. doi: 10.1038/s41598-024-57446-8

Pham, Y. L., and Beauchamp, J. (2021). Breath biomarkers in diagnostic applications. *Mol. Basel Switz.* 26:5514. doi: 10.3390/molecules26185514

Phetsuksiri, B., Klayut, W., Rudeeaneksin, J., Srisungngam, S., Bunchoo, S., Toonkomdang, S., et al. (2020). The performance of an in-house loop-mediated isothermal amplification for the rapid detection of Mycobacterium tuberculosis in sputum samples in comparison with Xpert MTB/RIF, microscopy and culture. *Rev. Inst. Med. Trop. Sao Paulo* 62:e36. doi: 10.1590/S1678-9946202062036

Qi, X., Yang, Q., Cai, J., Wu, J., Gao, Y., Ruan, Q., et al. (2024). Transcriptional profiling of human peripheral blood mononuclear cells in household contacts of pulmonary tuberculosis patients provides insights into mechanisms of Mycobacterium tuberculosis control and elimination. *Emerg. Microbes Infect.* 13:2295387. doi: 10.1080/22221751.2023.2295387

Rajpurkar, P., Chen, E., Banerjee, O., and Topol, E. J. (2022). AI in health and medicine. *Nat. Med.* 28, 31–38. doi: 10.1038/s41591-021-01614-0

Rossau, R., Traore, H., De Beenhouwer, H., Mijs, W., Jannes, G., De Rijk, P., et al. (1997). Evaluation of the INNO-LiPA Rif. TB assay, a reverse hybridization assay for the simultaneous detection of Mycobacterium tuberculosis complex and its resistance to rifampin. *Antimicrob. Agents Chemother.* 41, 2093–2098. doi: 10.1128/AAC.41.10.2093

Ruiz-Tagle, C., García, P., Hernández, M., and Balcells, M. E. (2024). Evaluation of concordance of new QuantiFERON-TB Gold Plus platforms for Mycobacterium tuberculosis infection diagnosis in a prospective cohort of household contacts. *Microbiol. Spectr.* 12:e0046924. doi: 10.1128/spectrum.00469-24

Saktiawati, A. M. I., Putera, D. D., Setyawan, A., Mahendradhata, Y., and van der Werf, T. S. (2019). Diagnosis of tuberculosis through breath test: a systematic review. *EBioMedicine* 46, 202–214. doi: 10.1016/j.ebiom.2019.07.056

Sama, L. F., Sadjeu, S., Tchouangueu, T. F., Dabou, S., Kuh, G. F., Ngouateu, O. B., et al. (2023). Diabetes mellitus and HIV infection among newly diagnosed pulmonary tuberculosis patients in the north west region of Cameroon: a cross-sectional study. *Int. J. Clin. Pract.* 2023:5998727. doi: 10.1155/2023/5998727

Savage, H. R., Rickman, H. M., Burke, R. M., Odland, M. L., Savio, M., Ringwald, B., et al. (2023). Accuracy of upper respiratory tract samples to diagnose Mycobacterium tuberculosis: a systematic review and meta-analysis. *Lancet Microbe* 4, e811–e821. doi: 10.1016/S2666-5247(23)00190-8

Schwabe, D., Becker, K., Seyferth, M., Klaß, A., and Schaeffter, T. (2024). The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review. *NPJ Digit. Med.* 7, 1–30. doi: 10.1038/s41746-024-01196-4

Shahzad, M., Ali, F., Shirazi, S. H., Rasheed, A., Ahmad, A., Shah, B., et al. (2024). Blood cell image segmentation and classification: a systematic review. *PeerJ Comput. Sci.* 10:e1813. doi: 10.7717/peerj-cs.1813

Sharma, M., and Surani, S. (2020). Exploring novel technologies in lung cancer diagnosis: do we have room for improvement? *Cureus* 12:e6828. doi: 10.7759/cureus.6828

Sharma, S. K., and Mohan, A. (2004). Extrapulmonary tuberculosis. *Indian J. Med. Res.* 120, 316–353.

Simion, A.-M., Radu, Ş., and Florea, A. M. (2024). A Review of generative adversarial networks for computer vision tasks. *Electronics* 13:713. doi: 10.3390/electronics13040713

Sollai, S., Galli, L., de Martino, M., and Chiappini, E. (2014). Systematic review and meta-analysis on the utility of Interferon-gamma release assays for the diagnosis of Mycobacterium tuberculosis infection in children: a 2013 update. *BMC Infect. Dis.* 14(Suppl. 1):S6. doi: 10.1186/1471-2334-14-S1-S6

Steingart, K. R., Flores, L. L., Dendukuri, N., Schiller, I., Laal, S., Ramsay, A., et al. (2011). Commercial serological tests for the diagnosis of active pulmonary and extrapulmonary tuberculosis: an updated systematic review and meta-analysis. *PLoS Med.* 8:e1001062. doi: 10.1371/journal.pmed.1001062

Steingart, K. R., Henry, M., Ng, V., Hopewell, P. C., Ramsay, A., Cunningham, J., et al. (2006). Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect. Dis.* 6, 570–581. doi: 10.1016/S1473-3099(06)70578-3

Su, Z., Adam, A., Nasrudin, M. F., Ayob, M., and Punganan, G. (2023). Skeletal fracture detection with deep learning: a comprehensive review. *Diagn. Basel Switz.* 13:3245. doi: 10.3390/diagnostics13203245

Takasaki, J., Manabe, T., Morino, E., Muto, Y., Hashimoto, M., Iikura, M., et al. (2018). Sensitivity and specificity of QuantiFERON-TB Gold Plus compared with QuantiFERON-TB Gold In-Tube and T-SPOT.TB on active tuberculosis in Japan. *J. Infect. Chemother.* 24, 188–192. doi: 10.1016/j.jiac.2017.10.009

Tan, H., Bates, J. H. T., and Matthew Kinsey, C. (2022). Discriminating TB lung nodules from early lung cancers using deep learning. *BMC Med. Inform. Decis. Mak.* 22:161. doi: 10.1186/s12911-022-01904-8

Tb Diagnostics Market Analysis Consortium. (2014). Market assessment of tuberculosis diagnostics in Brazil in 2012. *PLoS One* 9:e104105. doi: 10.1371/journal.pone.0104105

Tb Diagnostics Market Analysis Consortium. (2015). Market assessment of tuberculosis diagnostics in South Africa, 2012-2013. *Int. J. Tuberc. Lung Dis.* 19, 216–222. doi: 10.5588/ijtld.14.0565

Thabtah, F., Spencer, R., Abdelhamid, N., Kamalov, F., Wentzel, C., Ye, Y., et al. (2022). Autism screening: an unsupervised machine learning approach. *Health Inf. Sci. Syst.* 10:26. doi: 10.1007/s13755-022-00191-x

Tseng, Y.-C., Kuo, C.-W., Peng, W.-C., and Hung, C.-C. (2024). al-BERT: a semi-supervised denoising technique for disease prediction. *BMC Med. Inform. Decis. Mak.* 24:127. doi: 10.1186/s12911-024-02528-w

Vilcapoma, P., Parra Meléndez, D., Fernández, A., Vásconez, I. N., Hillmann, N. C., Gatica, G., et al. (2024). Comparison of faster R-CNN, YOLO, and SSD for third molar angle detection in dental panoramic X-rays. *Sensors* 24:6053. doi: 10.3390/s24186053

Wajgi, R., Yenurkar, G., Nyangaresi, V. O., Wanjari, B., Verma, S., Deshmukh, A., et al. (2024). Optimized tuberculosis classification system for chest X-ray images: fusing hyperparameter tuning with transfer learning approaches. *Eng. Rep.* 6:e12906. doi: 10.1002/eng2.12906

Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Del Ojo Elias, C., Bradley, P., et al. (2015). Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect. Dis.* 15, 1193–1202. doi: 10.1016/S1473-3099(15)00062-6

Wang, L., Ding, W., Mo, Y., Shi, D., Zhang, S., Zhong, L., et al. (2021). Distinguishing nontuberculous mycobacteria from Mycobacterium tuberculosis lung disease from CT images using a deep learning framework. *Eur. J. Nucl. Med. Mol. Imaging* 48, 4293–4306. doi: 10.1007/s00259-021-05432-x

Wang, L., Zhou, H., Xu, N., Liu, Y., Jiang, X., Li, S., et al. (2023). A general approach for automatic segmentation of pneumonia, pulmonary nodule, and tuberculosis in CT images. *iScience* 26:107005. doi: 10.1016/j.isci.2023.107005

Wen, M., Zhou, Q., Tao, B., Shcherbakov, P., Xu, Y., and Zhang, X. (2023). Short-term and long-term memory self-attention network for segmentation of tumours in 3D medical images. *CAAI Trans. Intell. Technol.* 8, 1524–1537. doi: 10.1049/cit2.12179

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., et al. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* 155, 529–536. doi: 10.7326/0003-4819-155-8-201110180-00009

WHO (2022). *Global Tuberculosis Report 2022*. Available online at: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022 (Accessed October 6, 2024)

Wu, W., Li, X., Du, P., Lang, G., Xu, M., Xu, K., et al. (2019). A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. *Appl. Intell.* 51, 4082–4093. doi: 10.1007/s10489-020-02051-1

Xu, W., Wang, H.-H., and Bai, B. (2015). Emergency transcatheter arterial embolization for massive hemoptysis due to pulmonary tuberculosis and tuberculosis sequelae. *Cell Biochem. Biophys.* 71, 179–187. doi: 10.1007/s12013-014-0182-3

Yan, C., Lin, J., Li, H., Xu, J., Zhang, T., Chen, H., et al. (2021a). Cycle-consistent generative adversarial network: effect on radiation dose reduction and image quality improvement in ultralow-dose CT for evaluation of pulmonary tuberculosis. *Korean J. Radiol.* 22, 983–993. doi: 10.3348/kjr.2020.0988

Yan, C., Wang, L., Lin, J., Xu, J., Zhang, T., Qi, J., et al. (2021b). A fully automatic AI-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. *Eur Radiol.* 32, 2188–2199. doi: 10.21203/RS.3.RS-434810/V1

Yang, Y., Xu, L., Sun, L., Zhang, P., and Farid, S. S. (2022). Machine learning application in personalised lung cancer recurrence and survivability prediction. *Comput. Struct. Biotechnol. J.* 20, 1811–1820. doi: 10.1016/j.csbj.2022.03.035

Yoon, H. J., and Kim, J.-H. (2020). Lesion-based convolutional neural network in diagnosis of early gastric cancer. *Clin. Endosc.* 53, 127–131. doi: 10.5946/ce.2020.046

Yoon, I., Hong, J. H., Witanto, J. N., Yim, J.-J., Kwak, N., Goo, J. M., et al. (2023). Mycobacterial cavity on chest computed tomography: clinical implications and deep learning-based automatic detection with quantification. *Quant. Imaging Med. Surg.* 13, 747–762. doi: 10.21037/qims-22-620

Yuan, H., Zhu, J., Wang, Q., Cheng, M., and Cai, Z. (2022). An improved DeepLab v3+ deep learning network applied to the segmentation of grape leaf black rot spots. *Front. Plant Sci.* 13:795410. doi: 10.3389/fpls.2022.795410

Zhan, Y., Wang, Y., Zhang, W., Ying, B., and Wang, C. (2022). Diagnostic accuracy of the artificial intelligence methods in medical imaging for pulmonary tuberculosis: a systematic review and meta-analysis. *J. Clin. Med.* 12:303. doi: 10.3390/jcm12010303

Zhang, F. (2024a). *Pathogenesis and typical symptoms of TB*. Available at: https://BioRender.com/u45e559

Zhang, F. (2024b). *Distribution of MTB in the human body*. Available at: https://BioRender.com/v45p623

Zhang, F. (2024c). *Schematic representation of a typical CNN architecture*. Available at: https://BioRender.com/x63i847

Zhang, F. (2024d). *A workflow for PTB diagnosis using DL based on a CT pipeline*. Available at: https://BioRender.com/h97d178

Zhang, F. (2024e). *The application of DL in TB management*. Available at: https://BioRender.com/r70x243

Zhang, S., He, C., Wan, Z., Shi, N., Wang, B., Liu, X., et al. (2024). Diagnosis of pulmonary tuberculosis with 3D neural network based on multi-scale attention mechanism. *Med. Biol. Eng. Comput.* 62, 1589–1600. doi: 10.1007/s11517-024-03022-1

Zhang, Y. D., Nayak, D. R., Zhang, X., and Wang, S. (2020). Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization. *J. Ambient Intell. Humaniz. Comput.* doi: 10.1007/s12652-020-02612

Zhao, Y.-L., Pang, Y., Xia, H., Du, X., Chin, D., Huan, S.-T., et al. (2016). Market assessment of tuberculosis diagnostics in China in 2012. *Int. J. Tuberc. Lung Dis.* 20, 295–303. doi: 10.5588/ijtld.15.0156

Zhu, L., Xu, S., Guo, H., Lu, S., Gao, J., Hu, N., et al. (2024). Machine learning-based phenogroups and prediction model in patients with functional gastrointestinal disorders to reveal distinct disease subsets associated with gas production. *J. Transl. Intern. Med.* 12, 355–366. doi: 10.2478/jtim-2024-0009

APPENDIX

APPENDIX TABLE A1 Evaluation of bias risk and applicability in selected studies.

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Zhang et al., 2024	Unclear risk	Low risk	Low risk	Unclear risk	Unclear risk	Low concern	Low concern
Haq et al., 2022	Unclear risk	Low risk	Low risk	Unclear risk	Unclear risk	Low concern	Low concern
Ma et al., 2020	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low concern	Low concern
Huang et al., 2023	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low concern	Low concern
Zhang et al., 2020	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low concern	Low concern
Li X. K. et al., 2021	Unclear risk	Low risk	Low risk	Unclear risk	Unclear risk	Low concern	Low concern
Lu et al., 2022	Unclear risk	Low risk	Low risk	Unclear risk	Unclear risk	Low concern	Low concern

Frontiers in Microbiology

TYPE Original Research
PUBLISHED 23 January 2025
DOI 10.3389/fmicb.2024.1508030

OPEN ACCESS

EDITED BY
Yu-Dong Yao,
Stevens Institute of Technology, United States

REVIEWED BY
Lu Liu,
Temple University, United States
Johid Malik,
University of Nebraska Medical Center,
United States

*CORRESPONDENCE
Jingping Zhang
✉ zjp809302@163.com
Xin Zhang
✉ zhangxin800705@163.com

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

‡These authors have contributed equally to this work

RECEIVED 08 October 2024
ACCEPTED 30 December 2024
PUBLISHED 23 January 2025

CITATION
Zhang X, Zhang X, Zhang D, Xu J, Zhang J and Zhang X (2025) The clinical prediction model to distinguish between colonization and infection by *Klebsiella pneumoniae*. *Front. Microbiol.* 15:1508030.
doi: 10.3389/fmicb.2024.1508030

The clinical prediction model to distinguish between colonization and infection by *Klebsiella pneumoniae*

Xiaoyu Zhang[1†], Xifan Zhang[1†], Deng Zhang[2], Jing Xu[1],
Jingping Zhang[1*‡] and Xin Zhang[1*‡]

[1]First Department of Infectious Diseases, The First Affiliated Hospital of China Medical University, Shenyang, China, [2]Department of Infectious Diseases, The First Affiliated Hospital of Xiamen University, Xiamen, China

Objective: To develop a machine learning-based prediction model to assist clinicians in accurately determining whether the detection of *Klebsiella pneumoniae* (KP) in sputum samples indicates an infection, facilitating timely diagnosis and treatment.

Research methods: A retrospective analysis was conducted on 8,318 patients with KP cultures admitted to a tertiary hospital in Northeast China from January 2019 to December 2023. After excluding duplicates, other specimen types, cases with substandard specimen quality, and mixed infections, 286 cases with sputum cultures yielding only KP were included, comprising 67 cases in the colonization group and 219 cases in the infection group. Antimicrobial susceptibility testing was performed on the included strains, and through univariate logistic regression analysis, 15 key influencing factors were identified, including: age > 62 years, ESBL, CRKP, number of positive sputum cultures for KP, history of tracheostomy, use of mechanical ventilation for >96 h, indwelling gastric tube, history of craniotomy, recent local glucocorticoid application, altered consciousness, bedridden state, diagnosed with respiratory infectious disease upon admission, electrolyte disorder, hypoalbuminemia, and admission to ICU (all $p < 0.05$). These factors were used to construct the model, which was evaluated using accuracy, precision, recall, F1 score, AUC value, and Brier score.

Results: Antimicrobial susceptibility testing indicated that the resistance rates for penicillins, cephalosporins, carbapenems, and quinolones were significantly higher in the infection group compared to the colonization group (all $p < 0.05$). Six predictive models were constructed using 15 key influencing factors, including Classification and Regression Trees (CART), C5.0, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), Random Forest (RF), and Nomogram. The Random Forest model performed best among all indicators (accuracy 0.93, precision 0.98, Brier Score 0.06, recall 0.72, F1 Score 0.83, AUC 0.99). The importance of each factor was demonstrated using mean decrease in Gini. "Admitted with a diagnosis of respiratory infectious disease" (8.39) was identified as the most important factor in the model, followed by "Hypoalbuminemia" (7.83), then "ESBL" (7.06), "Electrolyte Imbalance" (5.81), "Age > 62 years" (5.24), "The number of Positive Sputum Cultures for KP > 2" (4.77), and being bedridden (4.24). Additionally, invasive procedures (such as history of tracheostomy, use of ventilators for >96 h, and craniotomy) were also significant predictive factors. The Nomogram indicated that CRKP, presence of a nasogastric tube, admission to the ICU, and history of tracheostomy were important factors in determining KP colonization.

Conclusion: The Random Forest model effectively distinguishes between infection and colonization status of KP, while the Nomogram visually presents the predictive value of various factors, providing clinicians with a reference for formulating treatment plans. In the future, the accuracy of infection diagnosis can be further enhanced through artificial intelligence technology to optimize treatment strategies, thereby improving patient prognosis and reducing healthcare burdens.

KEYWORDS

***Klebsiella pneumoniae*, colonization, infection, sputum culture, risk factors, machine learning, clinical prediction model**

1 Introduction

Klebsiella pneumoniae (commonly referred to as KP) is a frequently encountered opportunistic pathogen in clinical settings, typically colonizing the human gut and upper respiratory tract (Podschun and Ullmann, 1998). KP colonization can be detected in more than 40% of the population (Guilhen et al., 2019). At the same time, this bacterium is the third most common bacterial cause of hospital-acquired pneumonia (Jones, 2010). KP infections are also common in ventilator-associated pneumonia (VAP) (Alnimr, 2023). With the emergence of multi-drug resistant (MDR) strains, it has been classified by the World Health Organization (WHO) as a priority pathogen for which urgent new therapies are needed (Antimicrobial resistance: global report on surveillance, n.d.). Sputum culture is an important method for identifying respiratory tract infection pathogens in clinical practice; however, for the opportunistic pathogen KP, clinicians often find it challenging to distinguish between colonization and infection. Colonization refers to the adherence and growth of microorganisms on the surface of host tissues, while infection describes the process by which microorganisms invade host cells or tissues, leading to pathological conditions. Colonization and infection are closely related (Vornhagen et al., 2024). When the host's immune response is suppressed or there are other risk factors present, colonizing *Klebsiella pneumoniae* may begin to proliferate and invade surrounding tissues, leading to severe infection (Cai et al., 2024). Misclassifying colonization as infection can lead to overtreatment in clinical settings, increasing the incidence of adverse reactions and promoting the development of antibiotic resistance. This includes mechanisms such as enzymatic antibiotic inactivation and modification (e.g., ESBLs, AmpC, and NDM), spread of resistance genes, absence of outer membrane porin expression, and overexpression of active efflux pump systems, among others (Li et al., 2023) (as shown in Figure 1, which illustrates the main resistance mechanisms of KP). Conversely, misclassifying infection as colonization can delay clinical treatment and increase patient mortality rates. Therefore, there is a need for a tool to assist clinicians in better differentiating between KP colonization and infection.

Previous studies have typically used logistic regression analysis to identify risk factors associated with infection and colonization (Wang X. et al., 2024). However, logistic regression struggles to capture complex nonlinear relationships between features and lacks intuitive clarity. In recent years, artificial intelligence and machine learning have shown great potential in the diagnosis and treatment of bacterial infections. They can effectively handle complex data, integrate various patient characteristics to construct predictive models, and help doctors understand more intuitively how different clinical features influence the occurrence of diseases, providing more reliable support for clinical decision-making (Deo, 2015; Jiang et al., 2022; Seyer Cagatan et al., 2022). Therefore, the aim of this study is to develop a predictive model that can accurately identify colonization versus infection of KP in sputum cultures. This model is intended to provide clinicians with scientific evidence to optimize treatment strategies and reduce the risk of misdiagnosis and inappropriate use of antimicrobial agents.

2 Methods

2.1 Setting and participants

A retrospective analysis was conducted on 8,318 patients with microbiological cultures identified as KP admitted to a tertiary hospital in Northeast China from January 2019 to December 2023. After excluding duplicate cases, other types of specimens, cases with substandard specimen quality, and mixed infections, a total of 286 cases were included, which consisted of patients with sputum cultures yielding only *Klebsiella pneumoniae*. Among these, 67 cases were classified as the colonization group and 219 cases as the infection group (Figure 2). This study has been reviewed and approved by the Ethics Committee of the First Affiliated Hospital of China Medical University [Ethics ID(2023)2023-142-2]. Infectious disease specialists confirmed the diagnosis of KP infection or colonization (diagnostic criteria are shown in Table 1). (In Figure 3 we show the CT lung images and sputum cultures of a relatively typical case of pneumonia due to KP).

2.2 Microbiological methods

Clinical data for the included cases were collected through the hospital's electronic medical record system. Using a case–control study design, a retrospective analysis of the clinical data from all patients was conducted. Antimicrobial susceptibility testing was performed on the KP (KP) strains obtained from the included cases. Identification of all pathogens and antimicrobial susceptibility testing were carried out using the VITEK 2 automated bacterial identification and VITEK 2 Compact antimicrobial susceptibility testing system from bioMérieux, France. The results of the antimicrobial susceptibility tests were categorized as susceptible, intermediate, and resistant according to the standards set by the Clinical and Laboratory Standards Institute (CLSI) (CLSI, 2023; Vornhagen et al., 2024). Extended spectrum beta-lactamase (ESBL) strains were detected through double disk synergy tests and phenotypic confirmation tests (antimicrobial susceptibility

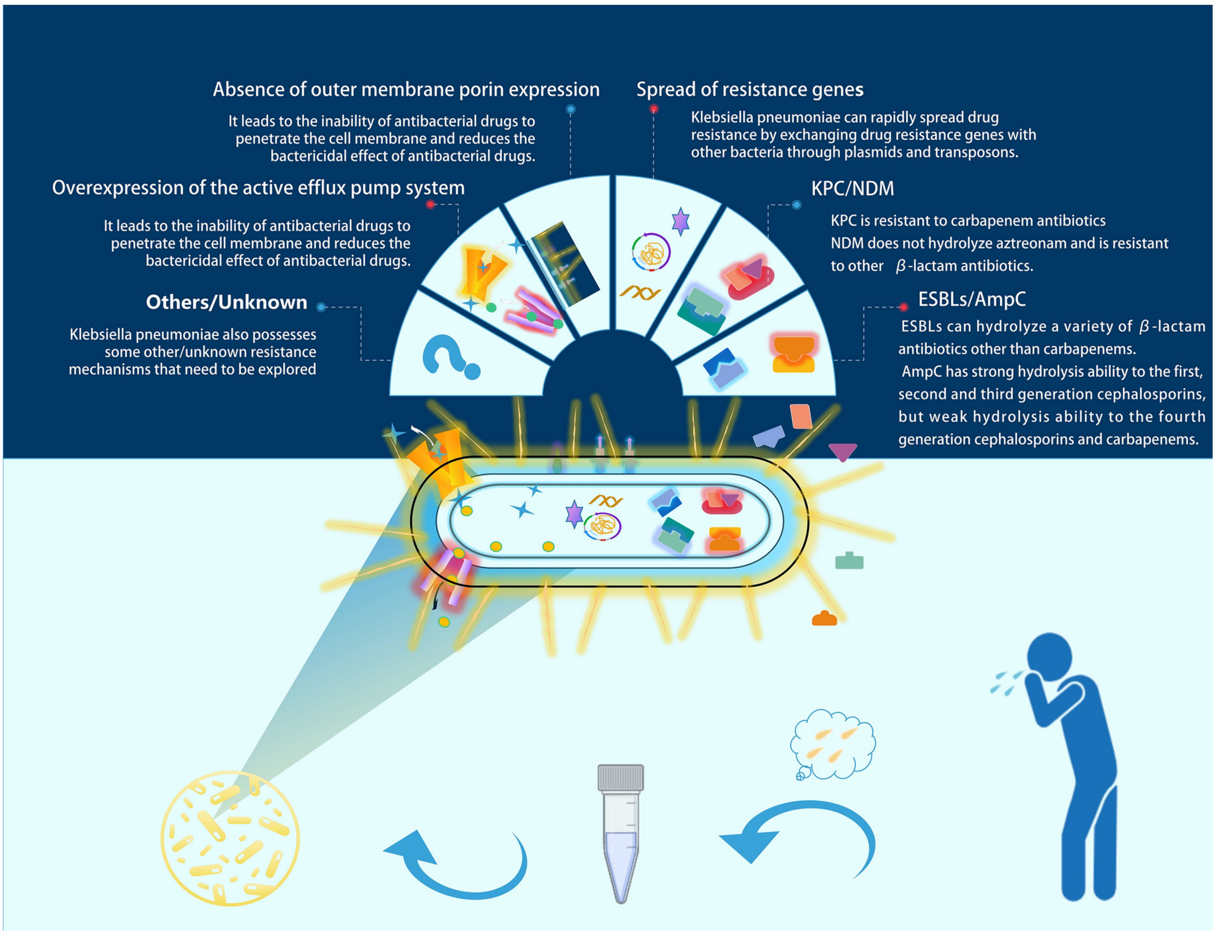


FIGURE 1
Major resistance mechanisms of KP. In clinical practice, sputum culture is an important tool for diagnosing respiratory tract infections. However, misinterpreting colonizing *Klebsiella pneumoniae* as an infectious pathogen may lead to unnecessary antimicrobial treatment. In such cases, this could contribute to the development of resistance in *Klebsiella pneumoniae*. We have illustrated several common resistance mechanisms of *Klebsiella pneumoniae* in the figure.

test disks were purchased from Oxoid, United Kingdom). In this study, carbapenem-resistant bacteria were defined as those resistant to either imipenem or meropenem.

2.3 Statistical method

Statistical analysis was conducted using SPSS 27.0 software. Categorical data were expressed as counts and percentages. The Chi-square test was used to compare the differences in resistance rates between the KP colonization group and the infection group ($p < 0.05$ was considered statistically significant). Univariate binary logistic regression analysis was performed to identify factors with statistically significant differences between the two groups ($p < 0.05$ was considered statistically significant), and these factors were included as predictive variables in the model. Predictive models to distinguish between KP colonization and infection were constructed using R 4.4.1 software, including Classification and Regression Trees (CART), C5.0, Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Random Forest (RF), and Nomogram models. The function set.seed(42) was used to ensure the reproducibility of the random process, allowing the same data to yield identical training and test set divisions. The trainControl function was employed to set the model training control parameters for 10-fold cross-validation. The performance of each model was evaluated using accuracy, precision, recall, F1 Score, AUC value, and Brier score.

3 Results

3.1 Antimicrobial susceptibility testing of KP

Antimicrobial susceptibility testing was performed on strains isolated from sputum cultures of the final 286 enrolled cases (Figure 4). Compared with the colonization group, the infection group exhibited generally higher resistance rates to various antibiotics, with all differences being statistically significant ($p < 0.05$) (see Supplementary Table S1 for detailed results). A total of 141 CRKP strains were isolated, including 17 strains from the colonization group and 124 strains from the infection group.

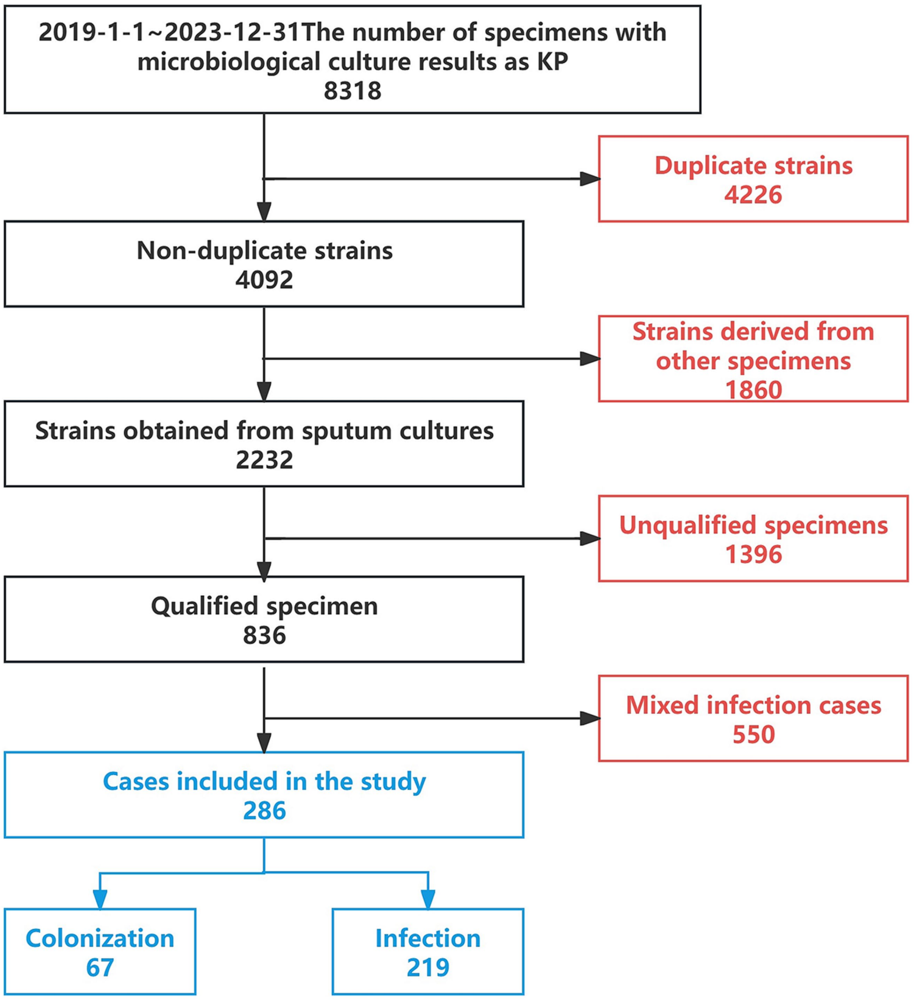


FIGURE 2
Case selection process.

TABLE 1 Inclusion, exclusion, and diagnostic criteria.

Inclusion criteria	Diagnostic criteria
• Sputum culture positive for KP • Sample quality meets standards	The diagnosis of KP infection or colonization was confirmed by infectious disease specialists based on the “Diagnostic Criteria for Hospital-Acquired Pneumonia and Ventilator-Associated Pneumonia in Chinese Adults” (2018 Edition) (Subspecialty Group of Infectious Diseases, Respiratory Society, Chinese Medical Association, 2018) and the “Guidelines for the Diagnosis and Treatment of Adult Community-Acquired Pneumonia” (Practical Edition, 2018) (Chinese Medical Association, Chinese Medical Association Journal, Chinese Medical Association General Medicine Society et al., 2019).
Duplicate strains	
• Non-sputum samples • Samples of unsatisfactory quality • Cases with mixed viral, other bacterial, or fungal infections	

3.2 Clinical characteristics of the participants

The clinical data of the included cases were collected (Table 2), and univariate logistic regression analysis revealed the following important factors influencing colonization and infection: age > 62 years, ESBL positivity, CRKP positivity, number of positive sputum cultures for KP, history of tracheostomy, use of ventilator for >96 h, presence of indwelling gastric tube, craniotomy, recent history of topical glucocorticoid use, altered consciousness, bedridden status, diagnosis of respiratory infectious disease upon admission, electrolyte disturbances, low serum protein levels, and admission to ICU (all $p < 0.05$).

3.3 Construction and evaluation of predictive models

The 15 important influencing factors obtained from univariate logistic regression analysis were used to construct models including Classification and Regression Trees (CART), C5.0, Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Random Forest

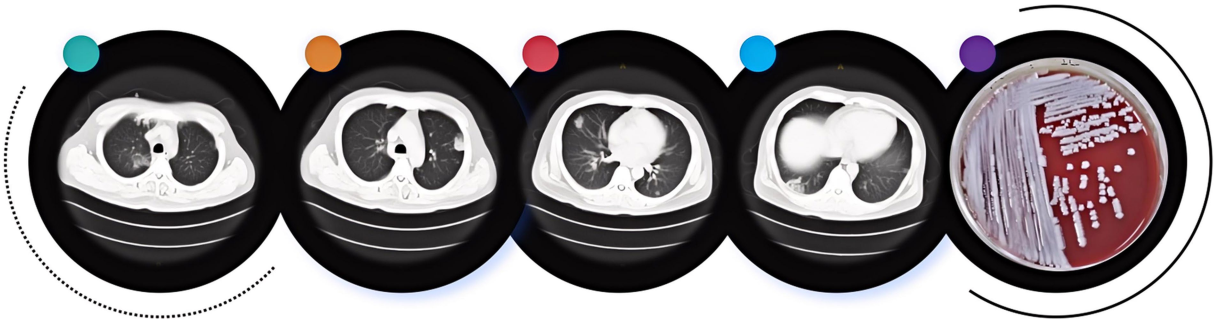


FIGURE 3
CT imaging and sputum culture of a typical case of pneumonia due to KP.

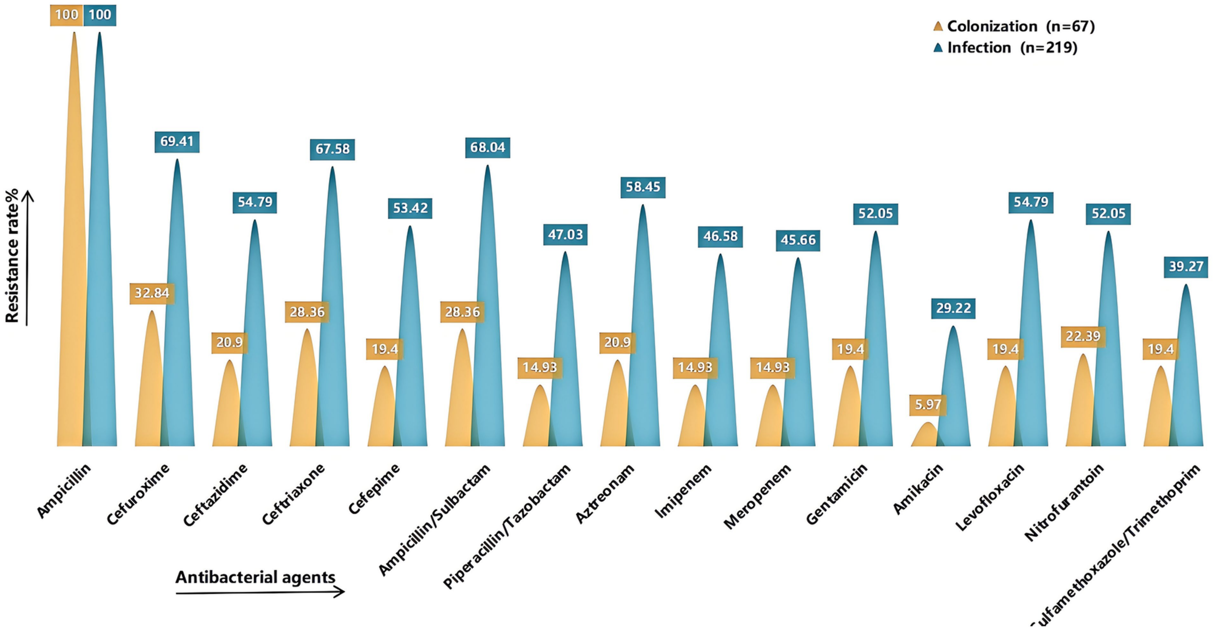


FIGURE 4
Comparison of drug resistance rates. It shows the rates of resistance to common antimicrobial agents in the colonized and infected groups. The X-axis represents the antimicrobial agents (Ampicillin, Cefuroxime, Ceftazidime, Ceftriaxone, Cefepime, Ampicillin/Sulbactam, Piperacillin/Tazobactam, Aztreonam, Imipenem, Meropenem, Gentamicin, Amikacin, Levofloxacin, Nitrofurantoin, Sulfamethoxazole/Trimethoprim), and the Y-axis represents the resistance rate. Yellow indicates the colonized group and green indicates the infected group. Detailed information can be found in Supplementary Table S1.

(RF), and Nomogram. These models were designed to differentiate between colonization and infection status of KP. The performance of the models was evaluated using several metrics, including accuracy, precision, recall, F1 Score, Area Under the Curve (AUC), and Brier Score.

In this study, the predictive models were evaluated, and the Random Forest model demonstrated the best performance across all metrics: accuracy of 0.93, precision of 0.98, Brier Score of 0.06, recall of 0.72, and F1 Score of 0.83, indicating excellent predictive performance. The C5 model also performed well, with an accuracy of 0.87, precision of 0.92, and an F1 Score of 0.65, suggesting its reliability in predicting KP infections. The CART model achieved an accuracy of 0.85, precision of 0.76, and an F1 Score of 0.61, showing performance similar to that of the C5 model. Although both C5 and CART effectively distinguished between colonization and infection, their recall rates were relatively low, and their Brier Scores were higher, indicating slightly insufficient predictive performance. The SVM model had an accuracy of 0.84 and effectively identified infection samples, reducing the risk of false negatives; however, its overall performance was still inferior to that of the C5 and CART models. The GBM model performed relatively weakly across all metrics, particularly in recall (0.35) and F1 Score (0.46), suggesting a higher probability of missed detections. The AUC (Area Under the Curve) is an important metric for measuring a model's discriminative ability. In this study, the Random Forest model achieved an AUC of 0.99, while the Nomogram model also exhibited a high AUC of 0.85. In contrast,

TABLE 2 Clinical characteristics of cases with KP colonization and infection and univariate logistic regression analysis.

Clinical characteristics/influencing factors			Colonization [*n* = 67, Strains(%)]	Infection [*n* = 219, Strains(%)]	Univariate logistic regression analysis	
					p	95%CI
Basic information	Sex	Male	53(79.10)	163(74.43)	0.437	0.396–1.491
		Female	14(20.90)	56(25.57)		
	Age[a]	>62	26(38.8)	124(56.62)	0.011	1.176–3.601
	ESBL		19(28.36)	149(68.04)	<0.01	2.944–9.822
	CRKP		17(25.37)	124(56.62)	<0.01	2.082–7.078
	The number of Positive Sputum Cultures for KP > 2[b]		5(7.46)	81(36.99)	<0.01	2.810–18.849
Underlying disease	Smoking for ≥ 10 Years		15(22.39)	46(21.00)	0.809	0.476–1.784
	Hypertension		17(25.37)	64(29.22)	0.541	0.652–2.263
	Diabetes		15(22.39)	61(27.85)	0.782	0.702–2.553
	Coronary heart disease		9(13.43)	20(9.13)	0.310	0.280–1.499
	Cerebrovascular disease		11(16.42)	62(28.31)	0.054	0.988–4.090
	Gastrointestinal bleeding		2(2.99)	10(4.57)	0.575	0.332–7.279
Invasive procedure	Tracheal intubation		31(46.27)	110(50.23)	0.571	0.677–2.028
	Tracheostomy		11(16.42)	65(29.68)	0.034	1.058–4.364
	Ventilator		32(47.76)	132(60.27)	0.071	0.957–2.878
	Ventilator use > 96 h		11(16.42)	85(38.81)	0.01	1.602–6.511
	Bronchoscope		7(10.45)	27(12.33)	0.678	0.500–2.907
	Peripherally inserted central catheter		20(29.85)	73(33.33)	0.595	0.649–2.128
	Gastric tube indwelling		31(46.27)	137(62.56)	0.019	1.116–3.372
	Recent surgery		25(37.31)	70(31.96)	0.416	0.446–1.397
	Craniotomy		4(5.97)	36(16.44)	0.039	1.061–9.050
Medication history in the last 2 weeks	History of proton pump inhibitor (PPI) use		38(56.72)	148(67.58)	0.104	0.909–2.785
	History of systemic steroid use		12(17.91)	73(33.33)	0.18	1.155–4.545
	Recent history of topical glucocorticoid use[c]		19(28.36)	98(44.75)	0.018	1.129–3.707
Others	Dysbiosis		12(17.91)	50(22.83)	0.394	0.674–2.730
	Disturbance of consciousness		10(14.93)	78(35.62)	0.002	1.525–6.521
	Bedridden status		43(64.18)	190(86.76)	<0.01	1.940–6.894
	Admitted with a diagnosis of respiratory infectious disease		13(19.40)	97(44.29)	<0.01	1.704–6.40
	Moderate to severe anemia		13(19.40)	51(23.29)	0.505	0.638–2.493
	Electrolyte imbalance		27(40.30)	144(65.75)	<0.01	1.621–4.991
	Hypoalbuminemia		22(32.84)	149(68.04)	<0.01	2.429–7.805
	Liver dysfunction		24(35.82)	107(48.86)	0.062	0.973–3.013
	Kidney dysfunction		8(11.94)	37(16.89)	0.332	0.661–3.400
	Heart Failure		4(5.97)	20(9.13)	0.418	0.522–4.804
	Admitted to the ICU		32(47.76)	145(66.21)	0.007	1.23–3.734
Prognosis	Die		7(10.45)	40(18.26)	0.136	0.815–4.502

[a] The average age of the subjects in this study is 62 years, which is used as the cutoff value.

[b] The number of KP sputum cultures: the average number of KP sputum cultures in all cases was 2, which was the critical value.

[c] Recent history of topical glucocorticoid use was defined as nebulized glucocorticoid inhalation.

the AUC values for the CART, GBM, and C5 models ranged from 0.75 to 0.80, indicating moderate discrimination ability. In summary, the Random Forest model performed best in distinguishing between colonization and infection of KP, while the Nomogram model also proved to be a reliable choice, suitable for the accurate identification of colonization and infection states in clinical applications (see Figure 5 and Table 3).

3.4 Evaluation of feature importance in the Random Forest model

To better understand the relationship between the model and the data, we visually analyzed the best-performing Random Forest model using Mean Decrease Gini. The X-axis represents the "Mean Decrease Accuracy" values for each feature; higher values indicate a more significant decline in model performance when that feature is removed. The influencing factors are ranked according to their importance, with features at the top contributing the most to the model. The most important factor in the model is "Admitted with a diagnosis of respiratory infectious disease" (8.39). The second most important feature is "Hypoalbuminemia" (7.83), followed by "ESBL" (7.06), "Electrolyte Imbalance" (5.81), "Age" (5.24), and "The number of Positive Sputum Cultures for KP > 2" (4.77). In contrast, factors such as "Craniotomy" (1.97), "Nasogastric Tube Placement" (2.62), and "ICU Admission" (2.99) contributed relatively little to determining whether the sputum culture for KP indicated infection (see Figure 6).

3.5 Nomogram model

Due to the good performance of the Nomogram, which is more intuitive than the Random Forest model, we also presented the drawn Nomogram (Figure 7). It clearly shows that "Admitted with a diagnosis of respiratory infectious disease" is the most important factor, followed by "bedridden status," "the number of positive sputum cultures for KP > 2," ESBL, and craniotomy. Additionally, CRKP, nasogastric tube placement, ICU admission, and tracheostomy are all important factors for determining KP colonization.

4 Discussion

In this study, the resistance rates of various antimicrobial drugs in the infection group were generally higher than those in

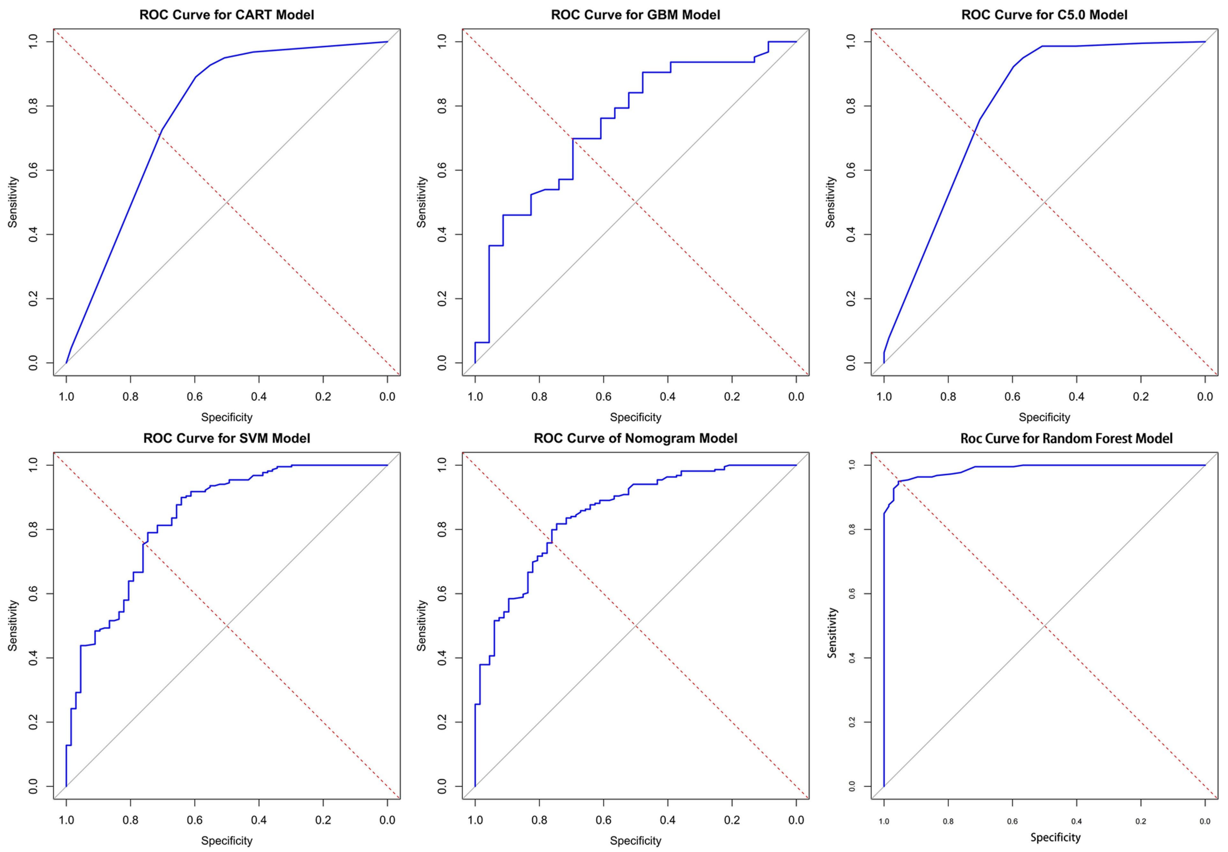


FIGURE 5
ROC curves for the six models. Here are the ROC (receiver operating characteristic) curves for six different models, with the horizontal axis of each plot representing specificity, the vertical axis representing sensitivity, and the diagonal as the baseline for random guesses. AUC is an important indicator of ROC curve, which is used to evaluate the classification performance of the model. The values of AUC range from 0 to 1 and are explained as follows: AUC = 1: perfect. $0.7 \leq \text{AUC} < 0.8$: good. $0.6 \leq \text{AUC} < 0.7$: general. AUC < 0.6 was considered poor. In the figure, the closer the curve is to the upper left corner, the higher the AUC value, indicating better model performance (the specific AUC values can be found in Table 3).

TABLE 3 Evaluation measures of the model.

	CART	GBM	C5	SVM	Nomogram	Random forest
Accuracy	0.85	0.78	0.87	0.84	0.84	0.93
Precision	0.76	0.67	0.92	0.73	0.72	0.98
Recall	0.51	0.35	0.51	0.49	0.51	0.72
F1 score	0.61	0.46	0.65	0.59	0.60	0.83
AUC	0.77	0.75	0.80	0.84	0.85	0.99
Brier score	0.12	0.17	0.11	6.56	0.12	0.06

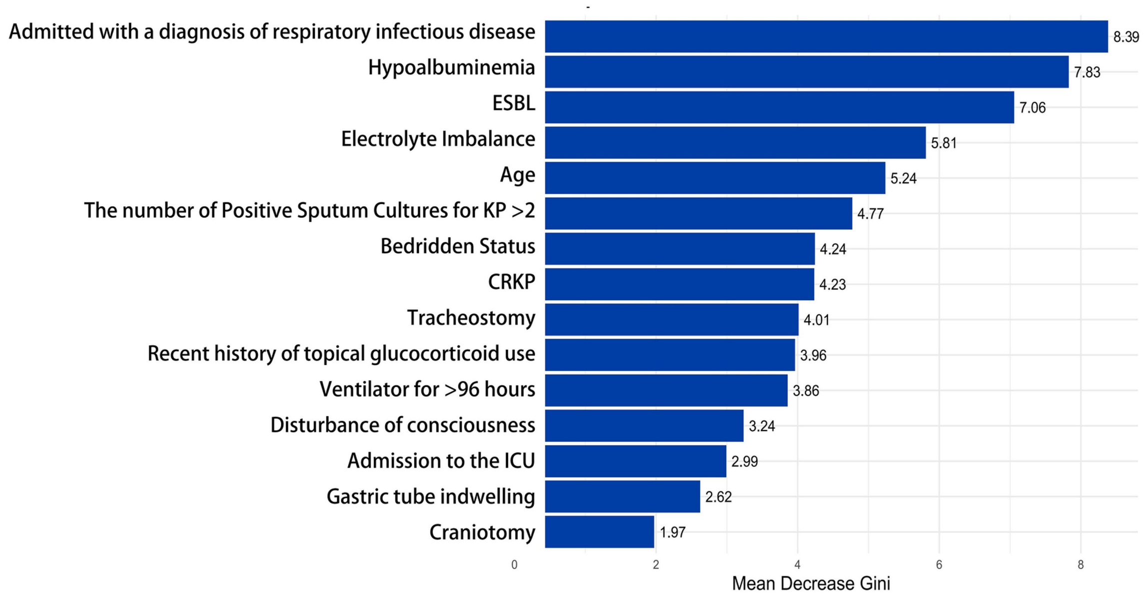


FIGURE 6
Evaluation of feature importance in the random forest model. This bar chart shows the importance of different clinical factors in the random forest model. The X-axis represents the "Mean Decrease Gini," while the Y-axis lists the clinical factors, arranged in order of their importance to the outcome from high to low. ≥ 8.0, High; 5.0–7.9, Medium-High; 3.0–4.9, Medium; 1.0–2.9, Low; 0.0–0.9, Very Low or None.

the colonization group, with statistically significant differences ($p < 0.05$). The proportion of ESBL and CRKP in the infection group was significantly higher than that in the colonization group, indicating that the infection status is highly likely associated with the extensive use of antimicrobial drugs. Previous studies have shown that the use of broad-spectrum antimicrobial agents and carbapenems are independent risk factors for the occurrence of ESBL and CRKP (Lou et al., 2022; Zhu et al., 2023; Lopera et al., 2024). High resistance rates will place significant pressure on the selection of antimicrobial agents in clinical practice. Infection status not only increases the demand for antimicrobial drugs among patients but may also accelerate the proliferation and transmission of resistant strains. Previous studies have indicated that colonizing strains of KP (KP) have a high degree of homology with infecting isolates, with 50% of KP infections arising from the patient's own microbiota (Gorrie et al., 2017). However, this study found that there was a significant difference in the resistance rates of the strains isolated from the colonization group and the infection group. From the host's perspective (Gonzalez-Ferrer et al., 2021), during an infection, the patient's immune system may be compromised, and the transmission of resistant strains through cross-infection in healthcare environments, such as hospitals (Córdova-Espinoza et al., 2023), may result in the strains in the infection group being more resistant. In contrast, the strains in the colonization group may not experience a significant increase in resistance due to the pressure from antimicrobial agents, as they exist in a relatively stable colonization state. From a microbiological perspective (Gomez-Simmonds and Uhlemann, 2017), Colonized strains may convert to infection by acquiring additional resistance genes through mechanisms such as horizontal gene transfer, gene mutation, and so on. In summary, the resistance rates in the infection group are significantly higher than those in the colonization group, indicating that we should enhance the monitoring and research of resistant strains in clinical practice. Particularly, preventing and controlling the spread of resistant bacteria will be an important focus for future research and practice.

In this study, we constructed Classification and Regression Trees (CART), C5.0, Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Random Forest (RF), and Nomogram

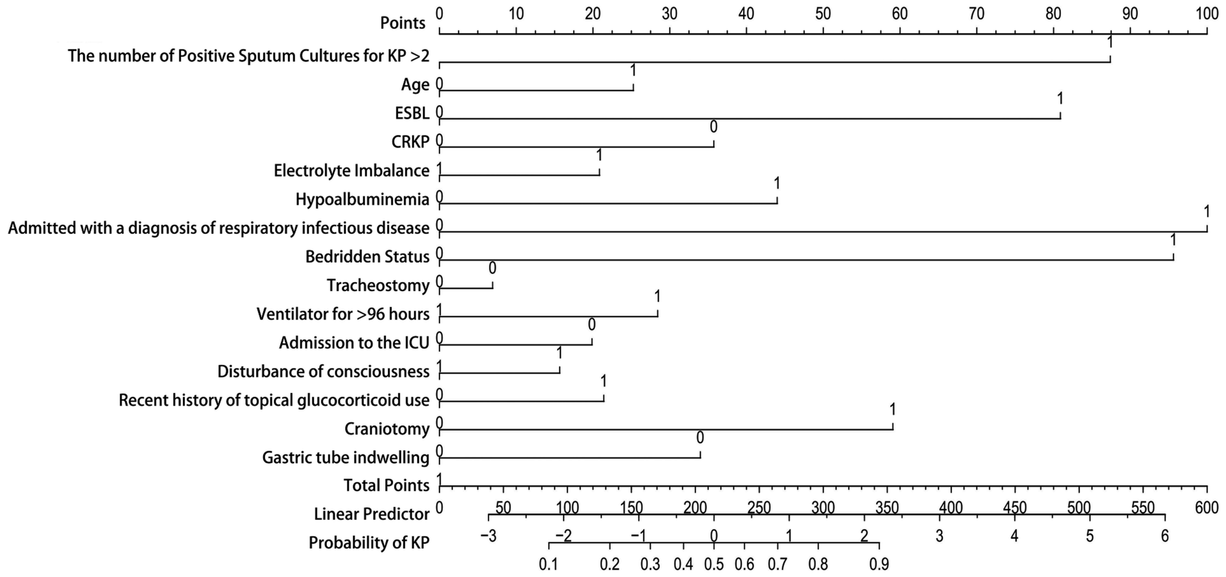


FIGURE 7
Nomogram model. The Y-axis of the nomogram represents various clinical factors, with each axis line corresponding to an input variable. The figure 0 on the axis indicates colonization, while the figure 1 indicates infection. The scale value on the X-axis represents the points or measurements of that clinical factors, and the corresponding point can be found to determine the contribution of that factor in the total score.

models through univariate logistic regression analysis. All six models achieved AUC values above 0.75, indicating that our research can provide a reference for differentiating between KP colonization and infection in clinical settings. The Random Forest model performed the best across all metrics, with an AUC of 0.99, an accuracy of 0.93, effectively distinguishing between KP colonization and infection status while significantly reducing the probability of misclassification. The precision was 0.98, indicating that nearly all samples predicted as infections were correct, thus reducing the occurrence of false positives. The Brier Score was 0.06, reflecting excellent predictive performance. In many practical scenarios, Random Forest has demonstrated outstanding performance, and this powerful ensemble learning algorithm is particularly well-suited for clinical data analysis and prediction (Jin et al., 2024; Mendapara, 2024; Wang Y. et al., 2024).

To better understand the relationship between the model and the data, we performed a visual analysis of the Random Forest model, which exhibited the best performance, by examining the importance of different factors using Mean Decrease Gini. These factors hold significant clinical relevance. Specifically, “Admitted with a diagnosis of respiratory infectious disease” is identified as the primary predictor of KP infection. The significant increase in infection risk for patients diagnosed with respiratory infectious diseases upon admission highlights the importance of early infection identification. This is particularly crucial as a large proportion of KP infections originate from the community, and studies have shown a strong correlation between KP-CAP (KP Community-Acquired Pneumonia) infections and higher, earlier mortality rates (Grosjean et al., 2024). This suggests that clinicians should promptly conduct pathogen detection and identification when seeing patients with respiratory infection symptoms, ensuring early detection, diagnosis, and treatment. Hypoalbuminemia, electrolyte imbalance, ESBL presence, and bedridden status are all important influencing factors for KP infection, closely related to the patient’s overall health status and immune function. Being bedridden indicates that the patient’s mobility is limited, which may lead to a series of health issues, such as muscle atrophy, venous thromboembolism, and complications like pneumonia. Hypoalbuminemia can result in malnutrition, reduced immunity, and poor healing capacity (Wang et al., 2023). This highlights the potential contribution of a patient’s nutritional and metabolic status to the risk of infection. For bedridden patients, it is crucial to regularly monitor serum albumin and electrolyte levels. An age greater than 62 years has been identified as an important risk factor, which is consistent with previously published studies indicating that nearly half of KP infections (45.7%) occur in the elderly population (Liu and Guo, 2019). This may be related to the decreased immune function and the prevalence of underlying diseases in the elderly population. The number of positive sputum cultures for KP (greater than 2) indicates a higher probability of infection, suggesting that multiple cultures are crucial for improving diagnostic accuracy in clinical practice. Invasive procedures such as tracheostomy and the placement of nasogastric tubes are also significant factors in KP infections. These invasive interventions can partially disrupt the normal structure of the upper respiratory tract, affecting the cleanliness and protective functions of the airway, making infections more likely. When the duration of mechanical ventilation exceeds 96 h, the risk of infection significantly increases. This may be related to the formation of respiratory biofilms, which can serve as breeding grounds for bacteria and elevate the risk of infection (Mishra et al., 2024). Craniotomy is also an important influencing factor, which may be because craniotomy is often difficult and lasting for a long time, and some patients are critically ill and need immediate surgery. Inadequate preoperative preparation and non-standard use of prophylactic antibiotics can lead to an increased probability of infection (He et al., 2024). Recent use of local corticosteroids can suppress local immune responses, reducing the body's defenses against infections and making it easier for infections to develop. This immunosuppressive effect can lead to an increased risk of KP infections, particularly in patients who may already be vulnerable due to other underlying health conditions or invasive procedures (Fernández Peláez et al., 2000). Patients with disorders of consciousness usually lack the cough reflex to effectively clear respiratory secretions, and at the same time, the probability of aspiration is increased, which increases the risk of KP infection (Kobata, 2023).

Although the Random Forest model demonstrated good overall performance, it is not very intuitive. In contrast, a Nomogram is an intuitive visual tool used to display the risks or outcomes of multivariable prediction models. It is easy to use and understand, and it clearly illustrates the relative contributions of various factors to the outcome, making it particularly suitable for personalized decision-making in clinical practice (Yang et al., 2022; Gao et al., 2024; Luo et al., 2024). This study found that certain factors can serve as important criteria for assessing whether KP is colonized. When a patient's sputum culture results indicate CRKP (Ceftazidime-Resistant KP), or if the patient has a nasogastric tube in place, is undergoing tracheostomy, or is in the ICU, we must exercise greater caution and thoroughness in assessment and judgment rather than blindly initiating antimicrobial therapy. Only by comprehensively considering the patient's specific circumstances can clinicians develop more rational treatment plans, significantly improving the targeting and effectiveness of treatment while minimizing the misuse of medications and the emergence of resistance.

Although this study provides important insights into the diagnosis and management of KP infections, several limitations remain, which we hope to overcome in future research. First, the study included only 286 cases of KP positivity, and the data were sourced from a single center. Future studies should aim to expand the sample size and conduct multi-center collaborative research to enhance the generalizability and reliability of the findings. Moreover, the determination of colonization and infection relies on clinicians' judgments and experiences, which may vary among different doctors and affect the accuracy of the results. We look forward to the development of more standardized clinical guidelines to reduce subjective differences among physicians and improve diagnostic consistency. Although this study identified 15 influencing factors, the occurrence of infections in clinical practice may be affected by other potential factors, such as the composition of the microbiome, environmental factors, and the psychological state of the patient. Additionally, molecular mechanisms, such as the development of bacterial resistance and the host's immune response, are also important influencing factors. However, due to limitations in laboratory conditions and funding, we were unable to include these factors in this study. We hope that future research can develop a more comprehensive clinical model to provide clinicians with a more scientific and accurate basis for their judgments. Additionally, we anticipate integrating our constructed model into a clinical decision support system (CDSS) in the future. Regular evaluation of the model's performance, along with the collection of feedback from healthcare professionals and patients, will be critical. This will allow us to update the data and retrain the model, continuously improving its usability and accuracy.

5 Conclusion

In this study, we successfully constructed six predictive models. Among them, the Random Forest model performed the best, achieving an area under the curve (AUC) of 0.99, with an accuracy of 0.93 and a precision of 0.98. This model effectively distinguishes between KP (KP) infection and colonization status, significantly reducing the rate of misjudgment. The Brier Score was 0.06, further validating the model's excellent predictive performance. Key factors such as being diagnosed with respiratory infections upon admission and the number of positive sputum cultures underscore the importance of early diagnosis and repeated cultures. Additionally, factors closely related to the overall health status of patients, such as being bedridden, electrolyte imbalances, and hypoalbuminemia, significantly impact the risk of KP infection. To enhance the model's applicability in clinical practice, we employed an intuitive nomogram, making it easier for clinicians to understand and utilize multivariable predictions. When faced with patients whose sputum cultures test positive for CRKP, or who have nasogastric tubes, tracheostomies, or are receiving ICU treatment, it is crucial to conduct a more cautious evaluation rather than initiating antimicrobial therapy blindly. This approach allows for more precise decision-making in patient management. We look forward to applying artificial intelligence technologies in future research to improve the accuracy of infection diagnoses, optimize antimicrobial treatment strategies, reduce unnecessary use of antibiotics, and thereby enhance patient outcomes and decrease the burden on healthcare systems.

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.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the First Affiliated Hospital of China Medical University [KLS(2023)2023-142-2]. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

XiaZ: Conceptualization, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. XifZ: Conceptualization, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. DZ: Conceptualization, Data curation, Formal analysis, Visualization, Writing – review & editing. JX: Conceptualization, Formal analysis, Methodology, Visualization, Writing – review & editing. JZ: Methodology, Resources, Writing – original draft, Writing – review & editing. XinZ: Methodology, Resources, 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 study was supported by the Community Infectious Diseases Research Capacity Building Program (J202201E026).

Acknowledgments

The images in this article are drawn with the help of BioRender and WPS office.

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2025.1508030/full#supplementary-material

References

Alnimr, A. (2023). Antimicrobial resistance in ventilator-associated pneumonia: predictive microbiology and evidence-based therapy. *Infect. Dis. Ther.* 12, 1527–1552. doi: 10.1007/s40121-023-00820-2

Antimicrobial resistance: global report on surveillance (n.d.). Available at: https://www.who.int/publications/i/item/9789241564748 (accessed July 23, 2024).

Cai, S., Wang, Z., Han, X., Hu, H., Quan, J., Jiang, Y., et al. (2024). The correlation between intestinal colonization and infection of carbapenem-resistant *Klebsiella pneumoniae*: a systematic review. *J. Glob. Antimicrob. Resist.* 38, 187–193. doi: 10.1016/j.jgar.2024.04.013

Chinese Medical Association, Chinese Medical Association Journal, Chinese Medical Association General Medicine Society, et al. (2019). Community-acquired pneumonia in adults basic guidelines (practice edition, 2018) [J]. *Chin. J. Gen. Pract.* 19, 127–133. doi: 10.3760/cma.j.issn.1671-7368.2019.02.006

CLSI (2023). Performance standards for antimicrobial susceptibility testing. 33rd ed.CLSI supplement M100. Clinical and Laboratory Standards Institute. Available at: https://iacld.com/UpFiles/Documents/672a1c7c-d4ad-404e-b10e-97c19e21cdce.pdf (Accessed April 7, 2024).

Córdova-Espinoza, M. G., Giono-Cerezo, S., Sierra-Atanacio, E. G., Escamilla-Gutiérrez, A., Carrillo-Tapia, E., Carrillo-Vázquez, L. I., et al. (2023). Isolation and identification of multidrug-resistant *Klebsiella pneumoniae* clones from the hospital environment. *Pathogens* 12:634. doi: 10.3390/pathogens12050634

Deo, R. C. (2015). Machine learning in medicine. *Circulation* 132, 1920–1930. doi: 10.1161/CIRCULATIONAHA.115.001593

Fernández Peláez, J. M., Sánchez Martín, E., Polo Romero, F. J., and Sáez Méndez, L. (2000). Lung infection caused by aspergillus fumigatus and *Nocardia asteroides* as complication of glucocorticoid treatment. *Med. Clin. (Barc.)* 114:358. doi: 10.1016/s0025-7753(00)71294-2

Gao, Y., Chen, L., Wen, Z., Jiang, H., and Feng, J. (2024). Risk factors for multidrug resistance in patients infected with Carbapenem-resistant *Klebsiella pneumoniae*: a nomogram. *Infect. Drug Resist.* 17, 4833–4841. doi: 10.2147/IDR.S479374

Gomez-Simmonds, A., and Uhlemann, A.-C. (2017). Clinical implications of genomic adaptation and evolution of Carbapenem-resistant *Klebsiella pneumoniae*. *J. Infect. Dis.* 215, S18–S27. doi: 10.1093/infdis/jiw378

Gonzalez-Ferrer, S., Peñaloza, H. F., Budnick, J. A., Bain, W. G., Nordstrom, H. R., Lee, J. S., et al. (2021). Finding order in the Chaos: outstanding questions in *Klebsiella pneumoniae* pathogenesis. *Infect. Immun.* 89, e00693–e00620. doi: 10.1128/IAI.00693-20

Gorrie, C. L., Mirceta, M., Wick, R. R., Edwards, D. J., Thomson, N. R., Strugnell, R. A., et al. (2017). Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin. Infect. Dis.* 65, 208–215. doi: 10.1093/cid/cix270

Grosjean, V., Gressens, S. B., Pham, T., Gaudry, S., Ait-Oufella, H., De Prost, N., et al. (2024). Community-acquired *Klebsiella pneumoniae* pneumonia in ICU: a multicenter retrospective study. *Ann. Intensive Care* 14:69. doi: 10.1186/s13613-024-01269-3

Guilhen, C., Miquel, S., Charbonnel, N., Joseph, L., Carrier, G., Forestier, C., et al. (2019). Colonization and immune modulation properties of *Klebsiella pneumoniae* biofilm-dispersed cells. *NPJ Biofilms Microbiomes* 5:25. doi: 10.1038/s41522-019-0098-1

He, K., Li, Y., and Liu, H. (2024). Risk and protective factors associated with wound infection after neurosurgical procedures: a meta-analysis. *Int. Wound J.* 21:e14699. doi: 10.1111/iwj.14699

Jiang, Y., Luo, J., Huang, D., Liu, Y., and Li, D.-D. (2022). Machine learning advances in microbiology: a review of methods and applications. *Front. Microbiol.* 13:925454. doi: 10.3389/fmicb.2022.925454

Jin, J., Lu, J., Su, X., Xiong, Y., Ma, S., Kong, Y., et al. (2024). Development and validation of an ICU-venous thromboembolism prediction model using machine learning approaches: a multicenter study. *Int. J. Gen. Med.* 17, 3279–3292. doi: 10.2147/IJGM.S467374

Jones, R. N. (2010). Microbial etiologies of hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia. *Clin. Infect. Dis.* 51, S81–S87. doi: 10.1086/653053

Kobata, H. (2023). Differential diagnosis and management of disturbance of consciousness: approach to comatose patients. *No Shinkei Geka* 51, 1009–1020. doi: 10.11477/mf.1436204847

Li, Y., Kumar, S., Zhang, L., Wu, H., and Wu, H. (2023). Characteristics of antibiotic resistance mechanisms and genes of *Klebsiella pneumoniae*. *Open Med. (Wars)* 18:20230707. doi: 10.1515/med-2023-0707

Liu, C., and Guo, J. (2019). Hypervirulent *Klebsiella pneumoniae* (hypermucoviscous and aerobactin positive) infection over 6 years in the elderly in China: antimicrobial resistance patterns, molecular epidemiology and risk factor. *Ann. Clin. Microbiol. Antimicrob.* 18:4. doi: 10.1186/s12941-018-0302-9

Lopera, C., Monzó, P., Aiello, T. F., Chumbita, M., Peyrony, O., Gallardo-Pizarro, A., et al. (2024). Prevalence and impact of multidrug-resistant bacteria in solid cancer patients with bloodstream infection: a 25-year trend analysis. *Microbiol. Spectr.* 12:e0296123. doi: 10.1128/spectrum.02961-23

Lou, T., Du, X., Zhang, P., Shi, Q., Han, X., Lan, P., et al. (2022). Risk factors for infection and mortality caused by carbapenem-resistant *Klebsiella pneumoniae*: a large multicentre case-control and cohort study. *J. Infect.* 84, 637–647. doi: 10.1016/j.jinf.2022.03.010

Luo, Y., Cui, X., Zhou, J., Zhuang, Y., Zheng, C., Su, Q., et al. (2024). Development and validation of a clinical nomogram for predicting complications from pediatric multiple magnet ingestion: a large retrospective study. *Am. J. Gastroenterol.* doi: 10.14309/ajg.0000000000002983

Mendapara, K. (2024). Development and evaluation of a chronic kidney disease risk prediction model using random forest. *Front. Genet.* 15:1409755. doi: 10.3389/fgene.2024.1409755

Mishra, A., Tabassum, N., Aggarwal, A., Kim, Y.-M., and Khan, F. (2024). Artificial intelligence-driven analysis of antimicrobial-resistant and biofilm-forming pathogens on biotic and abiotic surfaces. *Antibiotics (Basel)* 13:788. doi: 10.3390/antibiotics13080788

Podschun, R., and Ullmann, U. (1998). Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin. Microbiol. Rev.* 11, 589–603. doi: 10.1128/CMR.11.4.589

Seyer Cagatan, A., Taiwo Mustapha, M., Bagkur, C., Sanlidag, T., and Ozsahin, D. U. (2022). An alternative diagnostic method for *C. neoformans*: preliminary results of deep-learning based detection model. *Diagnostics (Basel)* 13:81. doi: 10.3390/diagnostics13010081

Subspecialty Group of Infectious Diseases, Respiratory Society, Chinese Medical Association (2018). Chinese guidelines for the diagnosis and treatment of hospital-acquired pneumonia and ventilator-associated pneumonia in adults (2018 edition) [J]. *Chin. J. Tubere. Respir. Dis.* 41, 255–280. doi: 10.3760/cma.j.issn.1001-0939.2018.04.006

Vornhagen, J., Rao, K., and Bachman, M. A. (2024). Gut community structure as a risk factor for infection in *Klebsiella pneumoniae*-colonized patients. *mSystems* 9, e00786–e00724. doi: 10.1128/msystems.00786-24

Wang, X., Liu, J., and Li, A. (2024). Incidence and risk factors for subsequent infections among rectal carriers with carbapenem-resistant *Klebsiella pneumoniae*: a systematic review and meta-analysis. *J. Hosp. Infect.* 145, 11–21. doi: 10.1016/j.jhin.2023.12.002

Wang, Q.-R., Long, J., Wang, C.-C., Hu, J.-L., Lin, N., and Tang, S.-H. (2023). Case report of atypical undernutrition of hypoproteinemia type. *Open Life Sci.* 18:20220766. doi: 10.1515/biol-2022-0766

Wang, Y., Yao, H.-X., Liu, Z.-Y., Wang, Y.-T., Zhang, S.-W., Song, Y.-Y., et al. (2024). Design of Machine Learning Algorithms and Internal Validation of a kidney risk prediction model for type 2 diabetes mellitus. *Int. J. Gen. Med.* 17, 2299–2309. doi: 10.2147/IJGM.S449397

Yang, L., Wang, Y., He, X., Liu, X., Sui, H., Wang, X., et al. (2022). Development and validation of a prognostic dynamic nomogram for in-hospital mortality in patients with Stanford type B aortic dissection. *Front. Cardiovasc. Med.* 9:1099055. doi: 10.3389/fcvm.2022.1099055

Zhu, L., Liang, L., Hui, J., Lu, J., Yang, R., He, Q., et al. (2023). Relationship between antibiotic exposure and carbapenem-resistant *Klebsiella pneumoniae* infection within four types of control patients: a systematic review and meta-analysis. *J. Glob. Antimicrob. Resist.* 33, 137–151. doi: 10.1016/j.jgar.2023.02.020

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 May 2025
DOI 10.3389/fmicb.2025.1510126

Check for updates

OPEN ACCESS

EDITED BY
Marcin Grzegorzek,
University of Lübeck, Germany

REVIEWED BY
Sasikala Muthusamy,
Brigham and Women's Hospital and Harvard Medical School, United States
Fanshuo Yin,
Ningbo First Hospital, China
Haodi Cao,
Capital Medical University, China
Kien Vu Trung,
University Hospital Leipzig, Germany

*CORRESPONDENCE
Xinglong Zhang
✉ zhang_cmu4h@163.com

RECEIVED 12 October 2024
ACCEPTED 22 April 2025
PUBLISHED 14 May 2025

CITATION
Zang W, Lin Z, Zhao Y, Jia T and Zhang X (2025) Construction of a predictive model for rebleeding risk in upper gastrointestinal bleeding patients based on clinical indicators such as *Helicobacter pylori* infection.
Front. Microbiol. 16:1510126.
doi: 10.3389/fmicb.2025.1510126

Construction of a predictive model for rebleeding risk in upper gastrointestinal bleeding patients based on clinical indicators such as *Helicobacter pylori* infection

Wei Zang[1], Ze Lin[2], Yanduo Zhao[2], Tianshi Jia[2] and Xinglong Zhang[3]*

[1]Department of Radiology, ShengJing Hospital of China Medical University, Shenyang, Liaoning, China, [2]The Second Clinical College of China Medical University, Shenyang, Liaoning, China, [3]The Fourth Affiliated Hospital of China Medical University, Shenyang, Liaoning, China

Background: The annual incidence of upper gastrointestinal hemorrhage (UGIB) is about 60 cases/100,000 people, and about 40% of UGIB patients have hemorrhagic ulcers. Ulcer formation is often associated with *Helicobacter pylori* (*H. pylori*) infection, non-steroidal anti-inflammatory drugs (NSAIDs) use and other factors, so ulcerative disease is the main cause of upper gastrointestinal bleeding. *H. pylori* induces chronic superficial gastritis with neutrophils infiltrating into the mucosa, so it is assumed that *H. pylori* infection is the basis of bleeding lesions. *H. pylori* infection is widespread worldwide, with about 50% of the population carrying the bacteria. Mortality during hospitalization is higher in patients with UGIB because rebleeding significantly increases the risk of death, especially if timely intervention is not provided. Rebleeding may also lead to severe complications such as shock and multiple organ failure. At present, the commonly used clinical scores for UGIB patients mainly include Rockall score (RS), AIMS65 score and Glasgow-Blatchford score (GBS). Because some hospitals are limited by local medical and health conditions, they lack timely and accurate endoscopic diagnosis and treatment equipment, and it is difficult to make accurate and timely judgments on patients.

Method: In this experiment, 254 patients with upper digestive tract hemorrhage from Shengjing Hospital affiliated to China Medical University were collected, and the clinical indicators and information of *H. pylori* infection, age, shock state, concomitant disease, *H. pylori* infection degree, systolic blood pressure, blood urea nitrogen, hemoglobin, pulse, black stool, syncope, liver disease and other patients were finally collected. We analyzed the correlation between various clinical indicators and rebleeding in hospitalized patients. Based on the collected clinical information and laboratory indicators, this study constructed a deep learning model, the data is divided into four categories (clinical information, vital signs, laboratory examination items, stool examination) as input, and Transformer is used as feature extractor. KAN as a classifier to predict the risk of rebleeding in patients with upper gastrointestinal bleeding. The model uses five-fold cross validation and calculates key metrics such as accuracy to evaluate its performance. In addition, the deep learning model was compared with a variety of machine learning methods (decision tree, random forest, logistic regression, K-nearest neighbor) and common clinical risk scores (Rockall score, AIMS65 score, Glasgow-Blatchford score) to verify its effectiveness and advantages. In order to highlight the importance of *H. pylori* infection degree to the model performance, we conducted a comparative experiment to observe the role of *H. pylori* infection degree in the model.

Results: In the correlation analysis between rebleeding and clinical data and related indicators, the risk of rebleeding in men (62.5%) was higher than that in women (43.47%), and the risk of rebleeding in patients with concurrent diseases (60.37%) was higher than that in patients without concurrent diseases. In the analysis of the correlation between the degree of infection and the laboratory test items, the hemoglobin level of patients will also change with the change of the degree of infection of patients ($p < 0.05$ in the above correlation analysis, all had statistical significance). The rebleeding detection rates of Rockall score, AIMS65 score and Glasgow Blatchford score were 16.14%, 0 and 77.17%, respectively. Of the four machine learning models, Random Forest (RF) had the highest accuracy on the test set at 0.68. The accuracy of the deep learning model on the verification set is the highest of 0.9750, and the accuracy of the test set is the highest of 0.9615. In addition, by exploring the influence of infection on the model prediction, it was found that the prediction accuracy of rebleeding in the non-*H. pylori* infection group (0.8989) was lower than that in the *H. pylori* infection group (0.9636), and other evaluation parameters were also lower than that in the infection group. In addition, by adding irrelevant random noise to mask the influence of infection degree on model output, it is found that the model prediction accuracy (0.7992) is significantly reduced.

Conclusion: Based on the degree of *H. pylori* infection in patients with upper gastrointestinal bleeding, combined with a number of clinical laboratory tests and clinical data, we developed a clinical model for predicting the risk of rebleeding in patients with upper gastrointestinal bleeding. It provides an early prediction of rebleeding during a patient's hospitalization and optimizes early intervention for patients to a certain extent. It provides a more concise, convenient and effective guidance scheme for small and medium-sized hospitals to make clinical decisions for UGIB patients.

KEYWORDS

deep learning, UGIB, *Helicobacter pylori*, Rockall score, Glasgow-Blatchford score, AIMS65 score

1 Introduction

Upper gastrointestinal hemorrhage (UGIB) refers to the occurrence of blood in the esophagus, stomach and duodenum, and the clinical symptoms of hematemesis or black stool (Lanas et al., 2009; Thiebaud et al., 2017). It is estimated that the annual incidence of UGIB is about 60 cases / 100,000 people (Gu et al., 2023; Laine et al., 2012), among which gastric and duodenal ulcers are the common causes of UGIB, according to statistics, about 30% to 50% of UGIB cases are related to peptic ulcers (Kamboj et al., 2019). Ulcer formation is often associated with *H. pylori* infection, the use of non-steroidal anti-inflammatory drugs (NSAIDs), and excessive alcohol consumption. Studies have shown that *H. pylori* can induce chronic superficial gastritis, and neutrophils infiltrate the mucosa (Popa et al., 2021). *H. pylori* infection will damage the inner wall of blood vessels, resulting in impaired vascular skin function, resulting in decreased vascular tension, and promoting the development of acute upper gastrointestinal bleeding. Moreover, under the action of *H. pylori*, a large number of inflammatory factors will be produced, resulting in increased inflammation and a large number of platelet aggregation in blood vessels. It can cause thrombosis, block blood vessels and affect blood clotting function (Toews et al., 2024). Therefore, it is speculated that *H. pylori* infection is the basis of bleeding lesions. According to current guidelines, patients with UGIB should undergo esophagogastroduodenoscopy (EGD) within 12 to 24 h after hemodynamic resuscitation (Gralnek et al., 2021; Karstensen et al., 2020). When erythema with recent bleeding is observed through EGD, endoscopic treatment is required to reduce mortality, recurrent bleeding, and surgical intervention rates (Veisman et al., 2022). Mortality and risk of rebleeding in UGIB patients should be evaluated after endoscopic and laboratory examinations. Mortality during hospitalization is higher in patients with UGIB because rebleeding significantly increases the risk of death, especially if timely intervention is not provided. Rebleeding may also lead to severe complications such as shock and multiple organ failure (Chen et al., 2021; Matsuhashi et al., 2021; Shung and Laine, 2024). Common clinical scores include Rockall score (RS), AIMS65 score, and Glasgow-Blatchford score (GBS). RS score is mainly used for risk assessment of rebleeding and death, and a number of clinical prediction models rely on RS score. The accuracy of this score has always been an advantage due to its endoscopy project (Redondo-Cerezo et al., 2020; Seo et al., 2020; Shung et al., 2020), but many hospitals are limited by their own conditions, and it is difficult to achieve timely and effective use of endoscopy. Therefore, GBS score and AIMS65 score independent of endoscopy project have wider application scenarios, high sensitivity, good generality, simple economy, suitable for early application in emergency treatment, and can also be used to predict UGIB inpatient mortality (Stanley et al., 2017), while GBS and AIMS65 score are not as accurate as RS score. This brings some difficulties to the diagnosis and treatment of patients (Alali et al., 2023; Cazacu et al., 2023). Deep learning is a kind of machine learning method based on artificial neural network, its core idea is to automatically extract and represent complex features of data through multi-layer neural network structure. Deep learning implements complex feature learning through invisible transformations between layers and deep learning models automatically extract useful features from inputs through multi-system neural networks, avoiding the strong dependence on feature engineering in traditional machine learning (Koetzier et al., 2023; LeCun et al., 2015; Schmidhuber, 2015). In recent years, the diagnosis and treatment of digestive system diseases has been fully developed through deep learning, which mostly focuses on the diagnosis and treatment of digestive system tumors and the establishment of deep learning models through endoscopic content, and the semi-automatic or fully automated diagnosis and treatment of diseases (Dong et al., 2022; Kim et al., 2023; Zhang et al., 2023; Zhuang et al., 2022).

However, the current deep learning models using clinical indicators are still not completely divorced from endoscopy, but there are few studies on *H. pylori* infection. Therefore, this study established a deep learning model by obtaining the *H. pylori* infection situation of patients with upper gastrointestinal bleeding and combining with a number of clinical indicators to assess the risk of rebleeding in patients with UGUB in the absence of endoscopic information.

2 Methods

2.1 Inclusion criteria

A total of 254 UGIB patients admitted to Shengjing Hospital Affiliated to China Medical University from January 2017 to June 2024 were selected and divided into *H. pylori* infection group and *H. pylori* non-infection group according to whether they were infected with *H. pylori*. There were 165 patients in *H. pylori* infection group and 89 patients in *H. pylori* non-infection group. The patients with *H. pylori* infection UGIB were further divided into weak positive (17 cases), positive (4 cases), weak positive for active infection (42 cases) and positive for active infection (102 cases). Inclusion criteria for UBIG patients: (1) Clinical manifestations included hematemesis, black stool, shock or other gastrointestinal bleeding symptoms; (2) Decreased hemoglobin level, decreased coagulation function, stool occult blood and other laboratory positive tests; (3) CT and endoscopy indicated upper gastrointestinal bleeding; (4) *H. pylori* infection was examined after hospitalization. Exclusion criteria: (1) Patients with past operations on upper gastrointestinal tract; (2) Anti-*H. pylori* infection treatment during hospitalization; (3) Patients with venous upper gastrointestinal bleeding. The flow chart of the row is shown in Figure 1. Detection of *H. pylori* infection in UGIB patients based on serological testing (antibody testing) and detection of *H. pylori* electrophoresis IgG antibodies. If the patient had *H. pylori* before and had anti-*H. pylori* infection, he was previously infected with *H. pylori*; if the patient had no *H. pylori* infection detected in the past and no anti-*H. pylori* infection in the past, he was currently infected with *H. pylori*. Combined with high titer warning and breath test, the infection was divided into positive and weak positive.

2.2 Rating

2.2.1 Rockall rating

Rockall belongs to the post-endoscopic risk score, which was published in 1996 and was designed to predict mortality based on 4,185 UGIB cases. The full score ranges from 0 to 11 and includes five variables, two of which depend on endoscopy results: age, hemodynamic stability, comorbidities, endoscopic diagnosis, and erythema of recent bleeding (Rockall et al., 1996). According to the score, patients can be divided into high risk (≥5 scores), medium risk (3–4 scores), and low risk (0–2 scores). Specific scores are shown in Table 1.

2.2.2 Blatchford rating

GBS belongs to the preendoscopic risk score, which was published in 2000 and is based on 1,748 UGIB patients with a full score of 0 to 23 and includes eight variables: blood urea nitrogen; Hemoglobin; Systolic blood pressure; Heart rate; Black stool; Fainting; Presence of liver disease (known history or clinical/laboratory evidence) and heart disease (known history or clinical/laboratory evidence). The objective is to determine whether the patient requires intervention, defined as transfusion, endoscopic or surgical intervention, death, or rebleeding (Blatchford et al., 2000). According to the score, patients can be divided into medium-high risk (≥6 scores) and low-risk (<6 scores). Specific scores are shown in Table 2.

2.2.3 AIMS65 score

The AIMS65 score, a preendoscopic risk score published in 2011, included 29,222 UIGB patients admitted to 187 hospitals from 2004 to 2005. AIMS65 was externally validated 1 year later, and a total of 32,504 patients were included in the database for its development. On a scale of 0 to 5, AIMS65 includes five clinical or laboratory variables: age of onset; Albumin, INR (as an indicator of coagulation function of patients, “INR” is an abbreviation of the international standardized ratio, derived from the international sensitivity index of prothrombin time and determination reagents); Changes in mental status and systolic blood pressure. Used to predict patient mortality (Saltzman et al., 2011). According to the score, patients can be classified into high risk (≥2 scores) and low risk (<2 scores). Specific scores are shown in Table 3.

Three scores were used to perform risk scores for the enrolled patients and compared with final patient outcomes and predictions in the deep learning model to show the efficacy of the current scoring model.

2.3 Methods for determining rebleeding

According to the clinical manifestations of the patient during the hospital, such as whether the patient has hematemesis, hematochezia or black stool, and the changes of the patient’s clinical indicators, such as changes in the hemoglobin level and coagulation function such as

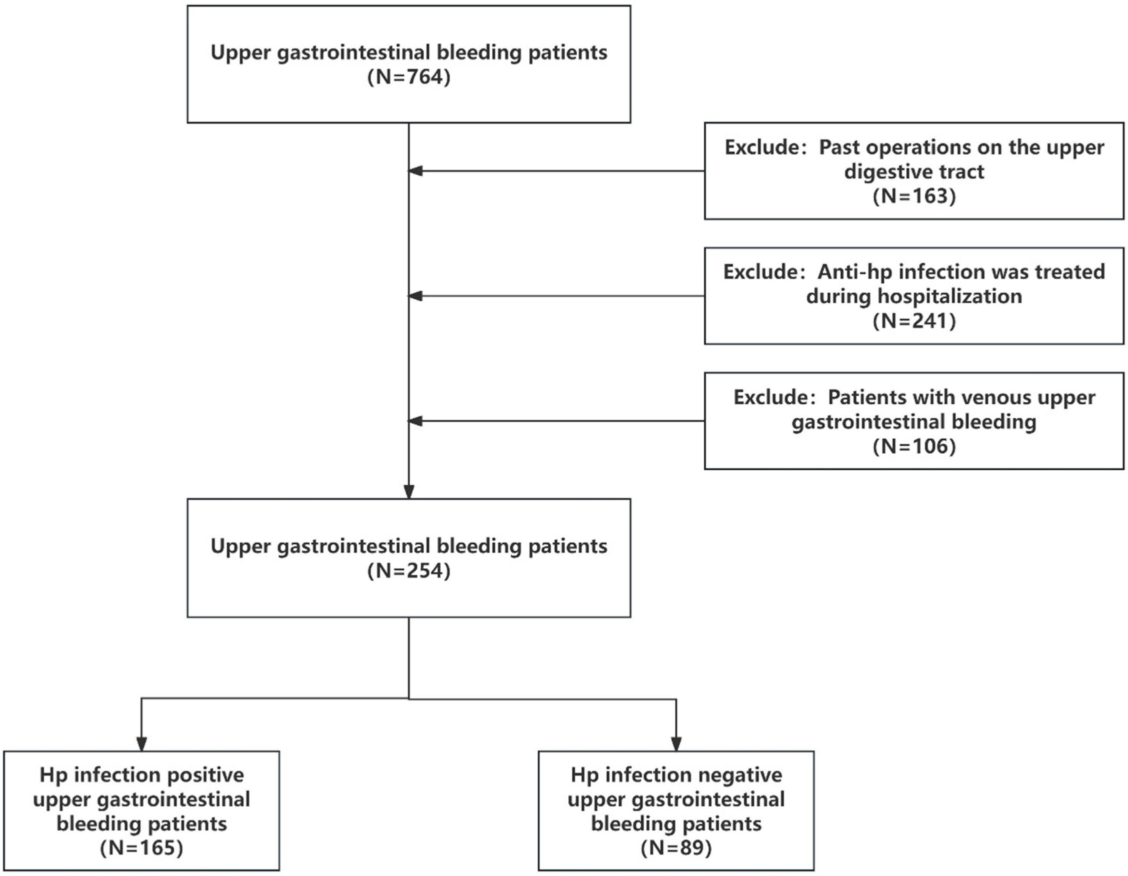


FIGURE 1
Flow chart of admission and discharge.

platelets during the hospital stay, if the patient has undergone endoscopy during the hospital stay, a comprehensive judgment will be made on whether the patient has rebleeding during the hospital stay combined with the contents of the endoscope.

2.4 Machine learning methods

Four machine learning methods were used in this study, including Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN).

DT is a classification or regression model with a tree-like structure that splits a sample into different branches by conditions to ultimately generate a prediction result. Maximize the information gain (or reduce the Gini coefficient) after each partition by dividing the input feature space. The gradient of the tree is generated from the root node to the leaf node, and the leaf node stores the prediction results. It is fast to train, easy to visualize, does not require data consumption (such as normalization), can handle multiple classification problems, but is sensitive to data noise. RF is an integrated learning algorithm based on decision trees that builds multiple decision trees and outputs the results by voting. The bagged method is used to randomly sample multiple sub-data sets from the original data. Each tree selects some features during training and obtains the final prediction result by voting (classification) or averaging (regression). It is suitable for classification and regression, and can handle nonlinear relationships between high-dimensional data and features. But computation costs are high, training and prediction are slow, and interpretative differences are output. LR is a linear model used for binary classification problems. It predicts the probability of an event occurring by using a linear combination of feature inputs and converts the output to a probability value between 0 and 1 using the Sigmoid function. The model is simple, easy to understand and implement, and can provide a probability value for each predicted output to explain and make decisions. At the same time, compared with other complex models, it has faster training and prediction speed and wider applicability, but it is more sensitive to noise and outliers, which may affect the model performance and can only handle binary classification problems. KNN algorithm is an instance-based learning method used for classification and regression problems. It does not go through an explicit training process, but instead calculates the distance (usually using Euclidean distance or Mahalanobis distance) between the predicted sample and the training set samples directly, and finds the nearest K neighbors. The prediction is made based on the categories (or numerical values) of these neighbors. Its transformation is simple to implement and easy to understand, and does not require a training process. It is suitable for problems without explicit distribution assumptions and can handle classification and regression problems, as well as multi-class classification. However, its computational cost is high and it is not suitable for high-dimensional data with irregularities.

TABLE 1 Content of Rockall rating.

Variables	Score
Age	
60–79 years old	1
>80 years old	2
Shock index	
Heart rate (beats per minute) > 100 and	1
Systolic blood pressure (mmHg) > 100	
Heart rate (beats per minute) > 100 and	2
Systolic blood pressure (mmHg) < 100	
Concomitant disease	
Heart failure, ischemic cardiomyopathy and other important concomitant diseases	2
Liver failure, kidney failure and disseminated malignancy	3
Endoscopic diagnosis	
Mallory Weiss had lacerations, no lesions	0
All other diagnoses	1
Malignant tumors of the upper digestive tract	2
Erythema of recent bleeding	
None, only dark spots	0
Blood, adhesion clot, ejection blood vessel	2

TABLE 2 Blatchford ratings.

Variables	Score
Heart rate (beats per minute)	
≥100	1
Heart rate (beats per minute)	
100–199	1
90–99	2
<90	3
Heart rate (beats per minute)	
6.5–7.9	2
8.0–9.9	3
10.0–24.9	4
>25.0	6
Hemoglobin (mmol/L)	
Male 120–130	1
110–119	3
<100	6
Female 100–120	1
<100	6
Other indicators	
Pulse ≥ 100	1
Accompanied by black stool	1
Present as syncope	2
Liver disease	2
Heart failure	2

TABLE 3 AIMS65 score content.

Variables	Score
Albumin < 30 g/L	1
INR > 1.5	1
Altered mental state	1
Systolic blood pressure (mmHg) ≤ 90	1
Age > 65 years	1

2.5 Deep learning model

This study classifies the information into four major categories: clinical information (including gender, age, concomitant diseases such as heart failure, ischemic heart disease, and other significant concomitant diseases, liver failure, renal failure, cancer metastasis, other manifestations, and diabetes history), vital signs (including systolic blood pressure, syncope, heart failure), blood indicators (including the degree of infection, hemoglobin, INR, and blood urea nitrogen), and stool examination (melena).

This study proposes a novel model based on Transformer and Kolmogorov–Arnold Networks (KAN), where the final linear layer of the Transformer is replaced by KAN. In the structure of the model, after each category of information is processed by the Transformer, one feature value is output. Finally, the feature values of the four types of information are concatenated and input into the final classification layer (KAN). This fusion structure not only combines the powerful feature extraction capability of Transformer, but also generates symbolic formula by introducing KAN to facilitate the analysis of the influence of different dimension information on the prediction results. The performance evaluation of the model is conducted by plotting the line graphs of the loss function (loss) and accuracy (accuracy). This approach visually demonstrates the convergence of the model during the training and validation processes as well as the changing trend of the classification performance, facilitating the assessment of the model's stability and generalization ability. In order to evaluate the generalization ability of deep learning models more comprehensively and stably, and make full use of limited data to avoid overfitting, this paper used five-fold cross validation to test the model.

In addition, we wanted to explore whether the infection situation would affect the prediction efficiency of the deep learning model. We designed two comparative experiments: one was to predict the best trained model in the *H. pylori* infected group and the *H. pylori* non-infected group; the other was to evaluate the prediction efficiency of the best model by setting the degree of infection as irrelevant random noise.

2.6 Statistics methods

IBM SPSS 26.0 software was used for statistical analysis. We first run the Kolmogorov–Smirnov normality test on all the data. For binary classification, the Mann–Whitney U test was used for inter-group differences of continuous variables, Chi-square test was used for inter-group differences of class variables, and Kruskal-Wallis H test was used for multi-class variables and continuous variables. The odds ratio and 95% confidence interval were increased for

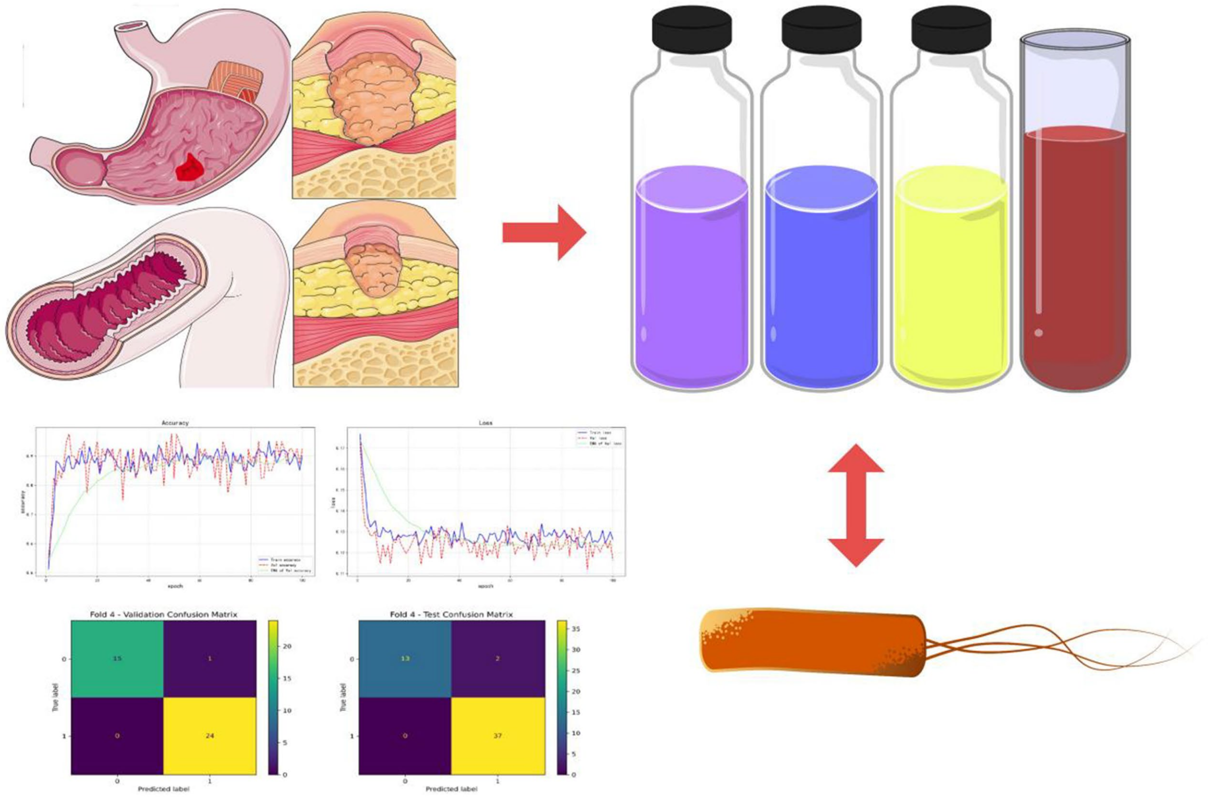


FIGURE 2
Experimental flow chart. Various laboratory indicators such as blood tests and stool tests were gathered from patients with acute upper gastrointestinal bleeding. In combination with the patients' *Helicobacter pylori* infection, a risk prediction model for rebleeding in inpatients with UGIB was constructed.

Kruskal-Wallis H test and Chi-square test. For Mann–Whitney U test, r effect size was increased. $p < 0.05$ was considered statistically significant.

The flow chart of the experiment is shown in Figure 2.

3 Results

3.1 Baseline

A total of 254 patients were included in the study, and variables with missing values greater than 20% were excluded. Table 4 shows clinical information and laboratory parameters for included patients.

3.2 Correlation between rebleeding and various indicators

3.2.1 Correlation between rebleeding and clinical information

Table 5 shows the correlation between rebleeding in included patients and clinical information. In the correlation analysis between rebleeding and clinical data, the risk of rebleeding in males (62.5%) was higher than that in females (43.47%) ($p < 0.05$, with statistical significance). The risk of rebleeding increased with age ($p > 0.05$, without statistical significance). The risk of rebleeding in patients with combined disease (60.37%) was higher than that in patients without combined disease (58.70%) ($p < 0.05$, with statistical significance). The risk of rebleeding was higher in patients without diabetes (59.21%) than in patients with diabetes (56.00%) ($p < 0.05$, without statistical significance).

3.2.2 Correlation between rebleeding and vital signs

Table 6 shows the correlation between rebleeding and vital signs in included patients. In the correlation analysis between rebleeding and vital signs, the risk of rebleeding gradually increased with the increase of systolic pressure ($p > 0.05$, without statistical significance), and the risk of rebleeding in patients without fainting (60.57%) was higher than that in patients with fainting (12.50%) ($p < 0.05$, with statistical significance).

3.2.3 Correlation between rebleeding and laboratory examination items

Table 7 shows the correlation between rebleeding in included patients and the items examined in the laboratory. In the correlation analysis between rebleeding and laboratory examination items, the probability of rebleeding did not change with the increase of *H. pylori* infection degree

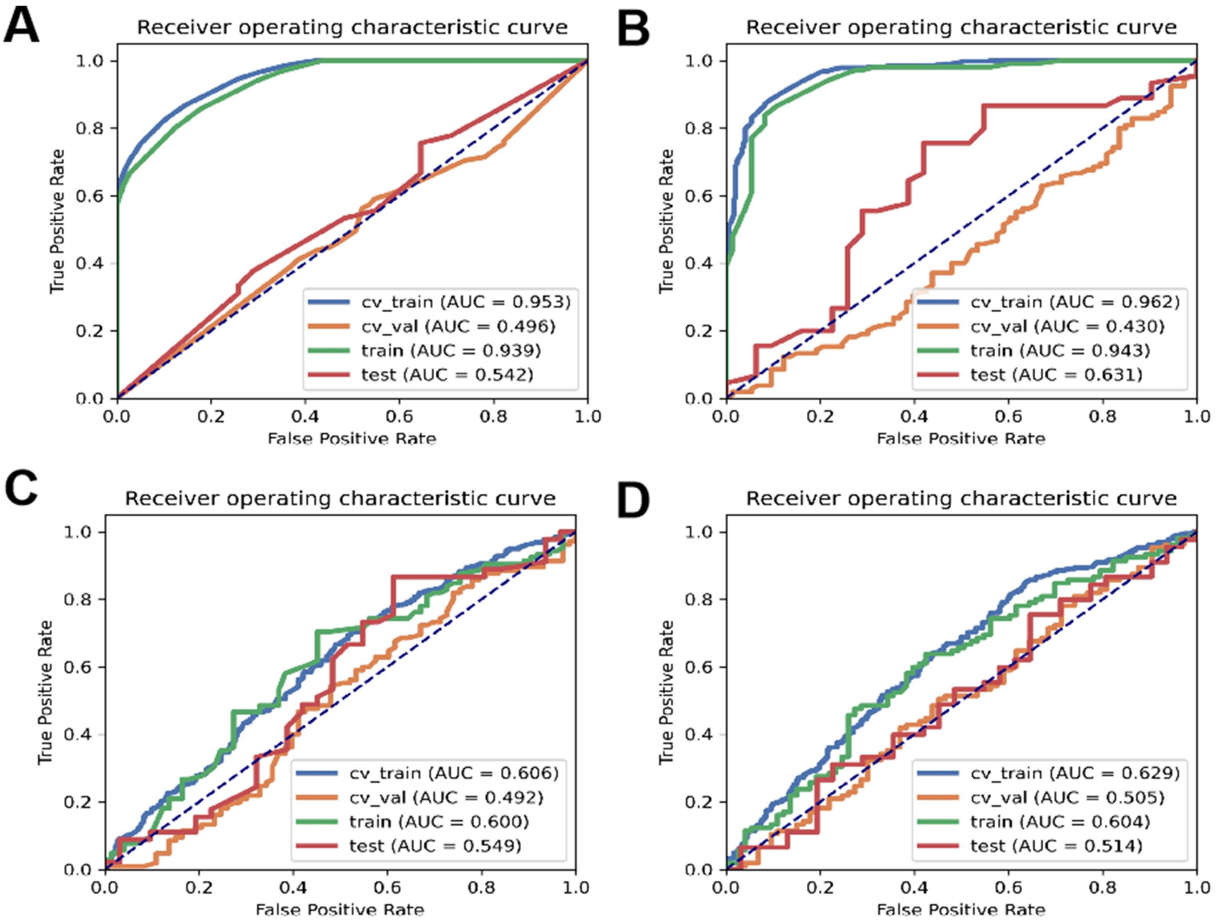


FIGURE 3
(A) ROC curve of DT, **(B)** ROC curve of RF, **(C)** ROC curve of LR, **(D)** ROC curve of KNN.

TABLE 4 Basic information of patients.

	Rehaemorrhagia	Normal
Sex	M130/F20	M78/F26
Age	52.23 (49.44,55.01)	53.85 (50.67,57.02)
Concomitant diseases	Y32/N118	Y21/N83
Whether have diabetes	Y15/N135	Y11/N93
Systolic pressure	121.01 (117.96,124.06)	123.28 (119.73,126.82)
Whether syncope	Y1/149	Y7/97
Whether heart failure	Y0/N150	Y0/N104
Degree of infection	N60/A9/B1/C22/D58	N29/A8/B3/C20/D44
Hemoglobin	92.08 (88.13,96.03)	93.32 (87.73,98.91)
INR	0.89 (0.77,1.01)	0.98 (0.77,1.20)
Blood urea nitrogen	10.72 (2.51,18.92)	6.32 (5.37,7.27)
Ulcer lesions	Y87/N63	Y56/N48
Tarry stool	Y131/N19	Y89/N15

($p > 0.05$, without statistical significance). With the changes of hemoglobin, INR and blood urea nitrogen, the probability of rebleeding would also change ($p > 0.05$, without statistical significance).

3.2.4 Correlation between rebleeding and stool examination

Table 8 shows the correlation between rebleeding and stool examination in included patients. In the correlation analysis between rebleeding and stool examination, the risk of rebleeding in patients with black stool (59.55%) was higher than that in patients without black stool (55.88%) ($p > 0.05$, without statistical significance).

3.2.5 Correlation between *Helicobacter pylori* infection degree and laboratory examination items

Table 9 shows the correlation between *H. pylori* infection degree and laboratory examination items of included patients. In the correlation analysis between the degree of *H. pylori* infection and laboratory examination items, the hemoglobin level of patients changed with the degree of infection ($p < 0.05$, with statistical significance), and the other laboratory indicators of patients also changed with the degree of infection ($p > 0.05$, without statistical significance).

TABLE 5 Correlation between rebleeding and clinical information.

	Statistic	*p* value	Statistical method	r effect size	Odds ratio	95%CI lower	95% CI upper
Sex	4.8776	0.0272*	Chi-square test		2.1667	1.1344	4.1380
Age	8449.5	0.2595	Mann–Whitney U test	0.0708			
Concomitant diseases	0.0040	0.9497	Chi-square test		1.0719	0.5778	1.9883
Whether have diabetes	0.7921	0.6730	Chi-square test		0.8768	0.3813	2.0162

*The difference is statistically significant ($p < 0.05$).

TABLE 6 Correlation between rebleeding and vital signs.

	Statistic	*p* value	Statistical method	r effect size	Odds ratio	95%CI lower	95% CI upper
Systolic pressure	8458.5	0.2510	Mann–Whitney U test	0.0720			
Whether fainting	5.5495	0.0185*	Chi-square test		0.0930	0.0113	0.7678

*The difference is statistically significant ($p < 0.05$).

TABLE 7 Correlation between rebleeding and laboratory examination items.

	Statistic	*p* value	Statistical method	r effect size	Odds ratio	95%CI lower	95% CI upper
Degree of infection	5.7306	0.2202	Chi-square test		0.5438	0.1902	1.5546
Hemoglobin	7916.5	0.8403	Mann–Whitney U test	0.0126			
INR	7,787	0.9822	Mann–Whitney U test	0.0014			
Blood urea nitrogen	7821.5	0.9709	Mann–Whitney U test	0.0023			

TABLE 8 Correlation between rebleeding and stool examination.

	Statistic	*p* value	Statistical method	Odds ratio	95%CI lower	95% CI upper
Tarry stool	0.0470	0.8283	Chi-square test	1.1620	0.5608	2.4078

TABLE 9 Correlation between degree of infection and stool examination.

	Statistic	*p* value	Statistical method	95%CI lower	95% CI upper
Hemoglobin	11.6905	0.0198*	Kruskal-Wallis H test	4.6882	30.3764
INR	1.8235	0.7682	Kruskal-Wallis H test	0.8621	14.6232
Blood urea nitrogen	2.3471	0.6722	Kruskal-Wallis H test	0.8160	15.7447

*The difference is statistically significant ($p < 0.05$).

TABLE 10 Three categories of ratings and their detection rates.

	RS	GBS	AIMS65 score
Low risk	213	58	254
Middle and high risk	41	196	0
Detection rate	16.14%	77.17%	0

3.3 Comparison between three common ratings, machine learning algorithm and deep learning algorithm model

3.3.1 Results of three common ratings

The three scores for the patients in this study are shown in Table 10.

3.3.2 Comparison of accuracy of four machine learning scoring methods

According to Table 11, the accuracy of the test set for DT analysis of machine learning methods is 0.53; the accuracy of the test set of RF machine learning method is 0.68; the accuracy of the test set of the machine learning method of LR machine learning method is 0.67; the accuracy of the test set of KNN machine learning method is 0.59.

TABLE 11 Results of machine learning algorithm.

	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC	95%CI Lower	95% CI Upper
DT	0.5337	0.5905	0.4521	0.6078	0.4342	0.4959	0.4145	0.5808
RF	0.6842	0.7556	0.5806	0.7234	0.6207	0.6312	0.4933	0.7667
LR	0.6711	0.8667	0.3871	0.6724	0.6667	0.5491	0.4073	0.6889
KNN	0.5921	0.7556	0.3548	0.6296	0.5	0.514	0.3777	0.6449

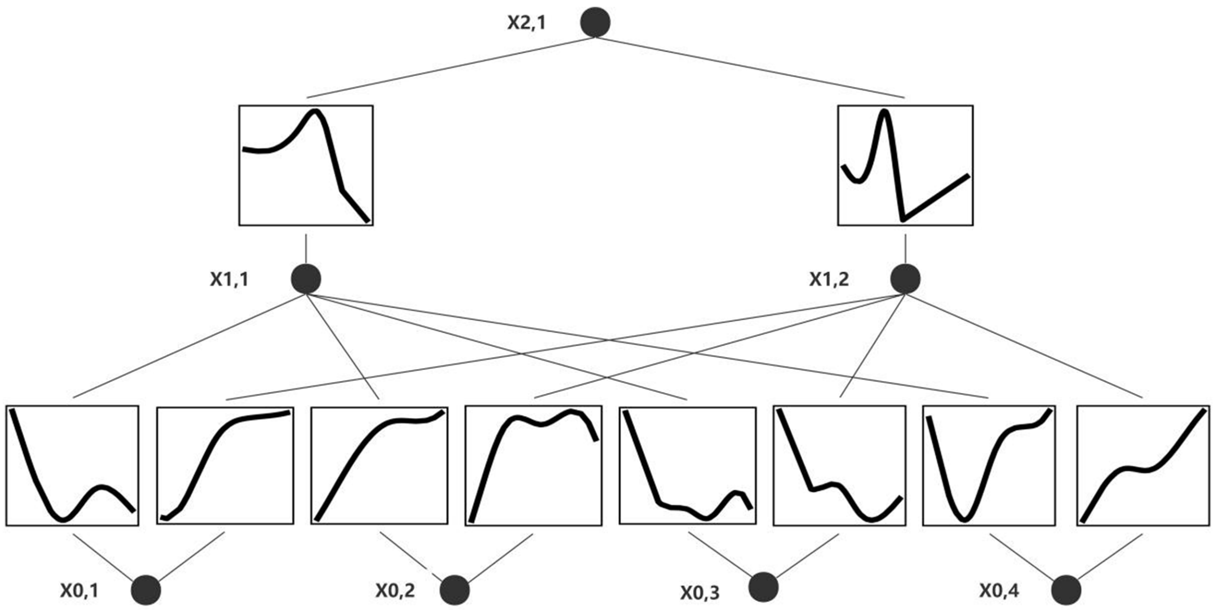


FIGURE 4
Deep learning model visualization. x0,1: clinical information, x0,2: vital signs, x0,3: blood parameters, x0,4: stool examination.

ROC curves for the four types of machine learning are shown in Figure 3.

3.3.3 Effect of deep learning model

In this study, after processing each type of information by transformer, a characteristic value will be output, and the characteristic value will be represented by x1-4, respectively. x1-4 represents clinical information, vital signs, laboratory examination items and stool examination respectively, which can be understood as dimensionally reduced by transformer. Finally, x1-4 is spliced and input into the final classification layer (KAN) to make predictions and provide symbolic formula: 0.32*x_1 + 1.61*x_2 − 0.42*x_3 + 0.23*x_4 − 0.23* tanh (8.13*x_1 + 1.14) + 0.54. The formula combines linear terms and hyperbolic tangent functions to capture the multilevel and nonlinear effects of the input variables (x_1, x_2, x_3, x_4) on the target variable (y). The formula can express complex variable relationships and reveal the mode of action of different factors in the risk of recurrent gastrointestinal bleeding. The flow chart of the KAN model is shown in Figure 4.

In the results of five-fold cross validation, the training performance of the fourth fold is particularly prominent, and the model achieves the best prediction accuracy on both the validation set (accuracy = 0.9750) and the test set (accuracy = 0.9615). In addition, by comparing the accuracy and loss curves in the training process, better fitting effect and more stable gradient descent trend can be observed, which further verifies the robustness of the training. The results of the five-fold cross validation and the results of the confusion matrix are shown in Figures 5–9 and Table 12.

In addition, by exploring the influence of infection on the model prediction, it was found that the prediction accuracy of rebleeding in the non-*H. pylori* infection group (0.8989) was lower than that in the *H. pylori* infection group (0.9636), and other evaluation parameters were also lower than that in the infection group. In addition, by adding irrelevant random noise to mask the influence of infection degree on model output, it is found that the model prediction accuracy (0.7992) is significantly reduced. The results of the comparison experiment are shown in Table 13.

4 Discussion

Based on four types of data (including vital signs, clinical information, blood indicators and stool detection) of UGIB patients infected with *Helicobacter pylori* (*H. pylori*), this study uses Transformer network as a feature extractor and KAN network as a classifier to build a deep learning prediction model. To predict the risk of rebleeding in UGIB patients during hospitalization, and compared with four machine learning

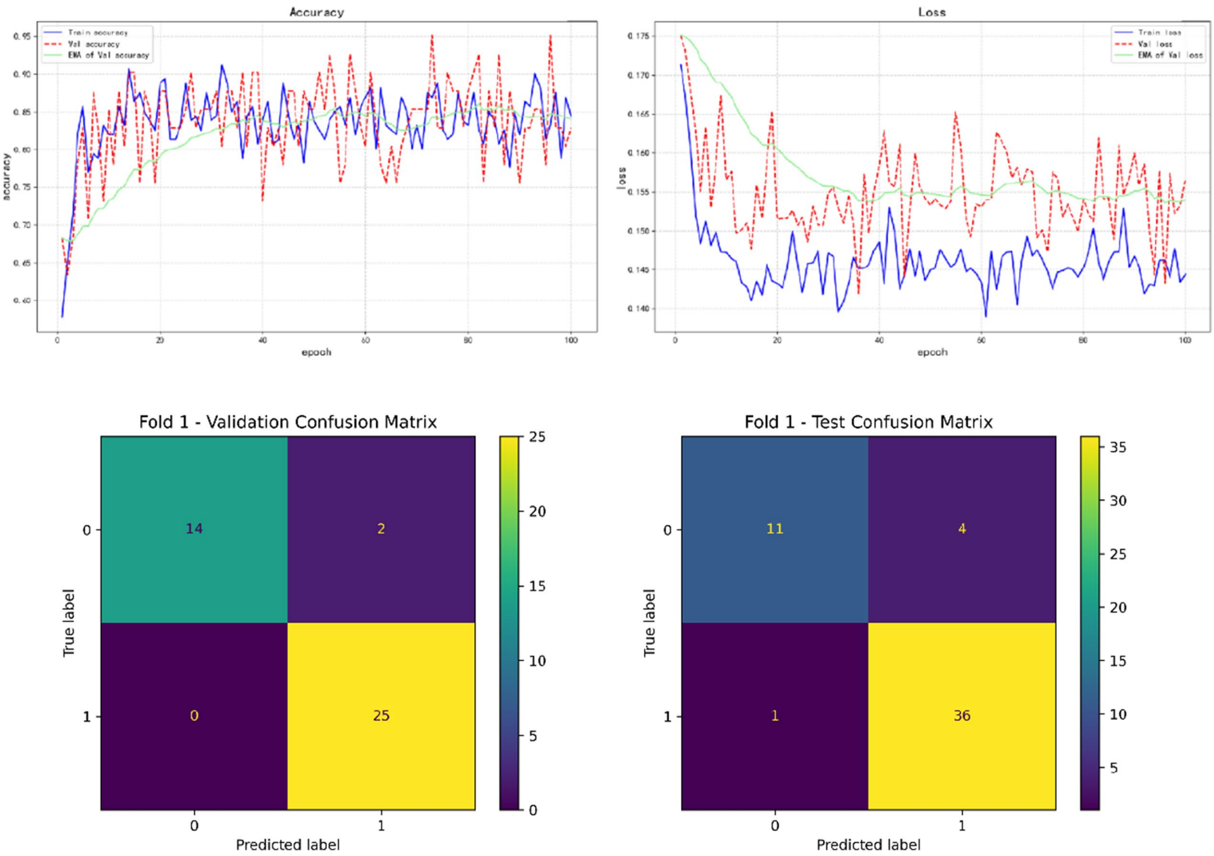


FIGURE 5
Fold 1 results and Validation confusion matrix and Test confusion matrix.

models, good results were obtained, and the accuracy of the final model reached 0.9615.

In a model that predicted the risk of gastrointestinal bleeding rebleeding, the effect of vital sign depth features was positive. Tari et al. (2023) found that UGIB patients with impaired hemodynamics were at increased risk for all associated adverse outcomes, such as higher rates of hospitalization and increased rates of re-bleeding within 30 days. In addition, UGIB patients with impaired hemodynamics require surgery more frequently. This result is consistent with the conclusion of the model in this study that abnormal vital signs are associated with an increased risk of rebleeding.

Secondly, although the clinical information depth feature presents a positive linear relationship, due to the hyperbolic tangent (tanh) nonlinear term contained in it, the information may lead to complex fluctuations under different circumstances. The depth characteristics of blood indicators showed a relatively small negative relationship. In most studies on the risk of UGIB rebleeding, different models analyzed different clinical information and blood markers and achieved better results, although the specific information and markers were different (Ungureanu et al., 2023; Uysal, 2021; Zhuang et al., 2023). However, in a recent study by Taylor et al. (2025), blood type was found to be associated with the risk of upper gastrointestinal bleeding, thrombosis and peptic ulcer disease. Blood type B is associated with a reduced risk of overall outcomes, including rebleeding, the need for surgery or embolization, and mortality, compared to non-B blood type. In part, this may explain that even when the same blood indicators or clinical information is used, their contribution to the prediction of rebleeding in patients may vary. Finally, fecal depth features showed a slight positive effect on the risk of recurrence, suggesting that fecal occult blood suggests the risk of rebleeding. Compared to deep learning models, the lower prediction accuracy of traditional machine learning models is due to their limited ability to handle highly nonlinear data or complex relationships, making it difficult to capture the intrinsic connections between various indicators. Deep learning, on the other hand, overcomes this limitation by seamlessly linking the internal relationships among all indicators, integrating them into a unified whole to accomplish tasks.

In the correlation between the risk of rebleeding and various indicators, we found that gender was related to the risk of rebleeding, that is, men were more likely to have rebleeding during hospitalization for upper gastrointestinal bleeding, which was also demonstrated in the study of Jeon et al. (2021), which found that gender could be used as a set of variables to predict rebleeding in UGIB patients. Similarly, the study results of Snipe et al. show that when women are in the follicular phase of the menstrual cycle, biological sex has no effect on intestinal epithelial injury and permeability, and has the least effect on gastrointestinal symptoms and the systemic cytokine spectrum in response to stress of exhaustion. However, the influence of males after exercise is greater than that of females (Snipe and Costa, 2018). Therefore, whether to keep calm after detecting the symptoms of upper gastrointestinal bleeding? Avoiding unnecessary exercise can reduce the risk of rebleeding, which is worth exploring. And in the

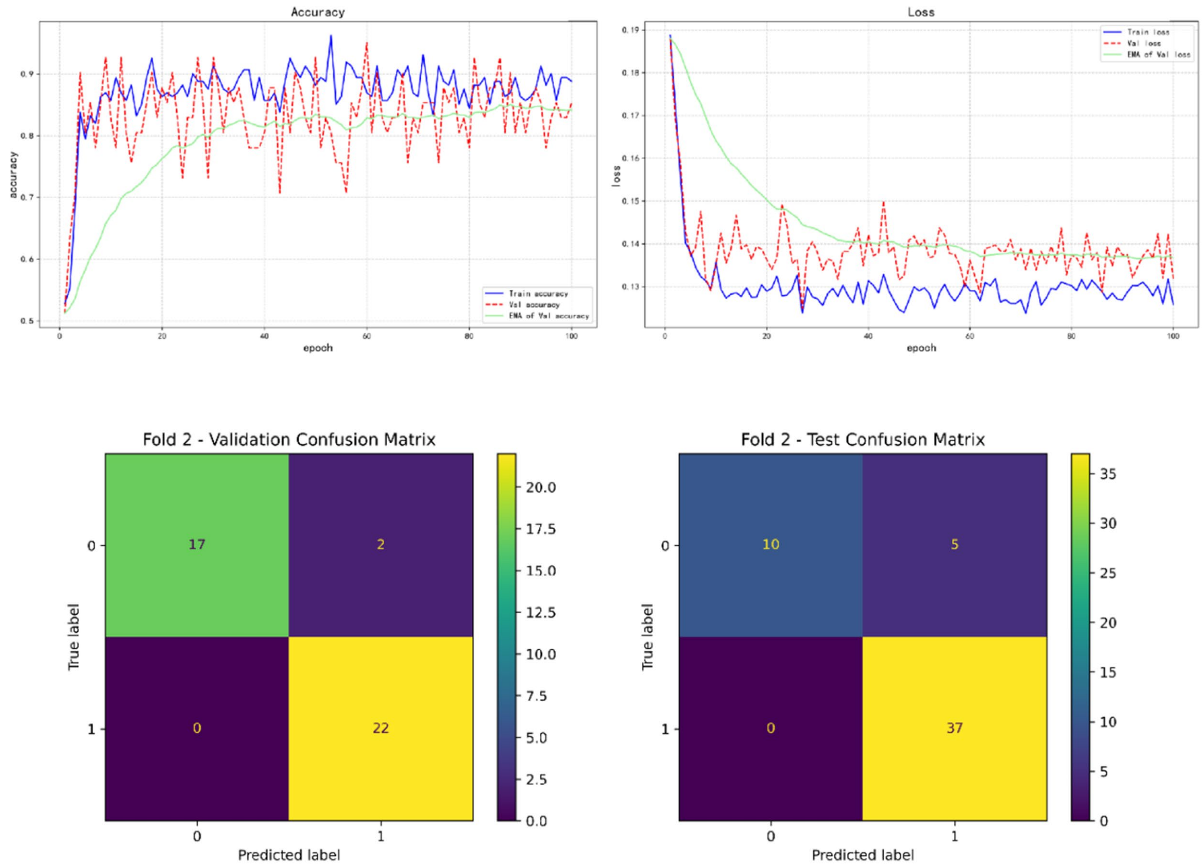


FIGURE 6
Fold 2 results and validation confusion matrix and test confusion matrix.

study by Zheng et al. (2022), it can be seen that although the incidence of UGIB has decreased in recent years, the age/sex-adjusted incidence of GIB increased from 378.4 per 100,000 people to 397.5 per 100,000 people between 2006 and 2019, so it is important to focus on the elderly. In particular, the incidence of UGIB in older men and the risk of rebleeding after hospitalization are necessary and meaningful behaviors. In the scoring system studied by Jeon et al. (2021), whether syncope is associated with mortality within 30 days after hospitalization in UGIB patients, syncope and rebleeding risk obtained in this experiment can supplement this experiment to some extent. Heyer et al. (2018) found that nausea associated with syncope was related to the change of gastric myoelectric activity and the increase of vasopressin and epinephrine in time, which could also explain that syncope patients were more likely to induce rebleeding to a certain extent, and the related mechanism of syncope and rebleeding could be further explored in future studies. In addition, in two reports on rare cases of upper gastrointestinal bleeding (Li et al., 2024; Ulbricht et al., 2022), syncope occurred in both patients. Although the internal correlation between syncope and rebleeding has not been actually explored, it can also suggest that current UGIB patients should pay attention to the clinical changes if syncope occurs. To provide clinical improvement measures as soon as possible. In the new UGIB treatment management, for all UGIB patients, it is recommended to resuscitate with intravenous infusion and red blood cell infusion according to the need, and the hemoglobin threshold is 70–80 g/L. When the hemoglobin of UGIB patients is lower than this threshold, resuscitation and red blood cell infusion therapy are performed (Stanley and Laine, 2019), which shows the influence of hemoglobin value on clinical treatment decision making. The new treatment management is based on the patient's current vital signs and does not investigate how the remaining risks during hospitalization are reflected in the hemoglobin value, however, a number of studies have been involved in the study of hemoglobin and the risk of rebleeding, and it is found that there is a certain correlation between hemoglobin and rebleeding. In addition, Extrat et al.'s study, hemoglobin level after arterial embolization in UGIB patients is more likely to reflect the early mortality and the risk of rebleeding in patients (Extrat et al., 2022; Tatlıparmak et al., 2022; Zheng et al., 2019). This has further defined the research direction for clinical research, which can be focused on in future studies. In addition, there is still some controversy about the hemoglobin threshold for red blood cell transfusion (Carson et al., 2021; Kola et al., 2021; Laine et al., 2021; Page et al., 2021; Teutsch et al., 2023), and the results of individual studies show that if the transfusion is started with a lower threshold, the incidence of transfusion reaction and post-transfusion intervention is lower. A hemoglobin threshold greater than 80 g/L may result in a higher rate of adverse outcomes. In conclusion, the correlation mechanism between hemoglobin value and clinical events during hospitalization should be explored, and then the optimal restrictive transfusion threshold should be further studied. As the

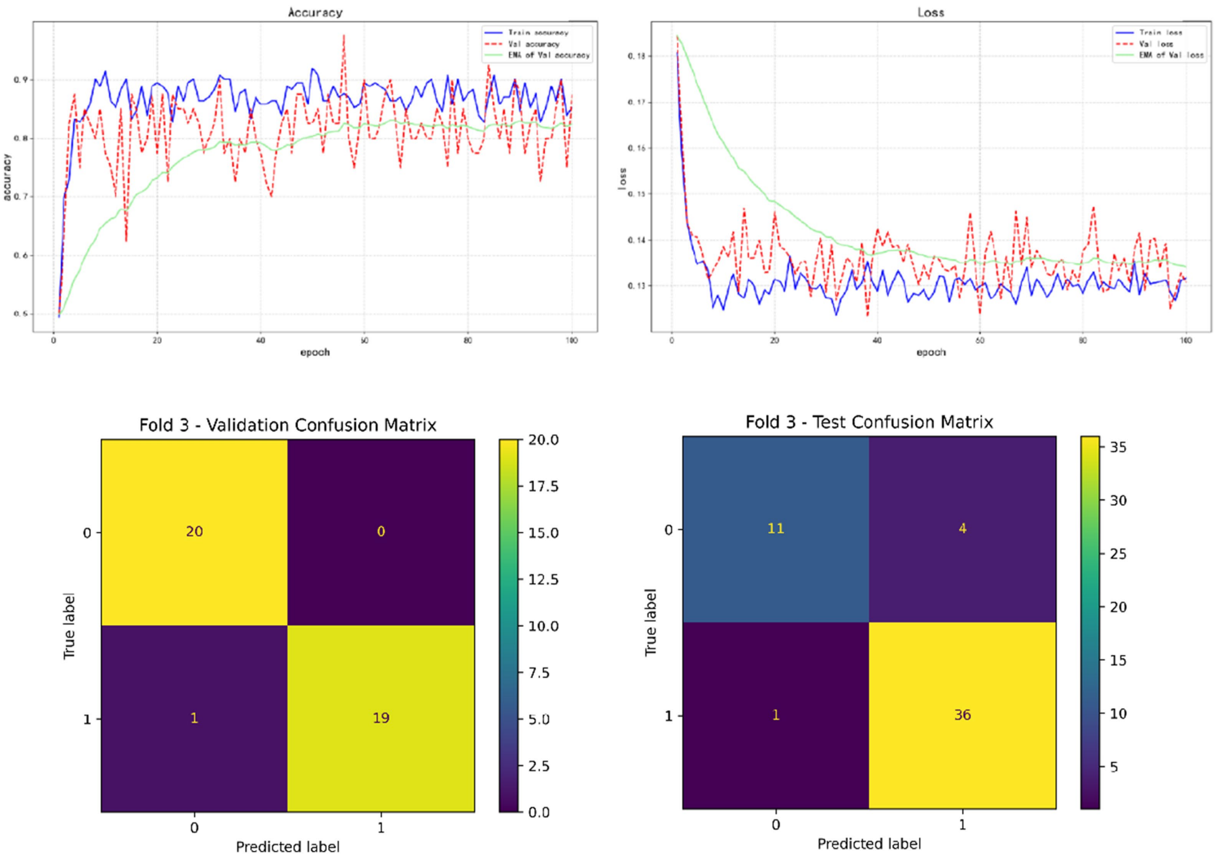


FIGURE 7
Fold 3 results and validation confusion matrix and test confusion matrix.

most direct sign of UGIB diagnosis, endoscopy is usually the most concerned result in clinic. However, limited by the scale and the use of instruments in some small and medium-sized hospitals, the current research focuses on the diagnosis of UGIB by skipping endoscopy or replacing endoscopy or predicting clinical outcome.

In addition, we wanted to explore whether the infection situation would affect the prediction efficiency of the deep learning model. We designed two comparative experiments: one was to predict the best trained model in the *H. pylori* infected group and the *H. pylori* non-infected group; the other was to evaluate the prediction efficiency of the best model by setting the degree of infection as irrelevant random noise. Although the correlation analysis did not find significant correlation between the degree of *H. pylori* infection and the risk of rebleeding. This may be due to the uneven distribution of data on the degree of certain infections or the fact that the infection affects the risk of rebleeding in a more complex non-linear relationship. However, compared with non-infected patients, it can be found from the results of the comparative experiment that the model has a higher prediction accuracy for rebleeding risk in infected patients with subdivided infection degree. Meanwhile, when the information of infection degree is ignored in the model, the prediction efficiency of the model is greatly reduced. Therefore, we believe that *H. pylori* infection plays a crucial role in the risk prediction of rebleeding, and *H. pylori* infection may affect the final prediction results in a non-linear manner by influencing other factors or complex combinations of variables. Although few studies have included the degree of *H. pylori* infection in the diagnosis and treatment of upper gastrointestinal bleeding (UGIB), there have been studies that have analyzed the correlation between the diagnosis of *H. pylori* infection and gastroscopy (EGD) results in other gastrointestinal diseases, such as gastric and duodenal ulcers. Attempts were made to replace gastroscopy to some extent by non-invasive detection of *H. pylori* infection (Liao et al., 2023). In the Pritchard DM study, the use of *H. pylori* infection testing as an alternative to gastroscopy and treatment was found to be the most cost-effective strategy (Pritchard et al., 2021). At the same time, *H. pylori* diagnostic methods have been increasingly improved, such as serological detection, fecal antigen detection and urea breath test, all of which are practical and highly sensitive (Ghazanfar et al., 2024). This study also explored the value of incorporating *H. pylori* infection in predicting the risk of rebleeding in patients with upper gastrointestinal hemorrhage during hospitalization.

At present, for the risk of rebleeding during hospitalization in UGIB patients, in addition to timely and accurate endoscopy, there are multiple scoring mechanisms. We found that the three types of traditional scoring are for different emergency time periods, and there are some problems in the detection efficiency, and the scoring also represents the risk in different time periods, and there is a certain lag or inaccuracy in clinical measures, but the four types of machine learning scoring methods can integrate all factors and examinations to evaluate the overall period of hospitalization of patients. Patients were monitored as a whole, but the accuracy of machine learning is not satisfactory. Meanwhile, deep learning models with higher testing

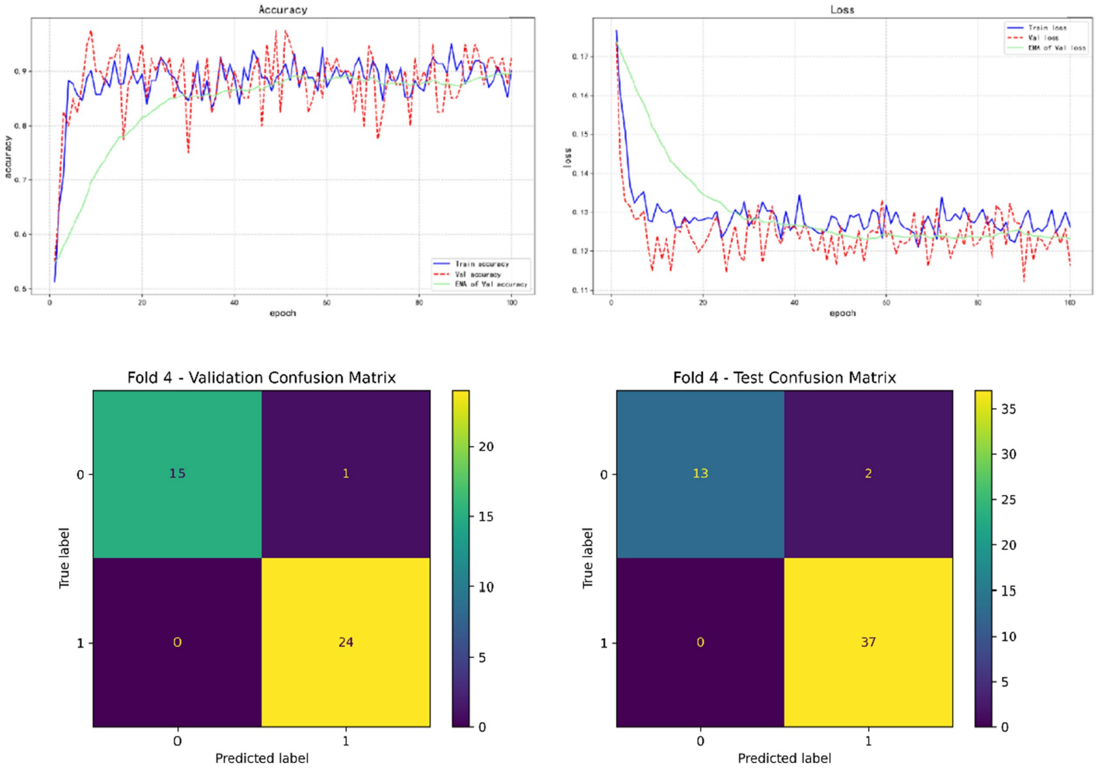


FIGURE 8
Fold 4 results and validation confusion matrix and test confusion matrix.

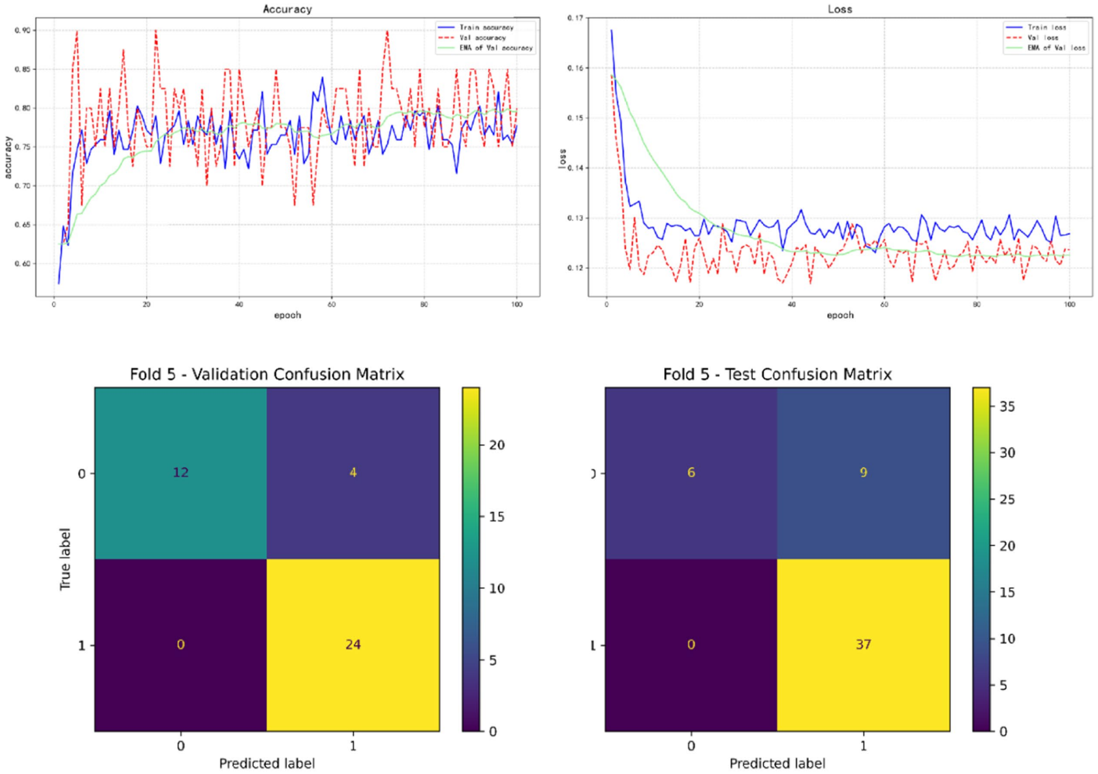


FIGURE 9
Fold 5 results and validation confusion matrix and test confusion matrix.

TABLE 12 Result of five-fold cross validation.

		Accuracy	Precision	Recall	F1-score
Fold 1	Validation set	0.9512	0.9259	1.0000	0.9615
	Test set	0.9038	0.9000	0.9730	0.9351
Fold 2	Validation set	0.9512	0.9167	1.0000	0.9565
	Test set	0.9038	0.8810	1.0000	0.9367
Fold 3	Validation set	0.9750	1.0000	0.9500	0.9744
	Test set	0.9038	0.9000	0.9730	0.9351
Fold 4	Validation set	0.9750	0.9600	1.0000	0.9796
	Test set	0.9615	0.9487	1.0000	0.9737
Fold 5	Validation set	0.9000	0.8571	1.0000	0.9231
	Test set	0.8269	0.8043	1.0000	0.8916

TABLE 13 The results of the comparison experiments.

	Accuracy	Precision	Recall	F1-score
Non-infection group	0.8989	0.9180	0.9333	0.9256
Infection group	0.9636	0.9565	0.9778	0.9670
Noised infection level group	0.7992	0.7765	0.9267	0.8450

efficiency provide stronger technical support for overall detection. For patients with UGIB, collecting relevant information based on deep learning models helps to make the fastest and most accurate judgments in the absence of endoscopy. However, this study is limited to a single center and lacks validation from multicenter experimental data, which will be an important direction for future research.

5 Conclusion

Based on the degree of *H. pylori* infection in patients with upper gastrointestinal bleeding, combined with a number of clinical laboratory tests and clinical data, we developed a clinical model for predicting the risk of rebleeding in patients with upper gastrointestinal bleeding. It provides an early prediction of rebleeding during a patient's hospitalization and optimizes early intervention for patients to a certain extent. It provides a more concise, convenient and effective guidance scheme for small and medium-sized hospitals to make clinical decisions for UGIB patients.

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 requirement of ethical approval was waived by Shengjing Hospital Ethics Committee for the studies involving humans because this experiment is a retrospective study, and the committee waived it. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because This experiment is a retrospective study, and the committee waived it.

Author contributions

WZ: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. ZL: Data curation, Resources, Validation, Writing – original draft, Supervision. YZ: Data curation, Investigation, Writing – original draft. TJ: Methodology, Project administration, Writing – review & editing. XZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research 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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

Alali, A. A., Boustany, A., Martel, M., and Barkun, A. N. (2023). Strengths and limitations of risk stratification tools for patients with upper gastrointestinal bleeding: a narrative review. *Expert Rev. Gastroenterol. Hepatol.* 17, 795–803. doi: 10.1080/17474124.2023.2242252

Blatchford, O., Murray, W. R., and Blatchford, M. (2000). A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. *Lancet* 356, 1318–1321. doi: 10.1016/S0140-6736(00)02816-6

Carson, J. L., Stanworth, S. J., Dennis, J. A., Trivella, M., Roubinian, N., Fergusson, D. A., et al. (2021). Transfusion thresholds for guiding red blood cell transfusion. *Cochrane Database Syst. Rev.* 2022:CD002042. doi: 10.1002/14651858.CD002042.pub5

Cazacu, S. M., Alexandru, D. O., Statie, R. C., Iordache, S., Ungureanu, B. S., Iovănescu, V. F., et al. (2023). The accuracy of pre-endoscopic scores for mortality prediction in patients with upper GI bleeding and no endoscopy performed. *Diagnostics* 13:1188. doi: 10.3390/diagnostics13061188

Chen, L., Zheng, H., and Wang, S. (2021). Prediction model of emergency mortality risk in patients with acute upper gastrointestinal bleeding: a retrospective study. *PeerJ* 9:e11656. doi: 10.7717/peerj.11656

Dong, Z., Wu, L., Mu, G., Zhou, W., Li, Y., Shi, Z., et al. (2022). A deep learning-based system for real-time image reporting during esophagogastroduodenoscopy: a multicenter study. *Endoscopy* 54, 771–777. doi: 10.1055/a-1731-9535

Extrat, C., Grange, S., Mayaud, A., Villeneuve, L., Chevalier, C., Williet, N., et al. (2022). Transarterial embolization for active gastrointestinal bleeding: predictors of early mortality and early Rebleeding. *J Personal Med* 12:1856. doi: 10.3390/jpm12111856

Ghazanfar, H., Javed, N., Reina, R., Thartori, O., Ghazanfar, A., and Patel, H. (2024). Advances in diagnostic modalities for *Helicobacter pylori* infection. *Life* 14:1170. doi: 10.3390/life14091170

Gralnek, I. M., Stanley, A. J., Morris, A. J., Camus, M., Lau, J., Lanas, A., et al. (2021). Endoscopic diagnosis and management of nonvariceal upper gastrointestinal hemorrhage (NVUGIH): European Society of Gastrointestinal Endoscopy (ESGE) guideline—update 2021. *Endoscopy* 53, 300–332. doi: 10.1055/a-1369-5274

Gu, S. X., Siddon, A. J., Huntington, S. F., and Jain, D. (2023). *Helicobacter pylori*-negative mucosa-associated lymphoid tissue (MALT) lymphoma of the stomach: a clinicopathologic analysis. *Am. J. Clin. Pathol.* 160, 612–619. doi: 10.1093/ajcp/aqad088

Heyer, G. L., Boles, L. H., Harvey, R. A., and Cismowski, M. J. (2018). Gastric myoelectrical and neurohormonal changes associated with nausea during tilt-induced syncope. *Neurogastroenterol Motil* 30:e13220. doi: 10.1111/nmo.13220

Jeon, H. J., Moon, H. S., Kwon, I. S., Kang, S. H., Sung, J. K., and Jeong, H. Y. (2021). Which scoring system should be used for non-variceal upper gastrointestinal bleeding? Old or new? *J. Gastroenterol. Hepatol.* 36, 2819–2827. doi: 10.1111/jgh.15555

Kamboj, A. K., Hoversten, P., and Leggett, C. L. (2019). Upper gastrointestinal bleeding: etiologies and management. *Mayo Clin. Proc.* 94, 697–703. doi: 10.1016/j.mayocp.2019.01.022

Karstensen, J. G., Ebigbo, A., Bhat, P., Dinis-Ribeiro, M., Gralnek, I., Guy, C., et al. (2020). Endoscopic treatment of variceal upper gastrointestinal bleeding: European Society of Gastrointestinal Endoscopy (ESGE) Cascade guideline. *Endosc Int Open* 8, E990–E997. doi: 10.1055/a-1187-1154

Kim, J. E., Choi, Y. H., Lee, Y. C., Seong, G., Song, J. H., Kim, T. J., et al. (2023). Deep learning model for distinguishing Mayo endoscopic subscore 0 and 1 in patients with ulcerative colitis. *Sci. Rep.* 13:11351. doi: 10.1038/s41598-023-38206-6

Koetzier, L. R., Mastrodicasa, D., Szczykutowicz, T. P., van der Werf, N. R., Wang, A. S., Sandfort, V., et al. (2023). Deep learning image reconstruction for CT: technical principles and clinical prospects. *Radiology* 306:e221257. doi: 10.1148/radiol.221257

Kola, G., Sureshkumar, S., Mohsina, S., Sreenath, G. S., and Kate, V. (2021). Restrictive versus liberal transfusion strategy in upper gastrointestinal bleeding: a randomized controlled trial. *Saudi J. Gastroenterol.* 27, 13–19. doi: 10.4103/sjg.SJG_152_20

Laine, L., Barkun, A. N., Saltzman, J. R., Martel, M., and Leontiadis, G. I. (2021). ACG clinical guideline: upper gastrointestinal and ulcer bleeding. *Am. J. Gastroenterol.* 116, 899–917. doi: 10.14309/ajg.0000000000001245

Laine, L., Yang, H., Chang, S. C., and Datto, C. (2012). Trends for incidence of hospitalization and death due to GI complications in the United States from 2001 to 2009. *Am. J. Gastroenterol.* 107, 1190–1195. doi: 10.1038/ajg.2012.168

Lanas, A., García-Rodríguez, L. A., Polo-Tomás, M., Ponce, M., Alonso-Abreu, I., Perez-Aisa, M. A., et al. (2009). Time trends and impact of upper and lower gastrointestinal bleeding and perforation in clinical practice. *Am. J. Gastroenterol.* 104, 1633–1641. doi: 10.1038/ajg.2009.164

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, M., Chen, C., Deng, D., Ni, X., Yin, J., Chen, L., et al. (2024). An unusual cause of life-threatening upper gastrointestinal bleeding. *Rev. Esp. Enferm. Dig.* 116, 54–55. doi: 10.17235/reed.2023.9645/2023

Liao, E.-C., Yu, C.-H., Lai, J.-H., Lin, C. C., Chen, C. J., Chang, W. H., et al. (2023). A pilot study of non-invasive diagnostic tools to detect *Helicobacter pylori* infection and peptic ulcer disease. *Sci. Rep.* 13:22800. doi: 10.1038/s41598-023-50266-2

Matsuhashi, T., Hatta, W., Hikichi, T., Fukuda, S., Mikami, T., Tatsuta, T., et al. (2021). A simple prediction score for in-hospital mortality in patients with nonvariceal upper gastrointestinal bleeding. *J. Gastroenterol.* 56, 758–768. doi: 10.1007/s00535-021-01797-w

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: 10.1136/bmj.n71

Popa, D. G., Obleagă, C. V., Socea, B., Serban, D., Ciurea, M. E., Diaconescu, M., et al. (2021). Role of *Helicobacter pylori* in the triggering and evolution of hemorrhagic gastro-duodenal lesions. *Exp. Ther. Med.* 22:1147. doi: 10.3892/etm.2021.10582

Pritchard, D. M., Bornschein, J., Beales, I., Beresniak, A., Salhi, H., and Malfertheiner, P. (2021). Cost-effectiveness modelling of use of urea breath test for the management of *Helicobacter pylori*-related dyspepsia and peptic ulcer in the UK. *BMJ Open Gastroenterol.* 8:e000685. doi: 10.1136/bmjgast-2021-000685

Redondo-Cerezo, E., Vadillo-Calles, F., Stanley, A. J., Laursen, S., Laine, L., Dalton, H. R., et al. (2020). MAP(ASH): a new scoring system for the prediction of intervention and mortality in upper gastrointestinal bleeding. *J. Gastroenterol. Hepatol.* 35, 82–89. doi: 10.1111/jgh.14811

Rockall, T. A., Logan, R. F., Devlin, H. B., and Northfield, T. C. (1996). Risk assessment after acute upper gastrointestinal haemorrhage. *Gut* 38, 316–321. doi: 10.1136/gut.38.3.316

Saltzman, J. R., Tabak, Y. P., Hyett, B. H., Sun, X., Travis, A. C., and Johannes, R. S. (2011). A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. *Gastrointest. Endosc.* 74, 1215–1224. doi: 10.1016/j.gie.2011.06.024

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Seo, D.-W., Yi, H., Park, B., Kim, Y. J., Jung, D. H., Woo, I., et al. (2020). Prediction of adverse events in stable non-variceal gastrointestinal bleeding using machine learning. *J. Clin. Med.* 9:2603. doi: 10.3390/jcm9082603

Shung, D. L., Au, B., Taylor, R. A., Tay, J. K., Laursen, S. B., Stanley, A. J., et al. (2020). Validation of a machine learning model that outperforms clinical risk scoring Systems for Upper Gastrointestinal Bleeding. *Gastroenterology* 158, 160–167. doi: 10.1053/j.gastro.2019.09.009

Shung, D. L., and Laine, L. (2024). Review article: upper gastrointestinal bleeding—review of current evidence and implications for management. *Aliment. Pharmacol. Ther.* 59, 1062–1081. doi: 10.1111/apt.17949

Snipe, R. M. J., and Costa, R. J. S. (2018). Does biological sex impact intestinal epithelial injury, small intestine permeability, gastrointestinal symptoms and systemic cytokine profile in response to exertional-heat stress? *J. Sports Sci.* 36, 2827–2835. doi: 10.1080/02640414.2018.1478612

Stanley, A. J., and Laine, L. (2019). Management of acute upper gastrointestinal bleeding. *BMJ* 364:l536. doi: 10.1136/bmj.l536

Stanley, A. J., Laine, L., Dalton, H. R., Schultz, M., Ngu, J. H., Abazi, R., et al. (2017). Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding: international multicentre prospective study. *BMJ* 356:i6432. doi: 10.1136/bmj.i6432

Tari, E., Frim, L., Stolcz, T., Teutsch, B., Veres, D. S., Hegyi, P., et al. (2023). At admission hemodynamic instability is associated with increased mortality and rebleeding rate in acute gastrointestinal bleeding: a systematic review and meta-analysis. *Ther. Adv. Gastroenterol.* 16:17562848231190970. doi: 10.1177/17562848231190970

Tatlıparmak, A. C., Dikme, Ö., Dikme, Ö., and Topaçoğlu, H. (2022). Cancer, platelet distribution width, and total protein levels as predictors of rebleeding in upper gastrointestinal bleeding. *PeerJ* 10:e14061. doi: 10.7717/peerj.14061

Taylor, S., Chen, L., Dutt, K., Boyapati, R., Ye, B., Ong, S., et al. (2025). Blood group B may reduce risk of Rebleeding in patients with upper gastrointestinal Haemorrhage due to peptic ulcer disease. *Aliment. Pharmacol. Ther.* 61, 706–711. doi: 10.1111/apt.18485

Teutsch, B., Veres, D. S., Pálinkás, D., Simon, O. A., Hegyi, P., and Erőss, B. (2023). Potential benefits of restrictive transfusion in upper gastrointestinal bleeding: a systematic review and meta-analysis of randomised controlled trials. *Sci. Rep.* 13:17301. doi: 10.1038/s41598-023-44271-8

Thiebaud, P.-C., Yordanov, Y., Galimard, J.-E., Raynal, P. A., Beaune, S., Jacquin, L., et al. (2017). Management of upper gastrointestinal bleeding in emergency departments, from bleeding symptoms to diagnosis: a prospective, multicenter, observational study. *Scand. J. Trauma Resusc. Emerg. Med.* 25:78. doi: 10.1186/s13049-017-0425-6

Toews, I., Hussain, S., Nyirenda, J. L. Z., Willis, M. A., Kantorová, L., Slezáková, S., et al. (2024). Pharmacological interventions for preventing upper gastrointestinal bleeding in people admitted to intensive care units: a network meta-analysis. *BMJ Evid Based Med* 30, 22–35. doi: 10.1136/bmjebm-2024-112886

Ulbricht, J., Madea, B., and Doberentz, E. (2022). Case report: fatal bleeding from a duodenal ulcer-Dieulafoy's lesion? *Int. J. Legal Med.* 136, 203–208. doi: 10.1007/s00414-021-02721-w

Ungureanu, B. S., Gheonea, D. I., Florescu, D. N., Iordache, S., Cazacu, S. M., Iovanescu, V. F., et al. (2023). Predicting mortality in patients with nonvariceal upper gastrointestinal bleeding using machine-learning. *Front. Med.* 10:1134835. doi: 10.3389/fmed.2023.1134835

Uysal, E. (2021). Features of patients with upper gastrointestinal bleeding and factors affecting the re-bleeding risk. *Turk J Trauma Emerg Surg* 28, 147–154. doi: 10.14744/tjtes.2021.00670

Veisman, I., Oppenheim, A., Maman, R., Kofman, N., Edri, I., Dar, L., et al. (2022). A novel prediction tool for endoscopic intervention in patients with acute upper gastrointestinal bleeding. *J. Clin. Med.* 11:5893. doi: 10.3390/jcm11195893

Zhang, L., Lu, Z., Yao, L., Dong, Z., Zhou, W., He, C., et al. (2023). Effect of a deep learning-based automatic upper GI endoscopic reporting system: a randomized crossover study (with video). *Gastrointest. Endosc.* 98, 181–190.e10. doi: 10.1016/j.gie.2023.02.025

Zheng, W., Jiang, L., Jia, X., Long, G., Shu, X., and Jiang, M. (2019). Analysis of risk factors and development of scoring system to predict severity of upper gastrointestinal bleeding in children. *J. Gastroenterol. Hepatol.* 34, 1035–1041. doi: 10.1111/jgh.14548

Zheng, N. S., Tsay, C., Laine, L., and Shung, D. L. (2022). Trends in characteristics, management, and outcomes of patients presenting with gastrointestinal bleeding to emergency departments in the United States from 2006 to 2019. *Aliment. Pharmacol. Ther.* 56, 1543–1555. doi: 10.1111/apt.17238

Zhuang, H., Bao, A., Tan, Y., Wang, H., Xie, Q., Qiu, M., et al. (2022). Application and prospect of artificial intelligence in digestive endoscopy. *Expert Rev. Gastroenterol. Hepatol.* 16, 21–31. doi: 10.1080/17474124.2022.2020646

Zhuang, Y., Xia, S., Chen, J., Ke, J., Lin, S., Lin, Q., et al. (2023). Construction of a prediction model for rebleeding in patients with acute upper gastrointestinal bleeding. *Eur. J. Med. Res.* 28:351. doi: 10.1186/s40001-023-01349-3