CONNECTED HEALTH: STATUS AND TRENDS

EDITED BY: Constantinos S. Pattichis, Andreas S. Panayides and Chris Nugent
PUBLISHED IN: Frontiers in Digital Health

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-853-5
DOI 10.3389/978-2-88971-853-5

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

CONNECTED HEALTH: STATUS AND TRENDS

Topic Editors:
Constantinos S. Pattichis, University of Cyprus, Cyprus
Andreas S. Panayides, University of Cyprus, Cyprus
Chris Nugent, Ulster University, United Kingdom

Citation: Pattichis, C. S., Panayides, A. S., Nugent, C., eds. (2021). Connected Health: Status and Trends. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-853-5

Table of Contents

frontiers in Digital Health

EDITORIAL
published: 20 October 2021
doi: 10.3389/fdgth.2021.762270

Editorial: Connected Health: Status and Trends

Constantinos S. Pattichis[1,2]*, *Andreas S. Panayides*[1] *and Chris Nugent*[3]

[1] *Department of Computer Science and Biomedical Engineering Research Centre, University of Cyprus, Nicosia, Cyprus,* [2] *CYENS Center of Excellence, Nicosia, Cyprus,* [3] *School of Computing, Ulster University, Coleraine, United Kingdom*

Keywords: connected health, digital health, eHealth, mHealth, telemedicine

Editorial on the Research Topic

Connected Health: Status and Trends

INTRODUCTION

Recent advances in information and communication technologies prescribe an emerging paradigm in the delivery of advanced healthcare services named connected health. The scope of this special issue, is fully aligned with the Frontiers in Digital Health, section on Connected Health (1). The aim of this special issue is to present selected papers describing recent advances in connected health systems, platforms and solutions targeted to the development of efficient and effective interventions toward the provision of healthcare services for the benefit of the citizen. Topics covered include, sensing and Internet of Things (IoT) in healthcare systems, personalized and well-being systems, early diagnostics and clinical diagnostics, electronic health records and integrated care, and security and data protection. The manuscripts cover the aforementioned technologies toward the development of citizen-centric personalized eHealth, mHealth, pHealth and uHealth solutions, in addition to telemedicine systems.

OPEN ACCESS

Edited and reviewed by:
Timo Jämsä,
University of Oulu, Finland

***Correspondence:**
Constantinos S. Pattichis
pattichi@ucy.ac.cy

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *21 August 2021*
Accepted: *16 September 2021*
Published: *20 October 2021*

Citation:
Pattichis CS, Panayides AS and Nugent C (2021) Editorial: Connected Health: Status and Trends. Front. Digit. Health 3:762270. doi: 10.3389/fdgth.2021.762270

A BRIEF OVERVIEW OF THE PAPERS IN THE SPECIAL ISSUE

The papers in this special issue can be grouped under the following three categories that are briefly presented in this section: patient engagement via sensing, monitoring and coaching systems; emerging disease biomarkers; and eHealth eco system enabling technologies.

Patient Engagement *via* Sensing, Monitoring, and Coaching Systems

The first paper by Androutsou et al. presents a mobile application that is part of a multifactorial intervention to support the aging with balance disorders. This application enables the users to self-evaluate their activity and progress, to communicate with each other and to engage with the system with undiminished interest for a long period of time. The application is expected to undergo an evaluation in four pilot studies with 160 participants.

The second paper by Schütz et al. introduces a calibration framework of passive infrared motion sensors that covers the quantification of simultaneously acquired data from wearable accelerometers and use the data to find a suitable correlation between in-home and out-of-home physical activity. The proposed framework was evaluated by 20 community-dwelling older adults for at least 60 days. It was found that passive infrared based wireless sensor systems can be calibrated to give largely better estimates of older adults' daily physical activity using data over 7–14 days only.

The paper by Ntracha et al. introduces smartphone interaction behavioral data, unobtrusively captured as an aid in the screening and monitoring of Mild Cognitive Impairment (MCI). Digital biomarkers drawn from Fine Motor Impairment and Spontaneous Written Speech related data analysis were investigated on matched groups of 11 MCI patients and 12 healthy controls, over a time span of 6 months. The results demonstrate the potential of the proposed biomarkers to detect early stages of cognitive decline.

The fifth paper in this subsection by Tsiouris et al. reviews 41 papers that were published in the last decade covering the topic of virtual coaching systems targeted to enhance healthcare interventions. The findings suggest that home coaching systems were mainly focused on physical activity and a healthier living based on IoT devices and sensor monitoring. The devices that were used were mostly activity trackers, pedometers and heart rate monitoring. It is documented that real-time performance evaluation and personalized feedback was found to be rather lacking. Future trends in home coaching systems target to close the loop with real-time automated performance evaluations, monitoring, feedback, and more advanced interventions.

Emerging Disease Biomarkers

Juvenile idiopathic arthritis (JIA) is the most common rheumatic disease of childhood. It is very difficult to assess JIA due to its highly variable presentation and the documentation of few reliable biomarkers to assess it. In the paper by Whittingslow et al. the joint acoustic emissions (JAEs) from the knees of children with JIA are quantified and proposed as a new biomarker for the non-invasive assessment of the disease. JAEs from 25 patients with JIA were evaluated with very promising findings suggesting the use of the proposed technique in clinical practice.

Stanitsas et al. propose a system that provides useful features that describe malignant regions in cancerous tissue. A Covariance-Kernel based patch descriptor is introduced that has the capability of describing tissue carcinoma. The proposed methodology was evaluated on a breast cancer dataset with very promising results.

Ehealth Eco System Enabling Technologies

The paper by Martin et al. proposes an immersive environment to track behaviors relevant to neuropsychiatric symptomatology. The proposed framework facilitates connected tele-psychiatry, providing quantitative effective assessment, and thus overcoming subjective symptomatology analysis that is the current practice.

The study by Christoforou et al. provides an overview of assistive robotics in nursing, summarizing the benefits of the proposed solutions in clinical practice. Moreover, the paper presents the end-users' perspective and identifies challenges, limitations and future directions of the robotic solutions.

The paper by Smith et al. reports the 15 years' experience (2000–2016) of the Queensland Telepaediatric Service (QTS) that was established in Brisbane, Australia. QTS was developed to support telehealth services in remote locations. A total of 23,054 telehealth consultations were delivered for 37 pediatric clinical specialties. The most common services covered the following: child and youth mental health, neurology, burns care, surgery, and ear nose and throat services. Most of the health services involved hospital video consultations for the delivery of the services.

A citizen-centric framework which enables EHR system integration with biobanks is proposed in the paper by Antoniades et al. The proposed framework enables retrospective lifelong prospective longitudinal studies, adhering strictly to legal and ethical requirements. Citizens would benefit through the proposed system, supporting them to make informed decisions and exercising their rights related to the use of their data.

Finally, the last paper in this special issue by Spanakis et al. aims to provide a framework of new technologies for the implementation of a secure information sharing platform for health data. The methodology presented in this paper documents strategies for information sharing, high-level requirements for the transfer of data between health-care organizations, technologies to support secure interconnectivity and trust, standards, guidelines, and interoperability specifications. Moreover, the use of cloud computing in sharing health data is covered, including also more advanced solutions such as block chain.

CONCLUDING REMARKS

The coronavirus disease (COVID-19) pandemic increases the need and demand for ongoing connected health and digital health interventions, solutions and tools (2). Virtual tele consultations have become the preferred mode of operation, rapidly replacing traditional face-to-face consultations given the increased risk of infection (3). Furthermore, in a recently published paper by Ding et al. (4), the following connected health enabling technologies and systems for handling the COVID-19 crisis are documented: "(i) wearable devices for monitoring populations at risk and those in quarantine that are needed for evaluating the health status of caregivers and for triggering triage processes for admission to hospitals; (ii) unobtrusive sensing systems for detecting the disease and for monitoring patients with mild symptoms whose clinical situation could suddenly worsen and trigger also admission to hospitals; and (iii) telehealth technologies for the remote monitoring and diagnosis of COVID-19 and related diseases."

The papers presented in this special issue provide connected health solutions and tools for minimizing the thread of the pandemic. Moreover, the papers presented provide a snapshot of recent advances in connected health, hoping to drive and accelerate the development, translation, and application of connected health into clinical practice toward improved disease management and treatment at the point of care, reduced hospitalization, offering a better quality of life to the citizen, at reduced cost. There is no doubt that connected health and digital health will continue to transform healthcare practice and services for the benefit of the citizen.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

CP acknowledges partial funding from the CYENS Centre of Excellence via the European Union's Horizon 2020 research and innovation programme under grant agreement No. 739578 and the funding received from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy. Invest Northern Ireland is acknowledged for partially supporting this work under the Competence Centre Programs Grant RD0513853 Connected Health Innovation Centre.

ACKNOWLEDGMENTS

The Guest Editors would like to thank all the contributors of the special issue and all reviewers for their thoughtful and valuable comments.

REFERENCES

1. Pattichis CS and Panayides AS. Connected health. *Front Digit Health.* (2019) 1:1. doi: 10.3389/fdgth.2019.00001
2. Cummins N and Schuller BW. Five crucial challenges in digital health. *Front Digit Health.* (2020) 2:536203. doi: 10.3389/fdgth.2020.536203
3. Webster P. Virtual health care in the era of COVID-19. *Lancet.* (2020) 395:1180–1. doi: 10.1016/S0140-6736(20)30818-7
4. Ding X, Clifton D, Ji N, Lovell NH, Bonato P, Chen W, et al. Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic. *IEEE Rev Biomed Eng.* (2020) 14:48–70. doi: 10.1109/RBME.2020.2992838

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 08 September 2020
doi: 10.3389/fdgth.2020.00015

A Smartphone Application Designed to Engage the Elderly in Home-Based Rehabilitation

Thelma Androutsou[1], Ioannis Kouris[1], Athanasios Anastasiou[1], Sotiris Pavlopoulos[1], Fariba Mostajeran[2], Doris-Eva Bamiou[3], Gregory J. Genna[3],Sergi G. Costafreda[4] and Dimitrios Koutsouris[1]*

[1] Biomedical Engineering Laboratory, National Technical University of Athens, School of Electrical and Computer Engineering, Athens, Greece, [2] Department of Informatics, Human-Computer Interaction, University of Hamburg, Hamburg, Germany, [3] University College London, UCL Ear Institute and UCLH Biomedical Research Centre, National Institute for Health Research, London, United Kingdom, [4] Division of Psychiatry, University College London, London, United Kingdom

As life expectancy increases, it is imperative that the elderly take advantage of the benefits of technology to remain active and independent. Mobile health applications are widely used nowadays as they promote a healthy lifestyle and self-management of diseases, opening new horizons in the interactive health service delivery. However, adapting these applications to the needs and requirements of the elderly is still a challenge. This article presents a smartphone application that is part of a multifactorial intervention to support older people with balance disorders. The application aims to enable users to self-evaluate their activity and progress, to communicate with each other and, through strategically selected motivational features, to engage with the system with undiminished interest for a long period of time. Mock-up interfaces were evaluated in semi-structured focus groups and interviews that were performed across three European countries. Further evaluation in the form of four pilot studies with 160 participants will be performed and qualitative and quantitative measures will be used to process the feedback about the use of the application.

Keywords: mHealth, smartphone, motivation, elderly, rehabilitation, user experience

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Lucia Rita Quitadamo, Aston University, United Kingdom
Mounir Mokhtari, Institut Mines-Télécom, France

****Correspondence:***
Thelma Androutsou
tandroutsou@biomed.ntua.gr

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *19 May 2020*
Accepted: *03 August 2020*
Published: *08 September 2020*

Citation:
Androutsou T, Kouris I, Anastasiou A, Pavlopoulos S, Mostajeran F, Bamiou D-E, Genna GJ, Costafreda SG and Koutsouris D (2020) A Smartphone Application Designed to Engage the Elderly in Home-Based Rehabilitation. Front. Digit. Health 2:15. doi: 10.3389/fdgth.2020.00015

INTRODUCTION

Advances in medicine and pharmacology, improvement in the quality of life and decline in mortality rates have led to people living longer and being more active than in the past. The population of people over 60 years old is rapidly growing in developed countries and thus a number of social and economic changes need to be addressed in the direction of inclusive societies. In order to maintain social cohesion and to balance the burden on health and care systems, this trend of aging population highlights the need for simultaneous extension of active aging. Elderly people should be enabled to be economically and socially independent and their needs and preferences should be taken into account in the process of designing and offering products and services.

Toward this direction, Information and Communications Technology (ICT) can play a major role in increasing the quality of life of the elders by solving the gap between their wishes and their needs. "Gerontechnology" is defined as the study of technology for ensuring good health, full social participation, and independent living throughout the entire life span, as long as it may extend (1, 2). It includes applications aimed at improving the quality of life, social interactions, medical and

psychological support, safety, and cognitive education of the elderly. Among its most dominant sectors is that of mobile communication, due to the rapid development of the technologies and capabilities of mobile devices, as well as the convenience the latter offer to people's daily lives. Mobile health (mHealth) applications, whether referring to stand-alone solutions or parts of an intervention, offer new capabilities and opportunities within health service delivery to patients and clinicians. mHealth solutions oriented to healthy lifestyle and self-management of diseases has attracted much of the research interest in recent years (3).

Despite the fact that, compared to younger people, older adults are much less familiar with technology, research has shown that they are nowadays much more digitally aware than they were in the past. They prefer to use a smartphone than other mobile devices and technological tools and the main functions they are interested in are the capability to communicate and get informed (4). They are also eager to learn how to use smartphones, especially when it comes to caring for their health (5). However, seniors' access to smartphone applications and assistive technology remains a challenge. Unfamiliarity and lack of appropriate training are undoubtedly some of the main obstacles. Nevertheless, an equally important factor in the low rate of technology adoption by the elderly is the exclusion of their needs and requirements when designing and implementing technological solutions (6).

Physical activity plays an important role in maintaining good physical and mental health and its promotion is especially important in the elderly, as it is the most sedentary population subgroup (7). There is a large number of mHealth applications aimed to stimulate physical activity, but most of them are not focused on older people. Advances in technology and capabilities, as well as greater accessibility to the benefits it offers, have led to studies directing to the needs of this target group. In (8) a training app that runs on a tablet and assists, monitors and motivates older people to follow personalized training plans autonomously at home is presented. The application includes both individual and social motivation techniques as well as a virtual training plan community. STARFISH is another smartphone-based application that enables the older users to accurately self-monitor their physical activity (9). Goal setting, action planning, feedback and social support are the main features that have been used in its context and showed potential to increase user acceptance. A smartphone and a heart rate belt were used as a mHealth system in order to monitor and promote elderly peoples' daily activity in a care home setting (10). The participants of this 10-week intervention increased their physical fitness levels. Gamification techniques in mobile applications can also play a quite beneficial role in promoting health in older adult populations. A game that embeds exercise and health education into a game context familiar to older adults was developed and the results of its utilization showed that it can engender high levels of adherence (11). In (12), a personalized behavioral intervention program based on theoretical constructs from the self-efficacy theory was developed. The program included personalized physical activity training, real-time physical activity self-monitoring, interactive prompts, and feedback with a smartwatch, phone consultation with an exercise trainer and research team members, and weekly financial incentives for achieving weekly physical activity goals. The SmartWalk project, on the other hand, focuses on creating a physical activity monitoring system for smart cities, based on the use of mHealth, where older people are motivated toward a physical active lifestyle and health care professionals can supervise them for best results.

This article introduces a smartphone application to support older people with balance disorders. The application was developed as part of the project HOLOBALANCE (HOLOgrams for personalized virtual coaching and motivation in an aging population with BALANCE disorders)[1], which aims to deliver a radically new cost-effective virtual coach to improve balance, cognition and physical activity. The main goal of the smartphone application is to promote physical activity and the user's engagement in the use of the HOLOBALANCE system. Design and motivation techniques have been incorporated into the application to help the user comply with their treatment plan, keeping the interest undiminished. The rest of the paper is structured as follows. The following section describes the HOLOBALANCE project and its modules. In the third section, the purpose, design and motivation features and evaluation of the Activation Planning Application are described in detail. Finally, conclusions and future steps are presented in the fourth section.

THE HOLOBALANCE PROJECT

Human balance is the result of many body systems working together: the visual and vestibular systems, proprioception and the musculoskeletal function are integrated in a complex and multifactorial way. The interaction between these systems and sophisticated mechanisms support anticipatory postural adjustments and adapt to changing environmental and balance task demands by means of sensory re-weighting (13). Age-related decline in all these sensory inputs and functions in older people is well-documented and leads to impaired postural and ambulation control and an increased risk of fall and injury (14). Gait and balance disorders can greatly affect daily life on a physical and psychological level and are therefore one of the leading causes of death among the senior citizens. Although medical conditions like arthritis and orthostatic hypotension are the common causes, most of the gait and balance disorders are multifactorial in origin and the consideration of both the contributing factors and the targeted interventions require a comprehensive evaluation (15). Deficits in executive functioning skills, which include a group of higher cognitive processes concerning one's ability to organize thoughts, prioritize tasks and make decisions, are associated with impairments in postural control, reduced gait speed, and increased falls risk (16, 17). Cognitive impairment and affective disorders or psychiatric conditions such as depression, anxiety, fear of falling, and sleep disorders have been also connected with balance dysfunctions and falls (18).

Due to the multifactorial nature of the majority of balance and gait disorders, there is a need to combine several interventions of treatment in order to restore and improve functional

[1]https://holobalance.eu/

capacity (19). Gait disorders that are related to chronic medical conditions can be treated to some extent by medical and surgical interventions. Moreover, the use of home environment assessment and intervention, especially when integrated in a multifactorial program, can significantly reduce falls and improve the quality of living of the elderly (20). On the other hand, balance training, including exercise and physical therapy, is gaining ground in the treatment of balance and gait disorders. The joint American and British Geriatric Society guidelines (21) recommend that a suitable balance exercise programme is a crucial component for balance rehabilitation. Through customized balance physiotherapy intervention, persons with postural deficits who have experienced falls or are at risk of falling, perform individualized exercises daily in a safe environment. The protocols commonly used in these interventions include a first visit to the physiotherapist for an initial assessment and the creation of a treatment plan. The patient then visits the specialist on a weekly basis, while also performing some exercises and a walking program at home. During treatment, the therapist assesses the patient's condition and progress and thus makes the necessary modifications in the program (22). The exercises included in the treatment plan are usually performed at physiotherapy and rehabilitation centers. Nonetheless, recent findings show that interventions that include exercise sessions in the home environment of the patient seem to be quite effective and present higher adherence rates than group-based community sessions (23). Cognitive training has also directly demonstrated benefits on balance and gait parameters in older adult fallers, while treatment strategies for anxiety and depression incorporated in balance training programs have proven to be effective (24, 25).

However, to date, there is a lack of personalized solutions that have been proposed to improve physical activity and the engagement of people with balance disorders to an exercise and physiotherapy program. The challenges in providing effective balance physiotherapy coaching include lack of compliance, difficulty in proper exercise performance and limited access to specialized physiotherapists and balance clinics. Despite evidence, suggesting that combined cognitive and functional training results in better outcome compared to either in isolation, there has been limited carryover into clinical practice.

The overall objective of HOLOBALANCE is to develop and validate a new personalized hologram coach platform for virtual coaching, motivation and empowerment of the aging population with balance disorders. The central idea of the intervention is the training of the users through a customized and personalized program, which will be formed through the use of ambient and wearable sensors and augmented reality interaction methods. The project engages experts related to the treatment of balance disorders, including physiotherapists, Ear Nose Throat experts, neurologists, psychologists, and gerontologists, who have the ability to track users' progress through an expert panel, formulate a comprehensive treatment plan and customize it with the help of progressing learning algorithms.

The HOLOBALANCE platform consists of the following coaching components:

- A hologram based surrogate balance physiotherapist

  The benefits of the traditional methods of physiotherapy and exercise training are fading because of lack of compliance or missed sessions as a result of the limited access to specialized clinics. Moreover, training programs that are performed at home environment often present poor adherence and fail to motivate the users. In order to overcome these challenges, we propose a virtual coach with daily presence in user's home. The physiotherapist, in the form of a hologram, can monitor, and assess the activities and the performance during balance training, while motivating and empowering the user. The reasoning of the virtual coach is implemented by a set of wearable and ambient sensors. This allows for an objective assessment of exercise performance and real-time feedback. For the physiotherapist and other healthcare professionals, the use of this sensor technology allows for ongoing daily activity coaching regarding quality of exercise performance, physical activity levels, activities of daily living performance and compliance with goals in order to provide feedback, update of instructions and exercise progression on a real time basis, which may further motivate and enhance self-management of users.

- The augmented reality cognitive games and exergames combined with auditory exercises

  The inclusion of the cognitive games and exergames module aims to empower and motivate people during their balance physiotherapy through augmented reality gamification. It includes games that incorporate key design factors from the entertainment video game world and aim to place both physical and cognitive training in a pleasant, motivating and engaging context. On the other hand, the main goal for the auditory exercises is to provide training tasks that improve the patient's auditory memory and perception of speech in noise. The auditory training is conducted through a smartphone application.

- The physical activity planning component

  The sensor technology used during physiotherapy sessions allows for continuous monitoring of patients' levels of physical activity, their progress and their degree of compliance with the treatment plan. The healthcare professionals engaged in the project can have access to these data through a specially designed dashboard interface. They can use this interface to register new patients in the HOLOBALANCE system, store useful information and define or adjust the physical activity and physiotherapy plans (26). The Activity Planning Application (APA) module, which is the main user interface, plays a crucial part in the physical activity planning. It is a smartphone application that allows users to constantly monitor their training plan as well as their progress, both in terms of physical activity and interaction with all parts of the system. It fulfills the key motivational strategies of the platform in order to engage the users and maintain their interest in the use of the system. Toward this direction, the Virtual Communities (VCs) dashboard integrates with the application. The main purpose of the VCs dashboard is to manage the interaction of the patients

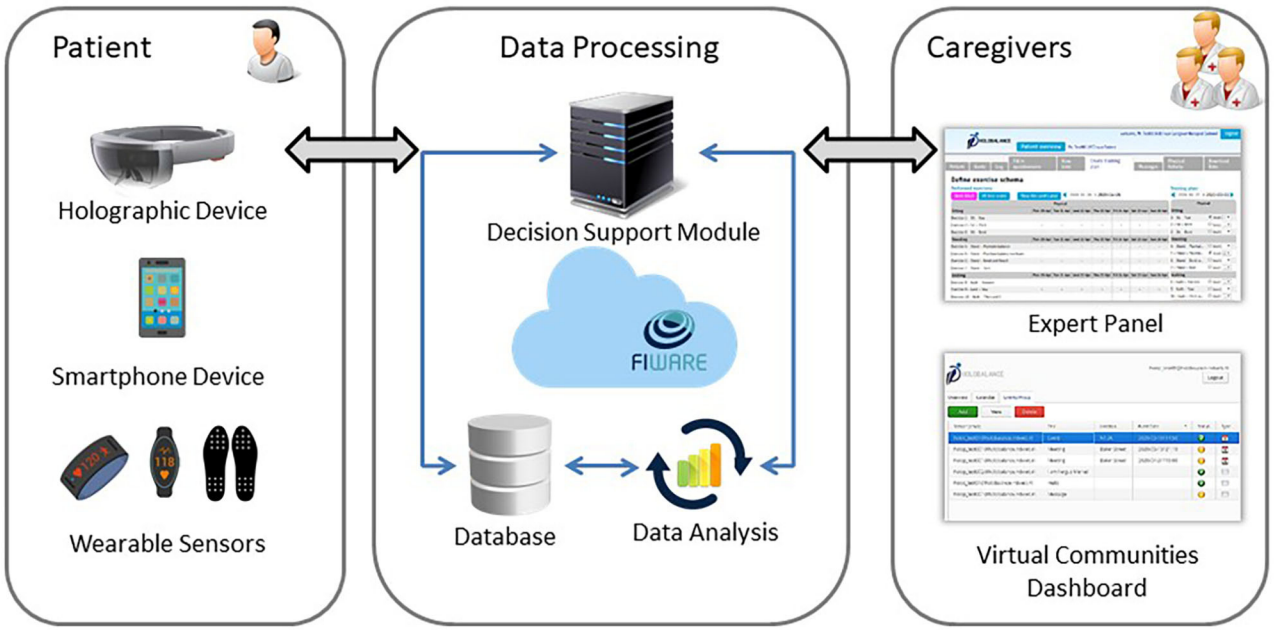

FIGURE 1 | Outline of HOLOBALANCE architecture.

with each other in a virtual community, where they can get in touch with other users of the HOLOBALANCE system, while being in touch with their clinicians. The authorized users of the dashboard, consisting of clinicians and local agent communities, can inform the patients about a variety of interesting issues and notify them about upcoming events. They can also satisfy a user's request to display a message or a suggested event in the community. The users' requests, the declaration of their willingness to attend an event and the display of the approved message and event list are performed through the APA that interacts with the VCs component.

An outline of the HOLOBALANCE architecture is shown in **Figure 1**. The part of the system that is placed in-home environment near the user consists of IoT devices that enable the monitoring of user's actions, behavior changes and level of activity. A set of hardware-software interfaces provide a standardized and interoperable layer to communicate with different sensor devices and acquire their data. A FIWARE-Orion enabled communication module handles the interaction between the components that are placed in the user's home and the cloud infrastructure. The latter integrates all the required services for the interaction between its different modules and performs all actions related to data exchange, processing, advanced analytics estimation and representation. It is consisted of a REST/JSON API over HTTPS that provides generic database queries for Create, Read, Update, and Delete (CRUD) operations. Authorization, authentication and user management are also provided. The data repository of the cloud supports various database engines and types of data as well as data replication and push notifications through Firebase. The advanced data analytics module of the cloud infrastructure allow the deployment of machine learning and deep learning algorithms to execute condition evaluation and user behavior tasks on data collected from the devices. The results of this multilevel analysis are used to feed back the system and the interfaces available to the involved clinical experts, through which they have overall oversight of the intervention. The main goal of the proposed interoperable platform is to remove restrictions regarding the system development and the design of any potential application to specific hardware and software components. It promotes a system that can be easily deployed in different areas, shifting the effort from design time investments to the process of developing innovative telehealth applications with advanced analytics and more efficient treatment plans (27).

THE HOLOBALANCE ACTIVITY PLANNING APPLICATION

Purpose

The APA is the main interface through which users communicate with the HOLOBALANCE platform. The main purpose of the application is for them to be able to constantly monitor their activity, as well as to determine to what extent they benefit from the use of the system. As poor adherence is one of the main challenges in long-term interventions, the application is intended to implement a significant part in inciting user engagement and motivation. Thus, all the design and implementation steps described in detail below were based on the objectives and the desired impact of the intervention, taking into account the characteristics and requirements of the target user group.

User-Experience Design

Determining the necessary functions and data that an application should contain is the first important step in creating a pleasant and meaningful user experience. This process is guided by the clarification of the target group and the main objectives of the application. The APA is aimed at adults over the age of 65 who suffer from balance disorders. Its main purpose is to offer patients with the capability to monitor their activity and have an overview of their progress resulting from their interaction with the components of the HOLOBALANCE platform. For that purpose, a number of functionalities regarding the display of information and the provision of education have been added. Through the smartphone application, users are enabled to keep track of the key elements regarding their activity recorded by an activity tracker, at any time of the day as well as of how close they are to achieving the goal set in their treatment plan. Social interaction functionalities were also determined as a powerful method for enhancing user health behaviors. The users of APA can communicate with other members of the HOLOBALANCE virtual community through the social interaction component of the application. They are informed about upcoming events and interesting issues and can exchange their ideas and concerns.

User-experience design has a major impact in the motivation of the users and in the effectiveness of the application. Thus, it is crucial to design and implement an overall motivation strategy that enhances user engagement. The purpose of APA is to motivate users both in terms of their daily physical activity and their engagement with the various components of the HOLOBALANCE tele-rehabilitation system.

The HOLOBALANCE is a complex intervention with several interacting components. It aims to bring about a series of behavioral changes concerning the health, the balance and the well-being of users. Theories of behavioral change aim to provide a theoretical framework for the factors influencing behavioral change and motivation (9) and which in turn are related to the success or failure of an intervention. A large number of theoretical frameworks have been proposed to conceptualize health-related behavioral change (10). These frameworks vary in their degree of empirical support, and many have been developed for specific health domains which many not be relevant to the specific characteristics of the HOLOBALANCE intervention (11).

Thus, a systematic approach was adopted for the selection of the theoretical framework to be used in HOLOBALANCE, based on the following criteria:

(a) The framework should be comprehensive, in the sense that its components include not only the intentional and motivational elements resulting in intervention adherence, but also the interplay of these psychological factors with the physical means, or physical capability to execute the intervention. This is important because HOLOBALANCE is an intervention delivered through technology.
(b) The framework should have been empirically validated through meta-analysis reporting that interventions informed by the model lead to changed behaviors.
(c) The framework should have been previously applied to other interventions similar to HOLOBALANCE, as assessed by publications on the following: promoting adherence to interventions with computer-based technological components, intervention involving physical activity, and population target of older adults.

Based on the above, the Capability, Opportunity, and Motivation (COM-B) model of behavior was chosen (12). The model posits that the interaction between Capability, Opportunity and Motivation (COM) causes the changes in Behavior (B), where: Capability is the "individual's psychological and physical capacity to engage in the activity concerned," Opportunity includes the "factors that lie outside the individual that make the behavior possible or prompt it" and Motivation includes Reflective Motivation (evaluations, intentions, and plans) and Automatic Motivation (emotions and impulses arising from learning and innate dispositions). The COM-B model incorporates the interplay between psychological and physical elements of the behavior through the components of physical capability (the physical ability of the person to engage in the target behavior), and physical opportunity (the environmental resources, including technology, that facilitate the target behavior).

This model has been developed as part of the Behavior Change Wheel (BCW), which is designed to help intervention designers move from a behavioral analysis of the problem to intervention design using the evidence-base. The BSW defines nine intervention functions (education, persuasion, incentivization, coercion, training, restriction, environmental restructuring, modeling, enablement) and seven policy categories (communication/marketing, guidelines. fiscal measures, regulation, legislation. environmental/social planning, service provision) and allows developers to identify systematically which of them could bring about behavioral change (28). Once the intervention functions and the policy categories are chosen, behavior change techniques and effective modes of delivery are identified.

The COM-B and BCW have already been used in areas relevant to the HOLOBALANCE intervention, including gamification of mobile health interventions (29, 30) and promoting adherence to hearing aids in older adults (31) while it has also been used in a recent meta-analysis to classify motivational interventions to promote exercise adherence for older fallers (32). After a detailed literature review of the parameters that promote physical exercise and exercise adherence, especially in the elderly, as well as after the feedback that was given in older adult workshops that were organized as part of the project, the motivation strategies of HOLOBALANCE were determined. These key strategies were related to the relevant components of COM-B and BCW models (33).

The motivational part of the APA is based on the overall motivational strategy of the project. Once the target behaviors and the appropriate intervention functions that will lead to behavioral changes were identified, according to the COM-B model, the specific Behavioral Change Techniques (BCTs) that the application will implement were defined. By BCT, we

mean an observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior; that is, a technique is proposed to be an "active ingredient" (e.g., feedback, self-monitoring, and reinforcement). BCTs can be used alone or in combination and in a variety of formats (34). The selection of techniques was made using the BCT Taxonomy (BCTTv1), which lists 93 BCTs with descriptions and examples of their application (35).

The motivation features that were derived from the above procedure and are included in the application can be categorized logically in informative, gamified, and social features (36).

Informative Features

These features allow the user to monitor their physical condition and progress in an easily accessible and understandable way. They are based on the fact that providing information to the user can itself be a powerful motivator for more effort. In the context of our application, the users can be informed at any moment about their daily step target, which is determined by the attending clinician, and about their longer-term general goals that have been formed in consultation with him/her. Their progress is displayed on the main screen when entering the application both in percentage and in numbers, while a card contains other useful information about their activity, such as the calories they burned and the distance they traveled. An arch chart has been chosen to visualize what part of the daily step goal has been fulfilled. It is also possible for them to monitor the long-term course of their activity through a 5-day chart. In this bar chart, the steps performed on each one of the past 5 days are represented by a bar, the color of which depends on whether the user has achieved the daily step target. Notifications are sent daily, informing and encouraging the users to complete their daily training plan and fulfill their goals. A message is displayed in the notification drawer and when tapped the user is redirected to a list with the exercises of the defined training program annotated based on whether they have already been performed or not. However, users are not only motivated when they are aware of their activity and progress but also when they receive general information and training about the goals and benefits of the intervention in which they participate. Through the functionalities that result from the integration of VCs dashboard with the application, the clinicians can provide easily accessible supportive educational material to users.

Gamified Features

Gaming methods are widely used in motivation strategies, as they enhance the maintenance of the user's interest and engagement in a pleasant way. This is the integration of methods encountered in

TABLE 1 | The motivational features of the application and their correspondence with the Behavioral Change Techniques Taxonomy (BCTTv1).

Motivational features of APA	COM-B component	BCT Label (BCTTv1)	BCT definition
Daily target/general goal/progress chart	Psychological capacity, reflective motivation	2.2 Feedback on behavior	Monitor and provide informative or evaluative feedback on performance of the behavior (e.g., form, frequency, duration, intensity)
		2.7 Feedback on outcomes of behavior	Monitor and provide feedback on the outcome of performance of the behavior
Notifications	Physical opportunity	7.1 Prompts/cues	Introduce or define environmental or social stimulus with the purpose of prompting or cueing the behavior. The prompt or cue would normally occur at the time or place of performance
Educational material through virtual communities	Psychological capacity	4.1 Instruction on how to perform a behavior	Advise or agree on how to perform the behavior (includes "skills training")
		5.1 Information about health consequences	Provide information (e.g., written, verbal, visual) about health consequences of performing the behavior
		5.3 Information about social and environmental consequences	Provide information (e.g., written, verbal, visual) about social, and environmental consequences of performing the behavior
		5.6 Information about emotional consequences	Provide information (e.g., written, verbal, visual) about emotional consequences of performing the behavior
Badges/achievements	Automatic motivation	10.4 Social reward	Arrange verbal or non-verbal reward if and only if there has been effort and/or progress in performing the behavior (includes "positive reinforcement")
		10.5 Social incentive	Inform that a verbal or non-verbal reward will be delivered if and only if there has been effort and/or progress in performing the behavior (includes "positive reinforcement")
Leaderboard	Social opportunity	6.2 Social comparison	Draw attention to others' performance to allow comparison with the person's own performance
Messages/events through virtual communities	Reflective motivation, psychological capacity	3.1 Social support (unspecified)	Advise on, arrange or provide social support (e.g., from friends, relatives, colleagues, "buddies" or staff) or non-contingent praise or reward for performance of the behavior It includes encouragement and counseling, but only when it is directed at the behavior

computer games, since they are first adapted to the requirements and needs of the users of each application. Consequently, it is crucial for an intervention for senior citizens to account for the most important age-related issues. Elderly users are not familiar with gaming systems and gamified features that are included in non-gaming applications. The lack of experience they are likely to have with gaming can be a barrier to understanding the metaphors derived from digital games and therefore can hinder their engagement (37). In our implementation, easily accessible and simple gamified features that do not assume user's prior gaming experience were chosen. A number of badges are calculated and given to the user as rewards for achieving several goals. The calculation is performed individually, taking into account the treatment plan and the goals set separately for each user by the clinician. The collection of badges covers the activities of all the interventions that compose the HOLOBALANCE project and is based on the logic of positive feedback, rewarding the effort more than the performance. It includes a number of predefined badges as well as some badges that are initially locked and can be unlocked after several achievements. By introducing these unlockable features, the users are encouraged to remain engaged over a longer period of time.

As part of the healthy competition between the members of the project community, a leaderboard was integrated in

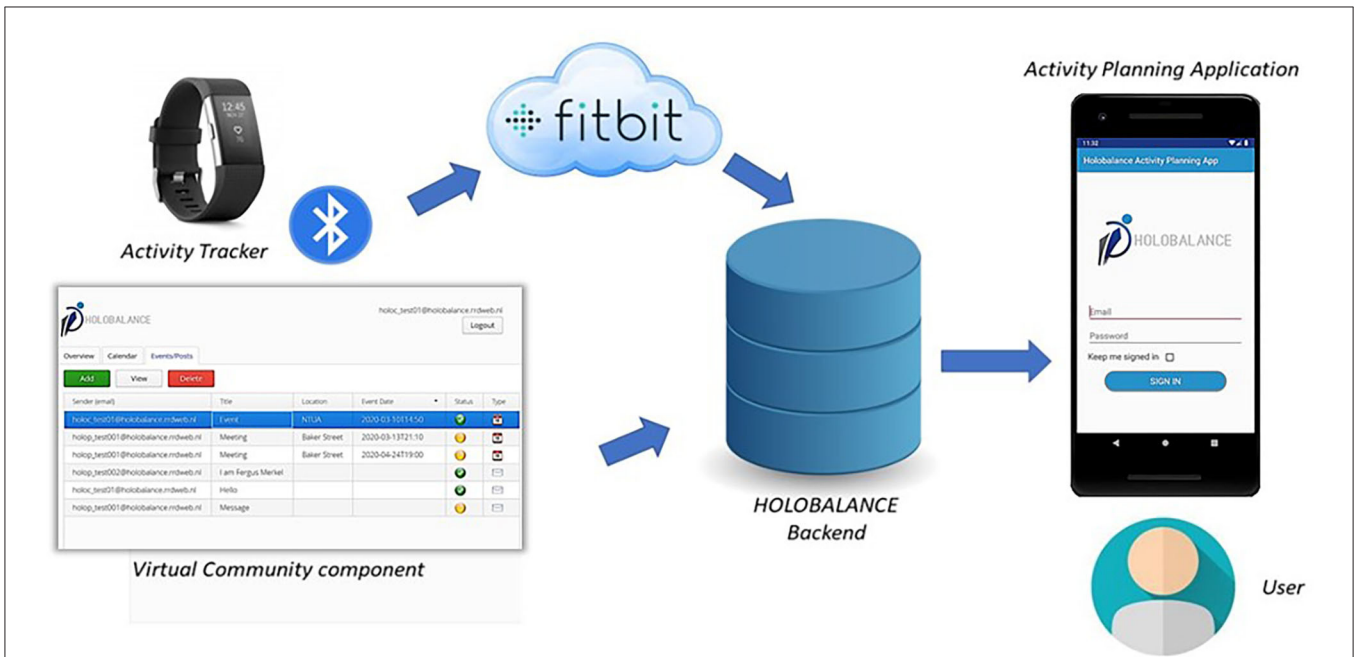

FIGURE 2 | Activity Planning Application in the HOLOBALANCE system.

TABLE 2 | Information about the participants and the evaluation protocol of the semi-structured focus groups performed in UK, Germany, and Greece.

	Focus group location	Protocol	Participants
Iteration I	London (UK)	1. Welcome, introduction, and obtaining participants' consent for participation 2. Demographic questionnaires 3. Demonstration of the 1st mockup of APA 4. Q&A with participants and clinical, technical, and usability experts 5. APA usability questionnaire (modified SUS) and open-ended questions	1. $N = 47$ 2. 29 females (63%) 17 males (37%) 3. Age: 60–84 years ($M = 73.61$, $SD = 6.12$, Mdn = 72.5) 4. $N = 13$ (28%) had balance disorder 5. 94% use computers daily 6. Confidence in using technologies = 3.32 ($SD = 0.91$, Mdn = 3) 7. 60% familiar with social media
Iteration II	Freiburg (Germany) and Athens (Greece)	1. Welcome, introduction, and obtaining participants' consent for participation 2. Participants used all Holobalance modules including APA for ∼15 min 3. Q&A with participants and clinical, technical, and usability experts 4. Questionnaires (Demographic, usability, etc.) and open-ended questions	1. $N = 24$ 2. Freiburg: $N = 8$, Athens: $N = 16$ 3. 21 females (63%) 3 males (12.5%) 4. Age: 60–84 years ($M = 71.42$, $SD = 7.99$, Mdn = 72) 5. $N = 12$ (50%) had balance disorder 6. 58.33% use computers daily 7. Confidence in using technologies = 2.33 ($SD = 1.34$, Mdn = 2.5)

the application. Through this component, which is reinitialized and updated every week, the users are able to see their weekly ranking between the other members of the community, without being able to see the names of the other competitors. The rankings are based on the users' badges. Specifically, possession of each of the badges is translated through an algorithm into the possession of "stars," which is our unit of scoring.

Social Features

Social influence is a great source of motivation. Through social support and pressure the users are motivated to be better and maintain their interest in the use of the system. Communication and interaction between users of the system is mainly through the functions of the Cs dashboard, which are accessible through the application. Specifically, users can exchange messages, be informed about events planned in their area as well as suggest the organization of an event. In this way, they are given the opportunity to communicate their concerns, build relationships and feel part of a team with a common goal, which is to encourage and motivate them. The leaderboard component, which was described above, can also be mentioned among social features, as it allows users to share their performance with others and compare their results.

The motivational features of the application and the BCTs to which each one corresponds are summarized in **Table 1**.

User-Interface Design

Designing a usable, efficient and enjoyable mobile application requires the combination of emotional and cognitive components. The basic usability and aesthetics principles should be satisfied and user needs and preferences should be constantly in the forefront throughout the process. As for the group of elderly users, the perception of their negative attitude toward technology often results in a lack of emphasis on their special needs during the design procedure. There is a general mismatch between perception about elders' needs and their actual needs, and a misinterpretation of elders' needs and demands by the non-elder users (38). The fact that people today live more and want to be active and independent as well as the fact that they are becoming more and more familiar with the technology and the benefits it offers them in their daily lives, makes it imperative to design technologies that will satisfy their needs.

The APA was created in order to be used by elderly people, therefore in the part of its design emphasis was placed on the needs and requirements of this age group. When aging, there are three main problem categories: sensory problems, motor problems, and cognitive problems (38). Below is an extensive description of each of the above categories of challenges, as well as the design decisions made in the application to address them:

- **Sensory problems**

  One of the characteristics of old age is the weakening of all the senses, but mainly those of hearing, sight and touch. The majority of people over the age of 40, regardless of gender, begin to show some degree of progressive hearing loss as time goes on. This decline is greater at higher frequencies (39). APA does not include any audio components or ring tones in its context. However, daily notifications are sent through the application. In order to ensure as much as possible that the user will be notified, even if there is difficulty in hearing the alert sound, vibration and light signals have been added when designing the notification module. Regarding the sensor of sight, older people present impairments regarding the breadth of visual field, the visual processing speed and the perpetual flexibility (38). They have difficulties in reading text in menu and text messages especially when font is not gothic, and when color contrast between text and background is not obvious (40). The application presented in this paper uses big font, big buttons and obvious color contrast. It includes many simple icons that represent different concepts, but there is a concern for their simultaneous description in the form of text. Thus, any difficulties related to reading the text or understanding the icons are addressed.

- **Motor problems**

  Movement control and manual dexterity is affected by aging, especially when diseases like multiple sclerosis or Parkinson's disorder make their appearance. Older people need more time to respond to a movement task and their moves are less precise, mainly due to a decrease in the muscle mass and strengths (38). They have difficulty in differentiating short push and long push on one button and they type slowly (40). APA is designed to include the absolutely necessary number of buttons, each of which performs a single function. Spacing is as large as possible, always following the design properties of each component, so as to minimize the possibility of pressing in the wrong place. Textual feedback is given when the effect of a movement is not clear to the user.

- **Cognitive problems**

  The deficits concerning the cognition level of seniors need to be carefully considered when designing a mobile application. Attention, memory and decision-making ability are affected

TABLE 3 | Means (standard deviations), medians, and percentage of the responses to a modified version of the SUS questionnaire.

Item	M (*SD*)	Mdn
I would like to use this system frequently if it meant that it would reduce my risk of falling	3.75 (1.28)	4
I think this system would be easy to use	3.46 (1.32)	3
I would imagine that most people would learn to use this system very quickly	3.53 (0.98)	3
I would feel very confident using this system	3.73 (1.12)	4
I think that I would need assistance to be able to use this system. If yes, what type?	**Percentage (%)**	
No assistance	30.6	
Physical assistance	5.6	
Technical assistance	50	
Educational assistance	16.7	
Medical assistance	2.8	

by aging, thus affecting human-machine interaction. Elders find it difficult to learn and understand new procedures or large amount of information at once, perform difficult tasks and solve problems (41). Regarding the above, the navigation process of APA has been kept simple and austere. Only task-relevant information is displayed, avoiding parallel material and tasks. Pop-up windows and other dialog view components are displayed longer than usual in smartphone applications, in order to facilitate their processing by the user. Memory-requiring actions and tasks that assume either user's familiarity with mobile applications or knowledge of metaphors and interaction techniques have also been avoided. Notifications are sent through the application and serve as reminders and memory aids for the users, informing them about their training program.

Technical Design

The application presented in this document was developed for Android platform. Most functionalities are implemented through Android software development kit and the use of external libraries has been minimized. Model-View-View-Model pattern was chosen as the main software architectural pattern in order to keep a reusable and testable code.

The APA is part of a complex and multifactorial intervention and it is necessary to retrieve and process data recorded from other technological modules of the system. The information

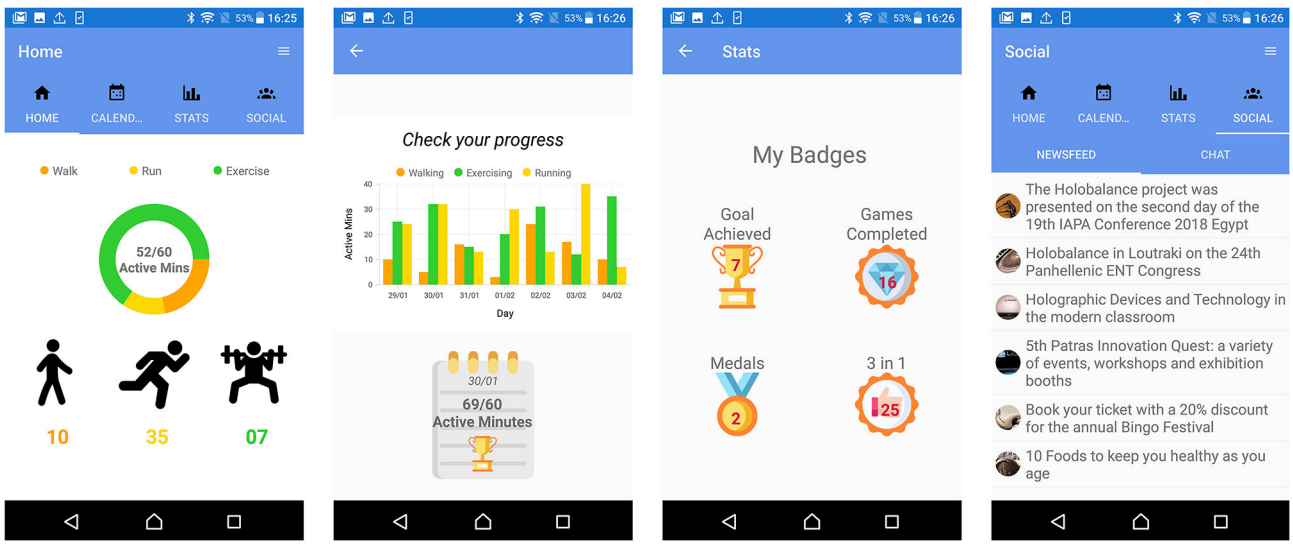

FIGURE 3 | Mock-ups of the Activity Planning Application.

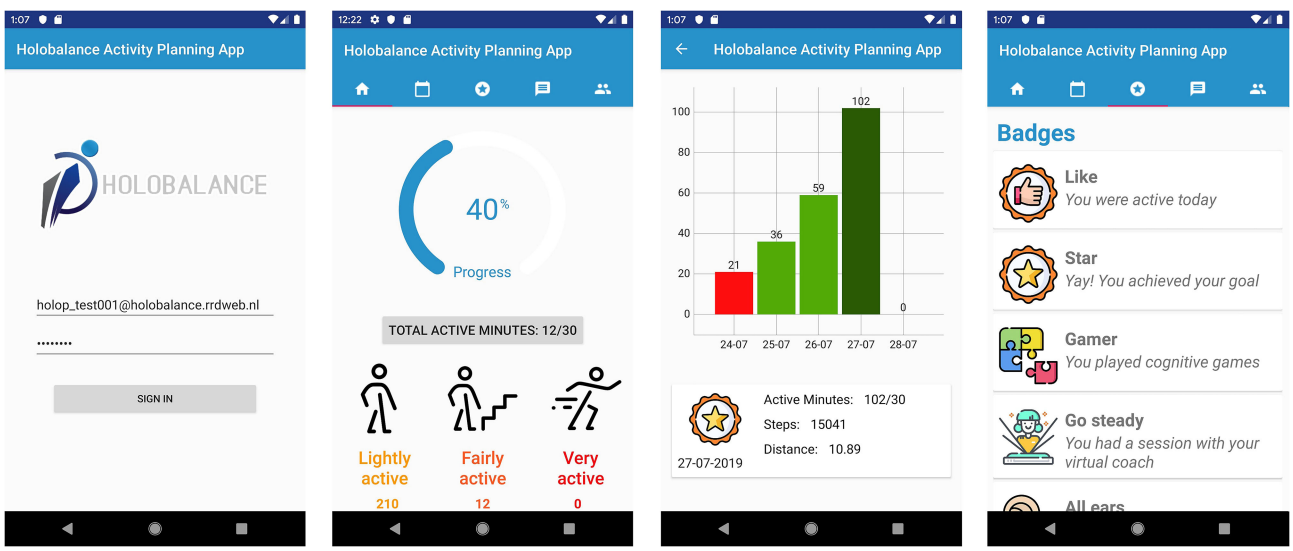

FIGURE 4 | First version of the Activity Planning Application.

about their daily activity recorded by the Fitbit activity tracker, as well as performance data regarding other platform's components are retrieved from the HOLOBALANCE backend infrastructure. The application also retrieves data about the posts and the events that should be displayed in the newsfeed page, where users can communicate and get informed about interesting topics. These data are available through the VCs interface, where the authorized clinicians and local agent communities can define the content of the newsfeed page that will be displayed to the profiles of a group of patients.

The communication between APA and the other modules uses HTTP protocol via JSON formatted messages in RESTful manner. As a result, the application receives real-time feedback concerning the activity and the training plan of the patient, as well as the content of the social component. Regarding the rewarding system of the application, data recorded by the activity tracker and results from physiotherapy, cognitive and auditory training are needed. APA consumes these data from the HOLOBALANCE cloud infrastructure, where they are stored. The interactions between the APA and the other modules of the system are shown in **Figure 2**.

Evaluation

The design and development of the APA followed a human-centered design approach. That is, through an iterative development process, the APA requirements were specified, new mock-ups and prototypes were implemented, and were then evaluated by the end users. To analyze the requirements and evaluate the mock-ups, several semi-structured focus groups and interviews were performed across three European countries (UK, Germany, and Greece). After each iteration, new insights have evolved which led to re-prototyping of the concepts and content elements.

Following this approach could ensure the user experience requirements of the final product and an easy-to-use tool that communicates harmoniously with the rest of the system and serves the purposes and strategies of the overall intervention. Details about the participants of the focus-groups and interviews that were performed as well as the evaluation protocol that was followed are presented in **Table 2**.

The initial requirements were defined through interviews and focus groups with the project's clinical experts. Based on these requirements, the first mock-up of APA was developed. This mock-up was then evaluated through a semi-structured focus group with older adults in London. During the focus group, participants could ask their questions, discuss the presented mock-up with the technical and clinical experts, and finally express their opinions through questionnaires and direct conversations with clinical, technical, and usability experts.

Forty-seven older adults (29 females) aged between 60 and 84 years (M = 73.61, SD = 6.12, Mdn = 72.5) participated in the focus group, 13 (28%) of whom had balance disorder. Regarding their familiarity with technology, most of them (94%) claimed that they use computer systems (including smart phones, tablet PCs, etc.) daily. On a scale of 1 (very bad) to 5 (very good), their confidence in using technologies was rated on average 3.32 (SD = 0.91, Mdn = 3). They were also asked about their experience in using social media. The average (60%) claimed that they have used social media and Facebook was named as the most frequently used application.

After the demographic questionnaire, the APA was presented and explained to the participants. In order to estimate the usability aspects of the APA, a modified version of the System Usability Scale (SUS) (42) was given to the participants. It consisted of the five most important items of the SUS for our case. Its four items were rated on a scale of 1 to 5 representing a range of Strongly Disagree to Strongly Agree

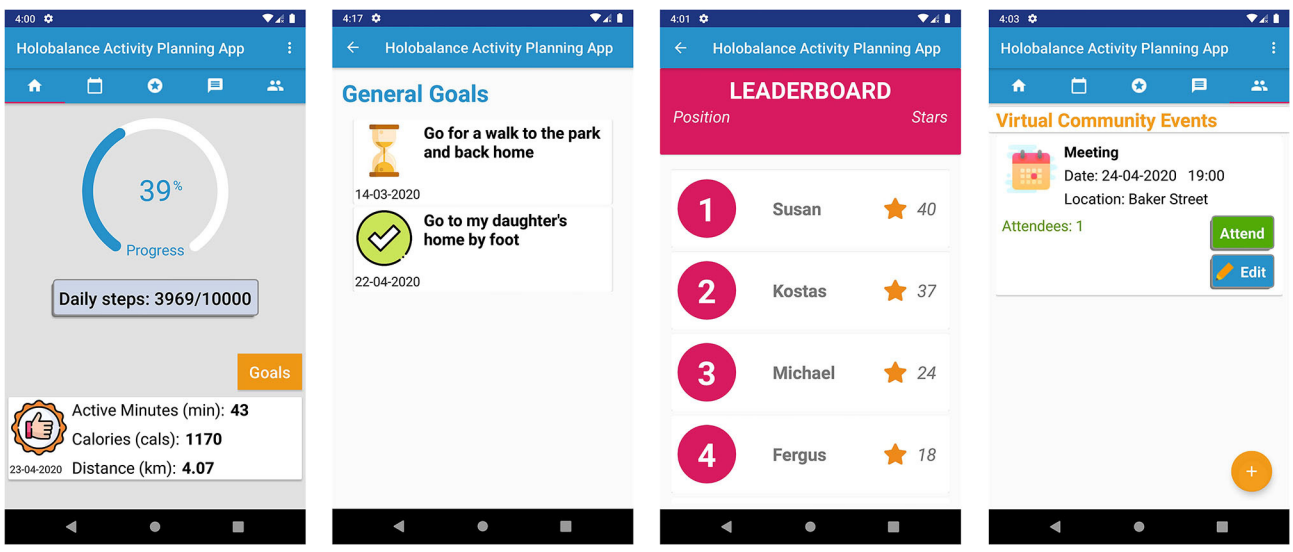

FIGURE 5 | Final version of the Activity Planning Application.

with the statement. The Assistance item, consisted of multiple choice questions and was served to estimate how many percent of the participants needed what type of assistance. The items of this questionnaire and the participants' responses are presented in **Table 3**.

Focus groups and interviews were also performed in Freiburg (Germany) and Athens (Greece). We interviewed in total 24 older adults (21 female) between 60 and 84 years old (M = 71.42, Mdn = 72, SD = 7.99), from whom eight were in Freiburg and 16 in Athens. Half of them believed that they have balance disorder. About two-thirds (66.66%) had access to broadband internet at home and used a computer (including smart phones, tablet PCs, etc.) on a daily (58.33%) or weekly (4.16%) basis. However, the other third (37.5%) reported that they never use any computer system. Moreover, the participants' average confidence with technology on a scale of 1 (very bad) to 5 (very good) was 2.33 (SD = 1.34, Mdn = 2.5). Among the other components of the HOLOBALANCE platform, mock-ups of the APA were presented to the participating older adults (**Figure 3**). They were then asked to comment on their experience of interacting with the application and to share their ideas, comments, and impressions. The followings are some examples of their comments:

- "You need to prepare a manual for the application"
- "It is simple, I like it"
- "I do not know if the competition between users would be a good idea"
- "I would like to see the others' progress"
- "I am not familiar with smartphones"
- "I don't have a mobile phone and don't know how to handle it"
- "It is good to know for how long I was active during the day"
- "I can see the letters but maybe the buttons could be bigger"
- "I do not know if I want to chat with the other members."

The above results from the focus groups organized in London, Freiburg, and Athens showed that the participants were neutral about using the APA. A large number of participants (about 50%) noted that they require technical assistance to use it. Most found the application simple and could navigate without problems. Some, however, suggested changes in the design that would make it easier to use, such as using bigger buttons and a larger font size. More detailed evaluation outcomes, especially regarding other aspects and modules of the Holobalance system can be found in (43, 44).

Based on the feedback from end users as well as from the clinical experts of the HOLOBALANCE project, different versions of the APA were created, as many features were evaluated and updated (**Figures 4**, **5**). The evaluation of the current version's adoption and acceptance will be conducted to the target users involved in the proof of concept at the pilot studies of the project. Four pilot sites were determined and each of them will recruit 20 participants and 20 control subjects. In total 80 participants in the intervention group and 80 in the control group will be recruited. Qualitative and quantitative measures will be used to process the results of these pilot studies and validate the use of the APA.

DISCUSSION

The rapid growth of mobile technology in recent years has paved the way for the development of innovative healthcare interactive systems. mHealth applications are widely used either as stand-alone or as part of a more comprehensive system for the purpose of disease management and healthy life-style. Smartphones are often involved in interventions for the elderly, as they are more familiar and prefer them over other technologies.

In this article, a smartphone application integrated in a complex intervention to support older people with balance disorders is described. The process of the application's design and development was human-centered and special emphasis was placed on the requirements of the target user group. The first mock-up interfaces were defined through interviews and focus groups with the project's clinical experts. They were subsequently evaluated in semi-structured focus groups and interviews that were performed across three European countries. The results were encouraging, as participants found the application simple to use and easy to navigate. However, they also pointed out their lack of familiarity with technology and their need for technical support. The feedback on the design features and the functionalities of the application led to the revision of several content elements, following an iterative development approach. The current version will be evaluated at four pilot studies, which will recruit a total of 160 people, divided equally into one intervention group and one control group.

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/s.

ETHICS STATEMENT

This study involved human participants and was reviewed and approved by the Ethics and Deontology committee of Hippokration General Hospital, Athens (reference number: 9769/24-6-2019), by the Ethics Committee of the Medical Association of Westphalia-Lippe and Westphalian Wilhelms University (reference number: 265/19), by Health Research Authority and Health and Care Research Wales (IRAS project ID: 248101) and by the Biomedical & Health Sciences, Dentistry, Medicine and Natural & Mathematical Sciences Research Ethics Subcommittee of King's College London. The Trial Registration Number of the study is NCT04053829 (https://clinicaltrials.gov/ct2/show/NCT04053829). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TA, AA, and IK made substantial contribution to the design and the development process of the smartphone application. D-EB, GG, and SC made substantial contribution to the

design of the motivation model. FM made substantial contribution to the evaluation of the application. SP and DK had general supervision of the development stage of the application. All authors contributed to revising the manuscript.

FUNDING

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 769574.

REFERENCES

1. Özsungur F. Gerontechnological factors affecting successful aging of elderly. *Aging Male.* (2019) 1–13. doi: 10.1080/13685538.2018.1539963. [Epub ahead of print].
2. Plaza I, Martín L, Martin S, Medrano C. Mobile applications in an aging society: status and trends. *J Syst Softw.* (2011) 84:1977–88. doi: 10.1016/j.jss.2011.05.035
3. Kalimullah K, Sushmitha D. Influence of design elements in mobile applications on user experience of elderly people. *Proc Comput Sci.* (2017) 113:352–9. doi: 10.1016/j.procs.2017.08.344
4. Klimova B, Maresova P. Elderly people and their attitude towards mobile phones and their applications—a review study. *Lect Notes Electr Eng.* (2016) 393:31–6. doi: 10.1007/978-981-10-1536-6_5
5. Zhao X, Wang L, Ge C, Zhen X, Chen Z, Wang J, et al. Smartphone application training program improves smartphone usage competency and quality of life among the elderly in an elder university in China: A randomized controlled trial. *Int J Med Inform.* (2020) 133:104010. doi: 10.1016/j.ijmedinf.2019.104010
6. Mallenius S, Rossi M, Tuunainen VK. Factors affecting the adoption and use of mobile devices and services by elderly people–results from a pilot study. In: *6th Annual Global Mobility Roundtable,* Vol. 31. Los Angeles, CA (2007). p. 12.
7. de Rezende LFM, Rey-López JP, Matsudo VKR, do Carmo Luiz O. Sedentary behavior and health outcomes among older adults: a systematic review. *BMC Public Health.* (2014) 14:333. doi: 10.1186/1471-2458-14-333
8. Kurniawan S. Older people and mobile phones: a multi-method investigation. *Int J Hum Comput Stud.* (2008) 66:889–901. doi: 10.1016/j.ijhcs.2008.03.002
9. Paul L, Brewster S, Wyke S, McFadyen AK, Sattar N, Gill JM, et al. Increasing physical activity in older adults using STARFISH, an interactive smartphone application (app); a pilot study. *J Rehabil Assist Technol Eng.* (2017) 4:1–10. doi: 10.1177/2055668317696236
10. Joosen P, Piette D, Buekers J, Taelman J, Berckmans D, De Boever P. A smartphone-based solution to monitor daily physical activity in a care home. *J Telemed Telecare.* (2019) 25:611–22. doi: 10.1177/1357633X18790170
11. Shake MC, Crandall KJ, Mathews RP, Falls DG, Dispennette AK. Efficacy of bingocize®: a game-centered mobile application to improve physical and cognitive performance in older adults. *Games Health J.* (2018) 7:253–61. doi: 10.1089/g4h.2017.0139
12. Li J, Hodgson N, Lyons MM, Chen K-C, Yu F, Gooneratne NS. A personalized behavioral intervention implementing mHealth technologies for older adults: a pilot feasibility study. *Geriatr Nurs.* (2019) 41:313–9. doi: 10.1016/j.gerinurse.2019.11.011
13. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? *Age Ageing.* (2006) 35:ii7–11. doi: 10.1093/ageing/afl077
14. Lord SR, Clark RD. Simple physiological and clinical tests for the accurate prediction of falling in older people. *Gerontology.* (1996) 42:199–203. doi: 10.1159/000213793
15. Salzman B. Gait and balance disorders in older adults. *Am Fam Physician.* (2010) 82:61–8.
16. Delbaere K, Kochan NA, Close JCT, Menant JC, Sturnieks DL, Brodaty H, et al. Mild cognitive impairment as a predictor of falls in community-dwelling older people. *Am J Geriatr Psychiatry.* (2012) 20:845–53. doi: 10.1097/JGP.0b013e31824afbc4
17. Vaughan L, Giovanello K. Executive function in daily life: age-related influences of executive processes on instrumental activities of daily living. *Psychol Aging.* (2010) 25:343–55. doi: 10.1037/a0017729
18. Muir SW, Gopaul K, Montero Odasso MM. The role of cognitive impairment in fall risk among older adults: a systematic review and meta-analysis. *Age Ageing.* (2012) 41:299–308. doi: 10.1093/ageing/afs012
19. Hausdorff JM, Nelson ME, Kaliton D, Layne JE, Bernstein MJ, Nuernberger A, et al. Etiology and modification of gait instability in older adults: a randomized controlled trial of exercise. *J Appl Physiol.* (2001) 90:2117–29. doi: 10.1152/jappl.2001.90.6.2117
20. Gillespie L, Gillespie W, Cumming R, Lamb S, Rowe B. American Geriatrics Society; British Geriatrics Society; American academy of orthopaedic surgeons panel on falls prevention. Guideline for the prevention of falls in older persons Interventions for preventing falls in the elderly. *J Am Geriatr Soc.* (2001) 49:664–72. doi: 10.1046/j.1532-5415.2001.49115.x
21. Panel on Prevention of Falls in Older Persons AGS BGS. Summary of the updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc.* (2011) 59:148–57. doi: 10.1111/j.1532-5415.2010.03234.x
22. Badke MB, Shea TA, Miedaner JA, Grove CR. Outcomes after rehabilitation for adults with balance dysfunction11No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the author(s) or on any organization with whic. *Arch Phys Med Rehabil.* (2004) 85:227–33. doi: 10.1053/j.apmr.2003.12.005
23. Thiamwong L, Suwanno J. Effects of simple balance training on balance performance and fear of falling in rural older adults. *Int J Gerontol.* (2014) 8:143–6. doi: 10.1016/j.ijge.2013.08.011
24. Iaboni A, Flint AJ. The complex interplay of depression and falls in older adults: a clinical review. *Am J Geriatr Psychiatry.* (2013) 21:484–92. doi: 10.1016/j.jagp.2013.01.008
25. Smith-Ray RL, Hughes SL, Prohaska TR, Little DM, Jurivich DA, Hedeker D. Impact of cognitive training on balance and gait in older adults. *Journals Gerontol Ser B Psychol Sci Soc Sci.* (2015) 70:357–66. doi: 10.1093/geronb/gbt097
26. Kouris I, Sarafidis M, Androutsou T, Koutsouris D. HOLOBALANCE: an augmented reality virtual trainer solution forbalance training and fall prevention. In*: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* Honolulu, HI: IEEE (2018). p. 4233–6. doi: 10.1109/EMBC.2018.8513357
27. Tsiouris KM, Gatsios D, Tsakanikas V, Pardalis AA, Kouris I, Androutsou T, et al. Designing interoperable telehealth platforms: bridging IoT devices with cloud infrastructures. *Enterp Inf Syst.* (2020) 1–25. doi: 10.1080/17517575.2020.1759146
28. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci.* (2011) 6:42. doi: 10.1186/1748-5908-6-42
29. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile h ealth (mHealth) technologies in dietetic practice: a three country study. *J Hum Nutr Diet.* (2017) 30:439–52. doi: 10.1111/jhn.12446
30. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. *J Med Internet Res.* (2014) 16:e9. doi: 10.2196/games.3413
31. Barker F, Atkins L, de Lusignan S. Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation. *Int J Audiol.* (2016) 55:S90–8. doi: 10.3109/14992027.2015.1120894
32. Hughes KJ, Salmon N, Galvin R, Casey B, Clifford AM. Interventions to improve adherence to exercise therapy for falls prevention in community-dwelling older adults: systematic review and meta-analysis. *Age Ageing.* (2019) 48:185–95. doi: 10.1093/ageing/afy164

33. Gatsios D, Georga EI, Kourou KK, Fotiadis DI, Kikidis D, Bibas A, et al. Achieving adherence in home-based rehabilitation with novel human machine interactions that stimulate community-dwelling older adults. In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. Rhodes (2019) p. 616–9. doi: 10.1145/3316782.3322762
34. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med.* (2013) 46:81–95. doi: 10.1007/s12160-013-9486-6
35. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative da. *Health Technol Assess.* (2015) 19:1–87. doi: 10.3310/hta19990
36. Bielik P, Tomlein M, Krátky P, Mitrík Š, Barla M, Bieliková M. Move2Play: an innovative approach to encouraging people to be more physically active. *IHI'12 - Proceedings of the 2nd ACM SIGHIT Int Heal Informatics Symposium.* (2012). p. 61–70. doi: 10.1145/2110363.2110374
37. Gerling K, Masuch M. "Exploring the potential of gamification among frail elderly persons", in *CHI 2011 Workshop on Gamification: Using Game Design Elements in Non-Game Contexts* (Vancouver, BC) (2011). p. 1–4.
38. Iancu I, Iancu B. Designing mobile technology for elderly. A theoretical overview. *Technol Forecast Soc Change.* (2020) 155:119977. doi: 10.1016/j.techfore.2020.119977
39. Fisk AD, Rogers WA. *Designing for Older Adults*. Boca Raton, FL: Taylor & Francis Group (2009).
40. Kim H, Heo J, Shim J, Kim M, Park S, Park S. Contextual research on elderly users' needs for developing universal design mobile phone. In: Stephanidis C, editor. *Universal Acess in Human Computer Interaction. Coping with Diversity. UAHCI 2007. Lecture Notes in Computer Science, Vol. 4554.* Berlin; Heidelberg: Springer (2007). p. 950–9. doi: 10.1007/978-3-540-73279-2_106
41. Lewis JE, Neider MB. Designing wearable technology for an aging population. *Ergon Des.* (2017) 25:4–10. doi: 10.1177/1064804616645488
42. Brooke J. SUS-A quick and dirty usability scale. *Usability Eval Ind.* (1996) 189:4–7.
43. Mostajeran F, Katzakis N, Ariza O, Freiwald JP, Steinicke F. Welcoming a holographic virtual coach for balance training at home: two focus groups with older adults. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Osaka: IEEE (2019). p. 1465–70. doi: 10.1109/VR.2019.8797813
44. Mostajeran F, Steinicke F, Ariza Nunez OJ, Gatsios D, Fotiadis D. Augmented reality for older adults: exploring acceptability of virtual coaches for home-based balance training in an aging population. In*: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI (2020). p. 1–12. doi: 10.1145/3313831.3376565

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.

frontiers in Digital Health

REVIEW
published: 30 September 2020
doi: 10.3389/fdgth.2020.567502

A Review of Virtual Coaching Systems in Healthcare: Closing the Loop With Real-Time Feedback

Kostas M. Tsiouris[1,2], *Vassilios D. Tsakanikas*[2], *Dimitrios Gatsios*[2,3] *and Dimitrios I. Fotiadis*[2,4]*

[1] Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece, [2] Unit of Medical Technology and Intelligent Information Systems, Department of Material Science and Engineering, University of Ioannina, Ioannina, Greece, [3] Department of Neurology, Medical School, University of Ioannina, Ioannina, Greece, [4] Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology - Hellas, Ioannina, Greece

This review focuses on virtual coaching systems that were designed to enhance healthcare interventions, combining the available sensing and system-user interaction technologies. In total, more than 1,200 research papers have been retrieved and evaluated for the purposes of this review, which were obtained from three online databases (i.e.,PubMed, Scopus and IEEE Xplore) using an extensive set of search keywords. After applying exclusion criteria, the remaining 41 research papers were used to evaluate the status of virtual coaching systems over the past 10 years and assess current and future trends in this field. The results suggest that in home coaching systems were mainly focused in promoting physical activity and a healthier lifestyle, while a wider range of medical domains was considered in systems that were evaluated in lab environment. In home patient monitoring with IoT devices and sensors was mostly limited to activity trackers, pedometers and heart rate monitoring. Real-time evaluations and personalized patient feedback was also found to be rather lacking in home coaching systems and this is the most alarming find of this analysis. Feasibility studies in controlled environment and an ongoing active research on Horizon 2020 funded projects, show that the future trends in this field are aiming to close the loop with automated patient monitoring, real-time evaluations and more precise interventions.

Keywords: virtual coaching, review, coaching systems, automated feedback, closed-loop systems, physical activity monitoring, rehabilitation coaching, behavioral change

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Alexander James Casson, The University of Manchester, United Kingdom
Maurice Mars, University of KwaZulu-Natal, South Africa

***Correspondence:**
Dimitrios I. Fotiadis
fotiadis@cc.uoi.gr

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *29 May 2020*
Accepted: *28 August 2020*
Published: *30 September 2020*

Citation:
Tsiouris KM, Tsakanikas VD, Gatsios D and Fotiadis DI (2020) A Review of Virtual Coaching Systems in Healthcare: Closing the Loop With Real-Time Feedback. Front. Digit. Health 2:567502. doi: 10.3389/fdgth.2020.567502

INTRODUCTION

Coaching systems are on a surge of increasing popularity for both every day and medical applications. The primary driving forces for advancement have been: (a) the enormous growth of access to cheaper smart devices, including smartphones, smartwatches, activity trackers, and in home sensing devices (1), (b) the prevalence of design and development focusing on seamless interconnectivity among them (i.e., Internet of Things, IoT) (2), (c) technology acceptance, as people are becoming increasingly aware and familiar with the added benefits of the functionality that is provided (3), (d) increasing reliable internet access, especially through wireless smartphone connectivity, and (e) the continuous increase in mobile and edge processing power, which raises the possibilities of what can be practically achieved (4). Fitness training and activity monitoring was the

killer application that brought massive attention and demand in the field, highlighting the need for more data gathering (i.e., better sensing technologies) and most importantly, the need for advanced data interpretation algorithms and personalized user feedback based on these data (i.e., precise coaching). Machine learning, artificial intelligence and advanced data analytics are the major contributors for making more sense from the sensing data. In the medical and healthcare field, coaching systems have been more relevant for chronic disease patient management and disease modifying behavioral changes, in an attempt to guide and improve intervention outcomes (5).

Virtual coaching is an emerging field in the healthcare domain, which is primarily aiming at improving the personalized user-system interaction. This is crucial for promoting patient engagement and compliance, both of which are necessary for achieving long-term behavioral changes and adaptation of a healthier life-style (6). A virtual interface (e.g., virtual environments, virtual, or augmented reality) can be more effective in creating a more enjoyable experience for the end user, while offering the opportunity to create bounding conditions between virtual objects or human-like avatars. The later can significantly affect the quality of the user-system interaction and adherence to intervention (7). Virtual coaching helps in creating a more immersive experience, which in turn makes patients being more focused in setting and achieving higher goals, instead of following a more lightly-hearted approach.

Virtual coaching can be also most affective in the ends of the age spectrum; children, adolescents, and younger adults on the one end and elders on the other. The first are typically the most aware regarding the latest technological trends and most actively willing to be exposed to new and enticing approaches, including virtual scenarios and gamified coaching solutions. To some extent this is also true for older adults on the other end, but virtual coaching is primarily beneficial for the ease of interpretation it can provide and the enhanced guidance in every step of the intervention. Considering that aging is directly associated with the majority of the neurological disorders that lead to cognitive impairment, coaching strategies for this population group are most often limited in terms of the level of complexity that can be supported (8). The usefulness of the straightforward interactions of a virtual system are also beneficial for young patient populations, who experience equally challenging cognitive and learning difficulties (e.g., Autism spectrum disorders) (9, 10).

Previous review studies in the field of virtual coaching have been conducting focusing on systems that promote self-care (11), physical activity and rehabilitation (12), managing insomnia (13), self-tracking and e-coaching for healthy lifestyles (14), conversational agents and avatars in psychology (15). The main objective of this review is to examine the current status and trends of virtual coaching systems in healthcare, with an emphasis on systems that provide automated patient evaluations and appropriate feedback through virtual environments based on the information being gathered. The primary outcome of this study is to showcase automated coaching systems, use cases of active patient monitoring solutions based on sensing data from IoT devices for advanced feedback, the inclusion of automated real-time evaluations of sensor data (e.g., posture and gait analysis as a subject is moving), and the initiation of appropriate intervention from the virtual coach based on this information (e.g., update instructions, dynamically guide and educate the patient, showcase correct movement when an inappropriate one is detected, etc.). A literature review over the past 10 years is performed to assess recent advancements in the adaptation of such novel technologies for virtual coaching systems that are designed to operate independently in home (i.e., uncontrolled environment), and to investigate future trends based on in lab patient monitoring and coaching technologies that are still under evaluation in feasibility-level studies.

METHODS

Literature Search Overview

Within this section, the methodology for conducting the scoping review is detailed. More specifically, starting from the research question "Which is, nowadays, the level of validity and maturity of virtual coaching systems and how close are we to fully automated closed-loop systems?," the methodology proposed in Arksey and O'Malley (16) was followed. The main concept of the present study refers to virtual coaching systems within the medical arena, which collect in an automated way data from the end user (e.g., motion data, speech, gestures, etc.) and, through an inference mechanism, provide feedback related to the scope of each system. Ideally, the coaching systems should be adjustable to each user (i.e., personalized feedback) and be designed to dynamically adapt to its sensing input data, user feedback or environmental conditions. In addition, this review will showcase the use of real-time evaluations and coaching feedback, as this functionality is necessary for more advanced smart virtual coaching systems. Real-time refers to the system's ability to be constantly aware of both the user's condition and its environment, dynamically evaluate his/her actions, behavior, movement along with different environmental factors (i.e., real-time evaluations), and being able to timely respond to such changes as they are being registered with appropriate interventions (i.e., real-time user feedback), to ensure compliance, correct system use as intended, and provide motivation.

The collection of the relative studies was limited to a time range from January 2010 and until April 2020, utilizing three online databases; PubMed, Scopus, and IEEE Xplore. Within this context, a panel of three experts chose the most relevant keywords, which were used for querying the aforementioned databases. The selected keywords were "virtual coach," "virtual coaching," "virtual trainer," "virtual therapist," "virtual nurse," "virtual intervention," "persuasive system," "virtual reality coaching," "augmented reality coaching," "assistive robotics coaching," "avatar coaching," "e coach," and "robotic coaching," and were used to produce the search terminology for this review, by considering all possible variations. Using these keywords, appropriate search queries were formulated, according to the specifications of each database. It should be noted that the literature review for virtual coaching systems was focused on the medical and clinical domain. In addition, the research papers had

to be written in English in order to be included in this review. The type of publication was not considered as a limitation, and all studies that were either published in international journals or conference proceedings were included. The next steps involved the identification of the relevant studies and screening of the search results using a set of exclusion criteria as described below.

Study Exclusion Criteria

After collecting the literature, three of the authors (i.e., KT, VT and DG) independently screened the titles and the abstracts of all papers, aiming to apply a set of specific exclusion criteria. These criteria involved four main pillars:

- **Domain**: Only coaching systems relating to clinical and medical topics were considered. Thus, coaching systems in the area of e.g., sports were excluded.
- **End users**: Systems should target users who face a morbidity. Thus, any other target group (e.g., medical professionals) were excluded from the next phase.
- **Scope**: The main scope of the presented system should be active virtual coaching. Thus, systems in the domain of telehealth, e-learning, etc. were excluded. Also, systems that did not include any automated inference and feedback mechanism were excluded. The absence of real-time evaluations or real-time coaching feedback was not considered as a factor for excluding a study. Finally, review papers were also excluded.
- **Validity**: Systems with no evaluation/validation study were excluded. In addition, studies reporting findings of coaching systems that were tested on different subject populations than their intended end users were also excluded (e.g., if a post-stroke rehabilitation system was evaluated with healthy individuals).

The review panel met three times, in order to discuss the exclusion criteria and form a consensus on the next steps. During the last meeting, a test among the authors was conducted by randomly selecting ten research papers, from the dataset created in the previous phase. Each author had to exclude all papers of the sample provided that met one or more of the exclusion criteria. All authors pointed out the exact same studies for exclusion. All studies that passed the exclusion criteria were forwarded to the next phase, where a full-text review was conducted, as discussed in the following section.

RESULTS

Selected Studies Overview

The literature search process from the three databases resulted in a total of 1,266 research papers. From these studies, 71.2% of them were retrieved by the Scopus database, 17.3% from the PubMed database, and 11.5% from the IEEE Xplore database. The steps of the screening methodology are presented in **Figure 1**, along with the number of studies that were excluded in each step. The first step consisted of removing duplicate records of common studies that were found across the three databases. All unique studies remaining were then considered in the initial title and abstract-based screening following the proposed exclusion criteria (i.e., section Study Exclusion Criteria). At this state, the main exclusion criteria applied involved the domain and the scope criteria. Full text assessment of the remaining papers was then performed. Each one of the three authors involved assessed a total of 50 papers, thus a subset was reviewed more than once to control for inconsistences in the revision process among the panel. At this state, the most relevant exclusion factor was validity, as many studies included only preliminary methods and results, or presented coaching systems without any evaluation results (e.g., studies at pre-pilot state), or systems that were evaluated with subjects different to the intended patient populations.

The final subset of the 41 studies selected was then categorized into three classes depending on the targeted evaluation environment and equipment used: (i) home-based systems (56.1%), (ii) lab/hospital-based systems (36.6%), and (iii) systems including assistive robotics (7.3%). **Figures 2**, **3** present a quantitative schematic representation of the selected papers. More specifically, **Figure 2** presents the distribution of papers over time and also separated per class (i.e., home-based, lab/hospital-based, assistive robotics). An increasing trend in the number of studies per year, and therefore virtual coaching systems developed, can be observed within the last decade, and in fact 75.6% of the 41 studies included were published during the second half of the decade. This is a strong indicator that during the next few years, the number of virtual coaching systems will likely further increase. The number of studies per clinical/medical domain is presented in **Figure 3**, also separated per class.

In Home Virtual Coaching Systems

Table 1 presents remote virtual coaching systems that were designed and developed to operate at in home environments, as these are prime candidates to showcase the trends and advances in developing state-of-the-art automated models, which are capable of taking advantage of IoT solutions and cloud-edge computing interactions. There are 9 studies presenting coaching systems that focused on promoting and improving the levels of physical activity, with 5 of them focusing on young and middle-aged adults (19, 20, 22, 23, 25), 3 targeting physical activity motivation for the elder population (21, 26, 27), and one study focusing on families with overweight children (24). These applications consisted the typical use case for including continuous active subject monitoring with wearable sensing technology. Systems aiming at reducing the risk of developing type 2 diabetes and weight loss by increasing the level of physical activity in prediabetes subjects and overweight children, respectively, had also implemented continuous monitoring technologies in their intervention programs (24, 33). In most cases, the level of activity was determined by the number of daily steps, while the virtual coaches provided motivational content to help users reach their targeted goals as set during the intervention, in an attempt to keep them engaged in active exercise scheduling, motivate behavioral and lifestyle changes (i.e., especially toward healthier dietary habits) and increase adherence.

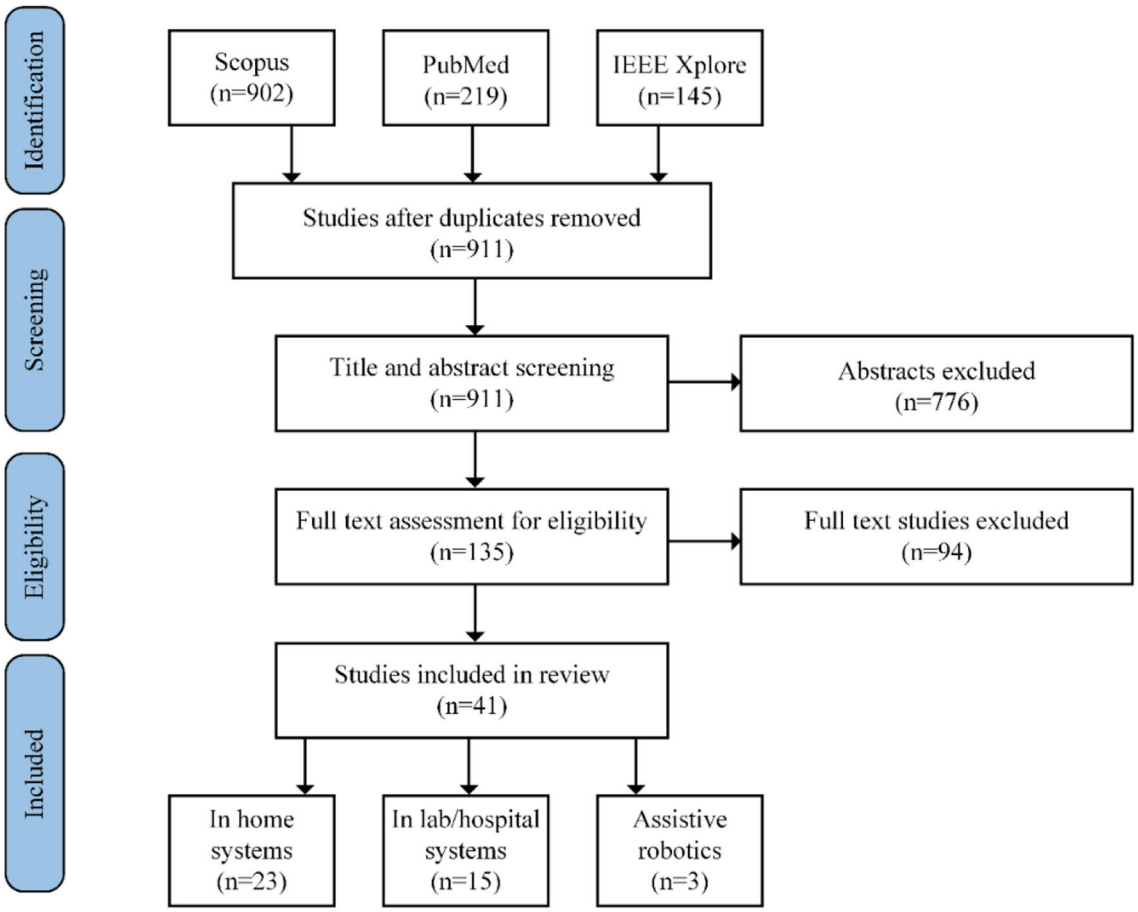

FIGURE 1 | Literature search and screening methodology along with paper selection outcome.

Insomnia was found to be another common domain for healthcare coaching systems. All interventions relied on conversational systems, delivering personalized advice either through animated coaching avatars (28, 30), or through automated text dialogs (i.e., chatbots) (29). None of these coaching systems included active sleep monitoring and analysis, but relied on self-reported sleep diaries and user information. Availability of sensor-based sleep tracking technologies however is on the rise as it is currently supported by the majority of commercial smart watch manufacturers (e.g., Garmin, Fitbit, Withings, Samsung, Polar, etc.), and such systems could find their way in similar studies in the future, as sleep tracking algorithms continue to mature. The same means of user-coach interaction was used for the two systems that were developed to assist with smoking cessation in Veterans (32) and young smoking adults (31). A similar system based on interactions with virtual human avatars was also developed to develop a self-management program that could assist older women with overactive bladder symptoms, improving quality of life and perceived symptoms severity (37). The primary focus of these systems was to raise awareness and promote behavioral changes for a more healthy lifestyle overall, a concept that led to holistic approaches with conversational agents that targeted the promotion of physical activity, healthy diet and effective stress coping as in the study of (25).

Real-time feedback on the other hand was included in only a few studies. Baez et al., created a virtual gym environment to address both aspects of physical training and social user interactions between older adults, and used wearable sensors to measure adherence by tracking exercise completion (27). Rehabilitation training consisted another medical domain where real-time feedback was provided to the user, for exercise quality evaluation through tracking posture and body movement during home-based physiotherapy programs for patients after total knee arthroplasty (18). More common was the use of heart rate monitoring and cardiac-control during exercise. Coaching systems that were guiding the level of activity based on heart rate monitoring for improved rehabilitation were developed and tested for patients with myocardial infarction (17), and patients diagnosed with Parkinson's disease (36). Focusing on younger adults, Segerståhl and Kukkonen, developed a system that could be used to guide exercise execution and keep the users within a certain cardiac level (i.e., training intensity level), using a

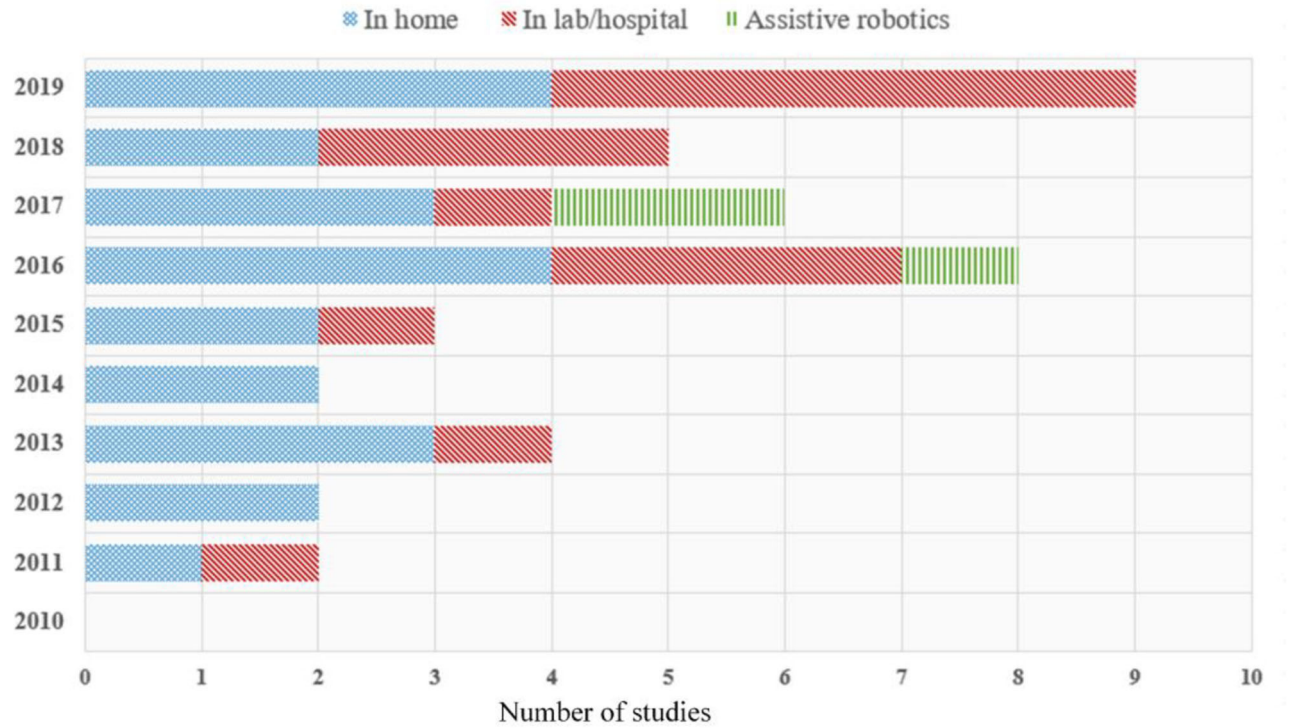

FIGURE 2 | Distribution of selected research papers according to publication year.

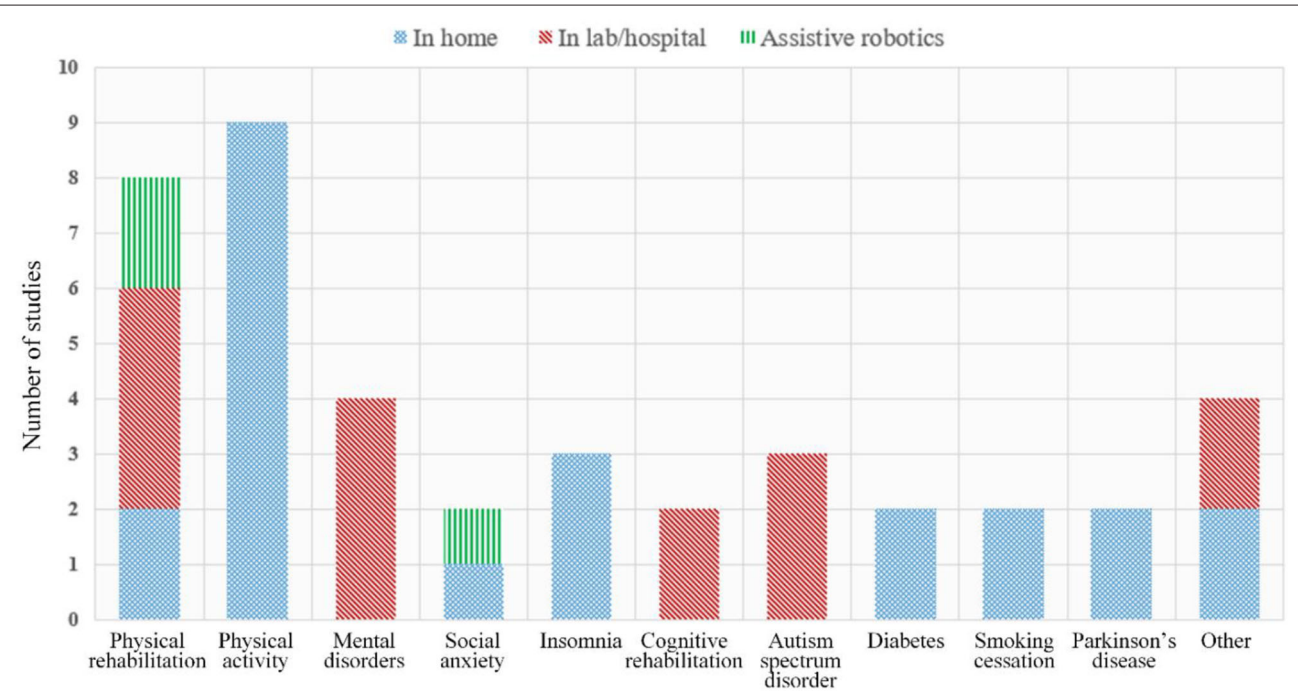

FIGURE 3 | Distribution of selected research papers according to medical domain.

TABLE 1 | Coaching systems that were designed for and evaluated in home environment.

Study	Healthcare domain	UI technologies	Coach-user interaction	Subject monitoring technologies	Continuous monitoring	Real-time evaluations and feedback	Number of subjects	Duration of trial	Evaluation outcomes
Salvi et al. (17)	Physical rehabilitation	Tablet	Motivational virtual coach avatar	ECG, heart rate, respiration rate, accelerometers	-	Heart rate feedback for cardiac rehabilitation	118 adults with myocardial infarction	21 weeks	Improved educational levels but not adherence or exercise habits significantly
Prvu Bettger et al. (18)	Physical rehabilitation	On board screen	Virtual coach avatar to demonstrate and guide activity	3D pose and motion tracking	-	Feedback on exercise quality	287 patients after total knee arthroplasty	12 weeks	Significantly lower health-care costs while providing similar effectiveness
Segerståhl and Oinas-Kukkonen (19)	Physical activity	Wristband display and PC	Exercise level guidance	Heart rate		Heart rate feedback	30 young adults	3 weeks	Feasibility, usability and acceptance of technology study
Watson et al. (20)	Physical activity	PC	Virtual embodied exercise coach	Digital pedometer	Daily number of steps	-	70 overweight adults	12 weeks	Effective in increasing short-term physical activity, long-term effects need further evaluation
Bickmore et al. (21)	Physical activity	Tablet	Virtual embodied conversational exercise coach to motivate walking	Digital pedometer	Daily number of steps	-	263 sedentary older adults	6 months	Effective in increasing short-term physical activity, but problematic long-term maintenance of behavior change
Friederichs et al. (22)	Physical activity	Online program	Motivational interviewing with coach avatar	-	-	-	500 Dutch adults	1 month	If avatar failed to strengthen social relationship with user the impact of the intervention was not enhanced
op den Akker et al. (23)	Physical activity	Smartphone and PC	Feedback by virtual human coach or text	Activity tracker, digital pillbox	Amount of physical activity	-	43 adult office workers	6 weeks	Virtual coach had lower effectiveness compared to text messaging
Gabrielli et al. (24)	Physical activity	Smartphone	Virtual coaching for behavioral change	Activity tracker	Daily number of steps	-	6 families with overweight children	6 weeks	Acceptable solution to support health promotion interventions in primary care
Fadhil et al. (25)	Physical activity	Smartphone	Text-based conversational agent (chatbot)	-	-	-	19 sedentary adults	4 weeks	Preference of using virtual human agent or a combination depends health domain
Brandenburgh et al. (26)	Physical and social activity	Tablet and PC	Virtual coach providing motivational feedback	Accelerometers	Amount of daily activity	-	7 lonely older adults	6 weeks	Reduced loneliness scores
Báez et al. (27)	Physical and social activity	Tablet	Virtual gym environments, personalized avatars, gym coach avatar	Accelerometers, barometric pressure	-	Participation and completion of exercises	34 older adults	10 weeks	High usability and technology acceptance of virtual fitness environment, but poor real-time user interaction
Espie et al. (28)	Insomnia	Online program	Animated personal therapist with personalized advice	-	-	-	164 adults with insomnia	6 weeks	Modest superiority over placebo on daytime sleepiness, improved sleep-wake functioning
Beun et al. (29)	Insomnia	Smartphone	Health coaching dialog system (chatbot)	-	-	-	151 adults with mild insomnia	6 weeks	Significantly more patients reached a meaningful clinical change on Insomnia severity

(Continued)

TABLE 1 | Continued

Study	Healthcare domain	UI technologies	Coach-user interaction	Subject monitoring technologies	Continuous monitoring	Real-time evaluations and feedback	Number of subjects	Duration of trial	Evaluation outcomes
Lorenz et al. (30)	Insomnia	Online program	Animated conversational coach	-	-	-	56 adults with insomnia	6 weeks	Online CBT had substantial long-term effects on relevant sleep-related outcome parameters
An et al. (31)	Smoking cessation	Online program	Personalized virtual avatar, online peer coach	-	-	-	1698 young adults	12 weeks	Increased smoking abstinence and positive changes in health behaviors
Abdullah et al. (32)	Smoking cessation	Tablet	Virtual conversational agent	-	-	-	6 Veterans	2 weeks	Feasibility study, agent helped in setting a quit date.
Connelly et al. (33)	Diabetes	Online program	Virtual conversational coach	Accelerometers	Level of physical activity	-	31 patients with type 2 Diabetes	6 months	Effective in promoting physical activity, interactive features did not affect activity
Block et al. (34)	Prediabetes	Online program, smartphone	Automated physical activity and eating suggestions and logging	-	-	-	339 adults at risk for developing diabetes	12 months	Improved glycemic control, body weight, BMI, waist circumference, TG/HDL ratio, and diabetes risk
Ellis et al. (35)	Parkinson's disease	Tablet	Virtual embodied conversational exercise coach to motivate walking	Digital pedometer	Daily number of steps	-	20 patients with PD	1 month	High retention, satisfaction and adherence to daily walking, significant improvement in mobility
van der Kolk et al. (36)	Parkinson's disease	PC	Virtual cycling targeting heart rate zone	Static bike with heart rate	-	Heart rate feedback	130 adults with mild PD	6 months	Aerobic exercise at home is feasible and reduced off-state motor signs
Andrade et al. (37)	Overactive bladder	Online program	Virtual human avatars and self-avatar	-	-	-	41 adult women	12 weeks	Improved QoL and overactive bladder symptoms
Hartanto et al. (38)	Social anxiety disorder	Head mounted display	Virtual health agent, dialogues with virtual characters	Head tracker, heart rate, microphone	-	Natural speech recognition, heart rate feedback	5 adults with social phobia	10 sessions	Anxiety was evoked and over time gradually decreased due to exposure therapy, but serious technical problems occurred
Wu et al. (39)	Powered wheelchair seating	Smartphone (mobile app)	Virtual seating coach promoting powered seating functions	Accelerometers	Seating angle monitoring	Reposition adjustment	5 powered wheelchair users	5 days	Feasibility study, accelerometers were unstable for seating angle while moving

UI, User Interface; Subject monitoring technologies: Type of sensing data/devices used, Continuous monitoring: Always-on tracking features, Real-time evaluations and feedback: Dynamic user evaluations using sensing data and provision of appropriate interventions.

wearable heart rate monitoring device (21). A wearable heart rate sensor was also implemented in another system that targeted virtual interventions for subjects with social anxiety disorders using virtual health agent and characters to simulate stressful immersive environments, where patients could train to improve their social behavior skills (38). Finally, in a more specialized medical application, Wu et al. developed a prototype coaching system that could assist powered wheelchair users to take full advantage of the provided powered seating functions (e.g., tilt in space, backrest recline, leg rest elevation), following clinically recommended adjusting protocols to maintain seating stability and prevent prolonged-seating complications (39).

Coaching Systems for Hospital/Lab Environment

This section includes 15 studies showcasing virtual coaching systems that were designed and evaluated in controlled hospital or lab environments, or virtual coaching systems that were designed for in home environment, but were evaluated in hospital/lab settings. As it shown in **Table 2**, 11 of these studies consisted of preliminary system evaluations focusing on feasibility and technology acceptance, performing single supervised experimental sessions in small patient populations. The primary focus of the virtual coaching systems that were tested was physical rehabilitation and psychological training for subjects with cognitive impairment, phobias, or mental disorders. Since the experiments were conducted under medical supervision, the clinical use cases extended to patients with more severe physical, mental or cognitive deficits, who were commonly excluded from studies and systems targeting in home evaluations, as the intervention could lead to uncontrolled and hard to predict adverse events. As expected, technologies for continuous subject monitoring could not be properly evaluated under these experimental conditions and, thus, none of these studies reported such measurements. On the other hand, real-time evaluations were extensively integrated in the proposed systems and tested in these studies, with "live" user feedback functionality being the key differentiating factor compared to the in home systems presented in the previous section.

Virtual reality (VR) wearable headsets were introduced into different medical domains, along with more advanced movement tracking and speech recognition systems. Freeman et al. used immersive VR environments to enhance psychological therapy interventions for adults with height phobias by simulating different scenarios of high altitude (48), while subjects with spider phobia were also introduced to a VR system that provided exposure therapy and educational material to help them cope with their fear (52). Systems offering social training for children with Autism Spectrum Disorders (ASD) (51), and physical rehabilitation for post-stroke patients (42) have also been evaluated, showing positive outcomes of the feasibility and acceptability of the immersive VR technology in medical coaching applications. Augmented reality (AR) devices in the form of smart glasses were also tested with children suffering from ASD, as a means to assist them in directing their attention toward a human companion during conversation scenarios, and promote social activities (49, 50).

As it is shown in **Table 2**, the controlled environment and the supervised experiments allowed researchers to evaluate more complex sensing and monitoring technologies. Virtual coaching and gamification of physical rehabilitation training with advanced movement tracking and real-time head and body movement evaluations were implemented using commercially available devices, including Microsoft's Kinect, Nintendo's Wii platform and its peripherals (40, 41, 43). In addition, in the cases where VR was also part of the intervention, head, and hand movement tracking was also performed using the included proprietary trackers of the commercial VR systems, or the integrated IMU sensor if a smartphone attached to a head-mounted device was used instead. Eye tracking through commercial smart glasses was also implemented in AR systems targeting children with ASD, in order to train them to direct their attention toward a fellow conversational companion and engage more in social activities (49, 50).

Despite being in a more laboratory state of technology readiness, custom tracking solutions were less common and only a couple of studies invested in alternative solutions. In a proof of concept study targeting cognitive rehabilitation, a sensing glove was utilized to test feasibility of a more realistic hand and finger movement tracking system with tactile feedback, while performing an intervention focusing on virtually simulated everyday tasks (e.g., meal preparation), showing mixed results in post-stroke patients (46). In another system for post-stroke rehabilitation developed by Lupu et al., biosignal analysis was used for eye movement tracking and motor imagery training, using electrophysiological data captured while recording patient's electrooculogram (EOG) and electroencephalogram (EEG) signals (42). As it shown by their inclusion in a single study, the use of such technologies for medical applications is currently in its infancy state, as the acquisition of biosignals can only be implemented under direct medical supervision and assistance.

Systems supporting natural speech recognition and text semantics analysis in the form of automated conversational agents either through virtual avatars of chatbots, respectively, were also found to gain research attention for coaching purposes. They were primarily used in systems targeting patients with mental and cognitive disorders, offering healthcare and motivational coaching, in order to help them overcome phobias, mental health care, and train in social activities (44, 45, 47, 53). Speech recognition was commonly used to complement systems with immersive VR, in order to provide an enhanced user-system interaction experience (48, 51).

Assistive Robotics in Coaching

This section describes the use of external assistive technologies and robotic-based interventions as part of the coaching process. As expected, these systems were evaluated under supervised conditions in lab environment, but are distinguished by the studies reported in section Coaching Systems for Hospital/Lab Environment above, since the added robotic devices enhance the realism of the interventions, especially when it is complemented with automated audio feedback and

TABLE 2 | Coaching systems that were evaluated in controlled laboratory or clinical environment under direct supervision.

Study	Healthcare domain	UI technologies	Coach-user interaction	Subject monitoring technologies	Continuous monitoring	Real-time evaluations and feedback	Number of subjects	Duration of trial	Evaluation outcomes
Pirovano et al. (40)	Physical rehabilitation	TV screen	Gamification of training with virtual therapist	Kinect, Wii Balance Board, Tymo Therapy Plate	-	Motion capture and body balance	7 older adults with multi-morbidities	1 session	Feasibility and usability study
Michel et al. (41)	Physical rehabilitation	TV screen	Gamification training with motivational messages from animal virtual coach	Kinect	-	Hand movements	20 children with Cerebral Palsy	4 sessions	Usability and acceptance of technology, game was enjoyable
Lupu et al. (42)	Physical rehabilitation	VR headset, PC monitor	VR avatar of a physiotherapist	Electrical stimulator in arms, EOG, EEG	-	Eye tracking (EOG), motor imagery hand movements	7 adults with post stroke central neuromotor syndrome	3 sessions	Feasibility and technology acceptance study
Pham et al. (43)	Physical rehabilitation	TV screen	Jintronix games and therapy modules	Kinect	-	Motion capture	20 hospitalized adults with burn injuries	1 session	Feasibility study, demonstrated good acceptability and safety in hospital environment
Bell and Weinstein (44)	Mental disorders	PC monitor	Virtual conversational agent providing feedback on interview responses	Headset with microphone	-	Speech recognition	10 subjects with psychiatric disabilities	1 session	Strongly positive response to simulated job interview skill training
Cameron et al. (45)	Mental disorders	PC monitor	Chatbot providing self-assessment and tips for stress, anxiety, depression, sleep, and self esteem	-	-	-	7 employees from a mental health social enterprise	1 session	Enjoyable, easy to use and consistent, but poor error management and intelligence
Fok et al. (46)	Cognitive rehabilitation	PC monitor	Audible and visual cues in virtual environment of a kitchen	Data gloves	-	Position and tactile information of hands	2 post-stroke patients with cognitive impairment	1 session	Observing therapists endorsed the system, but larger population study and improvements needed
Wirzberger et al. (47)	Cognitive rehabilitation	TV screen	Dialogue-based memory training with virtual agent	Microphone	-	Speech recognition (experimenter)	62 older adults	1 session	Training performance correlated with recall, longer system response times showed performance benefits
Freeman et al. (48)	Psychological therapy	VR headset	Avatar coach in VR environment	Microphone, head and hand VR trackers	-	Speech recognition, hand movements	100 adults with fear of heights	4 weeks	VR can increase treatment provision for mental health disorders
Liu et al. (49) Vahabzadeh et al. (50)	Autism spectrum disorder	AR headset	Interaction in augmented reality environment	Smart glasses	-	Gaze attention, head movement	10 children with ASD and their caregivers	1 session	Initial evidence in reducing hyperactivity, inattention and impulsivity in children with ASD
Rosenfield et al. (51)	Autism spectrum disorder	VR headset	Social training in VR environment	VR tracker, microphone	-	Head movement, speech recognition	2 children with ASD	1 session	Feasibility and technology acceptance study
Miloff et al. (52)	Exposure therapy	VR headset	Virtual therapist delivered voiceover psychoeducation and supportive feedback	Smartphone IMU sensors	-	Head movement	100 adults with spider phobia	1 session (52 weeks follow up)	VR therapy reduced short-term symptoms, no long-term advantage over *in-vivo* therapy
Kang et al. (53)	Virtual nurse	PC monitor	Virtual nurse offering healthcare and lifestyle advices	Microphone	-	Speech recognition	26 older adults	1 session	Virtual nurses based on adaptive persuasion models resulted in higher social presence and lower frustration
Shamekhi et al. (54)	Self-care management	PC monitor	Conversational agent as virtual coach	-	-	-	9 adults with spinal cord injury	1 session	Feasibility and technology acceptance study

UI, User Interface; Subject monitoring technologies: Type of sensing data/devices used, Continuous monitoring: Always-on tracking features, Real-time evaluations and feedback: Dynamic user evaluations using sensing data and provision of appropriate interventions.

speech recognition capabilities. Robotic rehabilitation systems could enhance the sensing input for the patients, providing force and haptic feedback, while interactions with actual robotic coaches and older adults that were used in socially assistive robotic systems, were found to be useful in improving social, physical, and cognitive training programs. The later was shown to be very effective in a recent study, where a commercial robot (i.e., NAO robot) was used as a coach to assist and provoke social interactions in older adults with and without cognitive impairment (55). The robotic coach was designed to provide one-on-one coach-user interaction, as well as multi-user and coach interactions, with individualized activity management and dynamically adaptive behavior for long-term engagement and was very received during its initial testing.

Then, there were also two feasibility evaluation studies that focused on robotic physical rehabilitation. Despite targeting different age population groups and medical conditions, both studies resulted in similar findings; showcasing the proposed coaching systems as useful training tools to complement the existing approaches in rehabilitation therapy. Chiang et al., used force feedback from two haptic devices to enhance real-time user-system interactions in non-immersive VR environments, training children and adolescents with upper limb disabilities to improve performance in activities of daily living, showing high level of satisfaction when using virtual rehabilitation systems (56). An arm exoskeleton that provided gravity-compensation in arm movements and dynamically adapted feedback in non-immersive VR environment was also tested for personalized rehabilitation in post-stroke patients with severely affected motor functionality (57). A closed-loop system that facilitated unsupervised motor learning providing adjustable level of supported depending on each patient's motor potential, and continuous visual feedback of a virtual arm was proposed to complement the perception of natural movement and improve the quality and range of the functional restoration.

DISCUSSION

As **Figure 3** suggests, 17 studies of the total of 41 included in this review of virtual coaching systems dealt with either physical rehabilitation or physical activity and motivation of healthier lifestyle choices. This was also found to be the most prominent field for using sensing devices as 15 out of the 17 systems included patient monitoring, especially with the capacity to capture level of exercise with activity trackers, accelerometers, and pedometers, and relevant body movement using depth cameras for precise posture and skeleton tracking, or even haptic and exoskeleton devices. Systems for motivation of physical activity could also be more easily implemented in larger patient populations as they did not engage specific and detailed real-time evaluations and relied mostly on user's self-reported input. Coaching systems for mental disorders and cognitive rehabilitation were also found to gain research attention in the past few years, but, as shown in **Table 2**, their applications were limited only within supervised testing conditions and have not yet been transported to independent, in home use scenarios. In addition, another notable remark refers to the under-representation of clinical conditions, like Parkinson's disease, cardiovascular diseases, ASD, diabetes, and orthopedics for which technology-aided virtual coaching systems should have a substantial impact to quality of life and everyday activities.

Integration of new technologies for in home coaching systems was also found to be rather limited, which is an alarming outcome for the field, overall. For example, immersive VR and AR interaction solutions were not implemented in systems that targeted in home use cases, and were rather only found in controlled lab environment. This is an indicator, however, that these technologies are currently being meticulously tested, and mass adoption in the future is to be expected. Especially as they continue to improve in terms of supported functionality, tracking accuracy and lower costs. The same could be applied to assistive robotics, as positive results from the studies reported in **Table 3** suggest that the robotic-based training has great potential to be used as a way to enhance existing interventions in occupational therapy. Although wide spread adoption is more challenging than VR and AR technologies, there is already evidence that the novelty factor that accompanies assistive robotic interventions, can be very beneficial in specific medical conditions such as stroke rehabilitation (58).

Subject monitoring and sensing technologies being used in coaching systems seem to be highly dependent on technology readiness levels of the various IoT devices. The level of physical activity as estimated through step tracking is found to be the most common means, which can be explained considering that it is nowadays a mature technology (i.e., in terms of both hardware and software), with increased availability and reduced associated costs, as competing commercial wearable devices pushed advancements in this field. Microsoft's Kinect was another landmark device that enabled advanced functionality in coaching systems, as it provided unobtrusive monitoring of whole body movement and more precise user-system interaction with upper extremities. Despite being discontinued at consumer level, its practicality allowed even non-highly tech educated people to familiarize with this technology and has inspired researchers to develop medical devices and systems that integrate similar functionality, especially in rehabilitation interventions (18, 59). Heart rate monitoring is another field where IoT devices have reached a level of functional robustness and monitoring accuracy that justified their widespread use for clinical use cases in coaching systems (**Table 1**). A similar transition in monitoring technology is currently being materialized following the advancements in smartwatches, which should be shown in more precise physical activity and sleep monitoring and enhanced evaluation feedback. This is the first link that will enable the development of personalized and precise monitoring systems that can seamlessly provide continuous patient tracking in their everyday life environment and living conditions.

The second important link in the chain of more advanced coaching systems is the feedback that they can provide to the user. As the distinction of in home and in lab systems in the results section revealed, the transition from offline and semi-real evaluation of user behavior to real-time feedback for in home environment is currently in its early stages (**Table 1**), despite the widespread use cases of real-time evaluation functionality

TABLE 3 | Coaching systems with assistive robotic functionality.

Study	Healthcare domain	Coach-user interaction	Subject monitoring technologies	Continuous monitoring	Real-time evaluations and feedback	Number of subjects	Duration of trial	Evaluation outcomes
Grimm et al. (57)	Physical rehabilitation	Training with adaptation of task difficulty in virtual environments	Arm exoskeleton (7 degree-of-freedom)	-	Hand movement tracking in interactive games	5 post-stroke patients with upper extremity motor deficits	4 weeks	Gravity-compensation and progressively challenging motor learning can facilitate functional restoration
Chiang et al. (56)	Physical rehabilitation	Interactive user interface with force feedback in virtual environments	Haptic devices (6 degrees of freedom inputs)	-	Interactions in virtual environments	20 students with upper limb dis-abilities	2 months	Useful training tool to complement conventional rehabilitation approaches
Fan et al. (55)	Socially assistive robotics	Interaction with robotic coach	NAO robot, Kinect, EEG, galvanic skin response	-	Gesture and speech recognition, gaze estimation	25 older adults (some with cognitive decline)	1 session	Good system functionality and positive perception of the robotic intervention

Subject monitoring technologies: Type of sensing data/devices used, Continuous monitoring: Always-on tracking features, Real-time evaluations and feedback: Dynamic user evaluations using sensing data and provision of appropriate interventions.

under controlled, supervised conditions (**Table 2**). The findings in **Table 2** however, are indicative of the future trends in the evaluation feedback of coaching systems, and the necessity to close the loop with automated and more precise real-time evaluations. This would ultimately allow the available clinical personnel to effectively supervise larger patient populations in their medical domain of expertise, and increase access to better healthcare services in rural areas.

The increased practical functionality that the improved monitoring technologies offer (section Coaching Systems for Hospital/Lab Environment), could define a new generation of coaching systems that are able to revolutionize patient perspective, as "smart" systems become progressively aware of user's environment, activities and behavior, improving the quality of user-system interactions and eventually adherence to the intervention protocols. A closed-loop coaching system that can dynamical adjust to each patient's needs, evaluate their behavior and actions in real-time, and intervene accordingly offers a holistic approach in motor and cognitive training and rehabilitation address. This will set a crucial milestone in coaching systems enabling the continuity of healthcare between subsequent visits with medical professionals.

The lack of holistic coaching systems was also acknowledged by the European Commission, which led to several recent projects responding to the H2020 call SC1-PM-15-2017—Personalized coaching for well-being and care of people as they age. The call tried to address the deployment of radically new solutions for personalized virtual coaching systems, building upon intelligent ICT environments, access to relevant physiological, and behavioral data and new forms of accessible interaction based on tangible user interaction concepts. Currently, these projects are evaluating the different approaches with the targeted aging population:

HOLOBALANCE[1] introduces a new personalized platform for virtual coaching, motivation, and empowerment of older citizens with balance disorders. The coaching part is realized with a holographic surrogate physiotherapist and augmented reality games, along with easy to use wearable sensors in an interoperable platform design (60).

Council of Coaches[2] introduces a radically new virtual coaching concept based on multiple autonomous, embodied virtual coaches, which form together a personal council addressing the needs of older adults in an integrated manner (61).

WellCo[3] provides an affective-aware coach that interacts through speech with the user in order to act as a virtually interface among the user and the platform managing the flow of all interactions and empower users in their behavior change process through simulation activities tailored to their current mood (62).

CAPTAIN[4] proposes a transparent technology designed to turn the home of the older adult into a ubiquitous assistant, based on

[1]https://holobalance.eu/

[2]https://council-of-coaches.eu/

[3]http://wellco-project.eu/

[4]https://www.captain-eu.org/

a projected augmented reality through use of micro-projectors, contextualized information and instructions on top of the real environment (63).
vCare[4] embeds clinical profiles and the pathways to drive the behavior of the virtual coach at home and provide well-elaborated services for tele-rehabilitation in neurology and cardiology (64).
SAAM[5] supports the aging population living at home, with a novel and practical emphasis on ambient sensing and learning of user needs and preferences, and effective coaching by leveraging the user's social support networks (65).
EMPATHIC[6] proposes multimodal face analytics, adaptive spoken dialogue systems and natural language interfaces along with non-intrusive technologies to extract physiological markers of emotional states in real-time, in order to support dependent aging persons and their caregivers (66).
NESTORE[7] leverages on novel ICT technologies for unobtrusive monitoring system, including wearable and environmental sensors, intelligent Decision Support System to provide personalized goals toward wellbeing, and active coaching as conversational agents, embodied in a physical companion to establish affective communication and engage user in personalized coaching activities (67).

Finally, it should be noted that besides the large amount of studies that has been considered during the literature search, it is still very likely that some relevant research papers have not been discovered or missed. From the studies that have been included however, it is clear that the current and near future research trends in virtual coaching systems are pushing to close the loop with advanced patient monitoring, improved virtual user-system interactions and most critically, support for automated feedback and dynamically adjusted precise interventions, based on personalized evaluations from the sensing data.

[4]https://vcare-project.eu/
[5]https://www.saam2020.eu/
[6]http://www.empathic-project.eu/
[7]https://nestore-coach.eu/home

CONCLUSION

The systematic evaluation of the literature within this review has shown that virtual coaching systems are gaining momentum, as new monitoring and system-user interaction technologies are being evaluated and their medical applications are expanding in more complex diseases. Technology integration is highly dependent on the current status and supported functionality of IoT sensing devices, with novel patient monitoring and intriguing virtual interfaces becoming increasingly popular in coaching systems, especially in controlled and supervised environments. Transition to fully automated in home coaching systems, however, is proving to be in its early stages and many obstacles still remain before holistic coaching solutions with real-time evaluations and personalized feedback can become widely available for unsupervised in home use. The added value of such systems in terms of effectiveness, technology acceptance, and reduced healthcare costs has been well documented though and proven in proof of concept studies over the past few years. The current research projects that were funded to address these issues are on track to set the foundations for a new generation of coaching systems.

AUTHOR CONTRIBUTIONS

KT, VT, and DG performed the screening process of the research papers. DF performed a final revision of the final manuscript. All authors contributed equally in the preparation of the manuscript, the formulation of the literature search methodology, the exclusion criteria used for this review, revision of the article, and approved the submitted version.

FUNDING

This study was supported by the Holobalance project (https://holobalance.eu/) which is funded within the European Union's Horizon 2020 research and innovation programme under grant agreement no. 769574.

REFERENCES

1. Farahani B, Firouzi F, Chakrabarty K. *Healthcare IoT. Intelligent Internet of Things: From Device to Fog and Cloud*. Firouzi Springer International Publishing. (2020).p. 515–45. doi: 10.1007/978-3-030-30367-9_11
2. Woo MW, Lee J, Park K. A reliable IoT system for personal healthcare devices. *Future Gen Comp Syst*. (2018) 78:626–40. doi: 10.1016/j.future.2017.04.004
3. Dimitrov DV. Medical internet of things and big data in healthcare. *Healthc Inform Res*. (2016) 22:156–63. doi: 10.4258/hir.2016.22.3.156
4. Ojo MO, Giordano S, Procissi G, Seitanidis IN. A review of low-end, middle-end, and high-end Iot devices. *IEEE Access*. (2018) 6:70528–54. doi: 10.1109/ACCESS.2018.2879615
5. Knickerbocker JU, Budd R, Dang B, Chen Q, Colgan E, Hung LW, et al. Heterogeneous integration technology demonstrations for future healthcare, IoT, and AI computing solutions. In: *2018 IEEE 68th Electronic Components and Technology Conference (ECTC)*. (2018) doi: 10.1109/ECTC.2018.00231
6. Kulyk O, op den Akker R, Klaassen R, van Gemert-Pijnen LJ. Personalized virtual coaching for lifestyle support: principles for design and evaluation. *Int J Adv Life Sci*. (2014) 6:300.
7. van Wissen A, Vinkers C, van Halteren A. *Developing a Virtual Coach for Chronic Patients: A User Study on the Impact of Similarity, Familiarity and Realism*. Cham:Springer International Publishing (2016). doi: 10.1007/978-3-319-31510-2_23
8. Klimova B, Valis M, Kuca K. Cognitive decline in normal aging and its prevention: a review on non-pharmacological lifestyle strategies. *Clin Intervent Aging*. (2017) 12:903–10. doi: 10.2147/CIA.S132963
9. Liu X, Zhao W. Buddy: a virtual life coaching system for children and adolescents with high functioning autism. In: *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress*. (2017). doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.62

10. Zhao W, Liu X, Qiu T, Luo X. Virtual avatar-based life coaching for children with autism spectrum disorder. *Computer.* (2020) 53:26–34. doi: 10.1109/MC.2019.2915979
11. Ding D, Liu HY, Cooper R, Cooper RA, Smailagic A, Siewiorek D. Virtual coach technology for supporting self-care. *Phys Med Rehabil Clin N Am.* (2010) 21:179–94. doi: 10.1016/j.pmr.2009.07.012
12. Tropea P, Schlieter H, Sterpi I, Judica E, Gand K, Caprino M, et al. Rehabilitation, the great absentee of virtual coaching in medical care: scoping review. *J Med Internet Res.* (2019) 21:e12805. doi: 10.2196/12805
13. Salvemini AG, D'Onofrio, Ciccone F, Greco A, Tullio A, et al. Insomnia and information and communication technologies (ICT) in elderly people: a systematic review. *Med Sci.* (2019) 7:70. doi: 10.3390/medsci7060070
14. Lentferink AJ, Oldenhuis H, de Groot KEM, Polstra L, Velthuijsen H, van Gemert-Pijnen JEWC. Key components in ehealth interventions combining self-tracking and persuasive ecoaching to promote a healthier lifestyle: a scoping review. *J Med Internet Res.* (2017) 19:e277. doi: 10.2196/jmir.7288
15. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. *J Med Internet Res.* (2017) 19:e151. doi: 10.2196/jmir.6553
16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soci Res Methodol.* (2005) 8:19–32. doi: 10.1080/1364557032000119616
17. Salvi D, Ottaviano M, Muuraiskangas SA, Martínez-Romero C, Vera-Muñoz MF, Cabrera Umpiérrez MT, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. *J Telemed Telecare.* (2018) 24:303–16. doi: 10.1177/1357633X17697501
18. Prvu Bettger J, Green CL, Holmes DN, Chokshi A, Mather R3rd C, Hoch BT, et al. Effects of virtual exercise rehabilitation in-home therapy compared with traditional care after total knee arthroplasty: VERITAS. a randomized controlled. *Trial J Bone Joint Surg Am.* (2020) 102:101–9. doi: 10.2106/JBJS.19.00695
19. Segerståhl K, Oinas-Kukkonen H. Designing personal exercise monitoring employing multiple modes of delivery: implications from a qualitative study on heart rate monitoring. *Int J Med Inform.* (2011) 80:e203–13. doi: 10.1016/j.ijmedinf.2011.08.011
20. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. *J Med Internet Res.* (2012) 14:e1. doi: 10.2196/jmir.1629
21. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. *J Am Geriatr Soc.* (2013) 61:1676–83. doi: 10.1111/jgs.12449
22. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a Web-based physical activity intervention with an avatar: randomized controlled trial. *J Med Internet Res.* (2014) 16:e48. doi: 10.2196/jmir.2974
23. Akker HJA, Klaassen R, Nijholt A. *Virtual Coaches for Healthy Lifestyle. Toward Robotic Socially Believable Behaving Systems - Volume II: Modeling Social Signals*. Cham: Springer International Publishing (2016). p. 121–49. doi: 10.1007/978-3-319-31053-4_8
24. Gabrielli S, Dianti M, Maimone R, Betta M, Filippi L, Ghezzi M, et al. Design of a mobile app for nutrition education (trec-lifestyle) and formative evaluation with families of overweight children. *JMIR Mhealth Uhealth.* (2017) 5:e48. doi: 10.2196/mhealth.7080
25. Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. *Methods Inf Med.* (2019) 58:9–23. doi: 10.1055/s-0039-1688757
26. Brandenburgh AW, van Breda W, van der Ham, Klein M, Moeskops L. *Increasing Physical and Social Activity through Virtual Coaching in an Ambient Environment*. Cham:Springer International Publishing (2014). doi: 10.1007/978-3-319-09912-5_41
27. Báez M, Ibarra F, Far IK, Ferron M, Casati F. Online group-exercises for older adults of different physical abilities. In: *2016 International Conference on Collaboration Technologies and Systems (CTS)*, Orlando, FL (2016). doi: 10.1109/CTS.2016.0098
28. Espie CA, Kyle SD, Williams C, Ong JC, Douglas NJ, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. *Sleep.* (2012) 35:769–81. doi: 10.5665/sleep.1872
29. Beun RJ, Fitrianie S, Griffioen-Both F, Spruit S, Horsch C, Lancee J, Brinkman W, et al. Tools: the best of both worlds in mobile user interfaces for E-coaching. *Personal Ubiquitous Comp.* (2017) 21:661–74. doi: 10.1007/s00779-017-1021-5
30. Lorenz N, Heim E, Roetger A, Birrer E, Maercker A. Randomized controlled trial to test the efficacy of an unguided online intervention with automated feedback for the treatment of insomnia. *Behav Cogn Psychother.* (2019) 47:287–302. doi: 10.1017/S1352465818000486
31. An LC, Demers MR, Kirch MA, Considine-Dunn S, Nair V, Dasgupta K, et al. A randomized trial of an avatar-hosted multiple behavior change intervention for young adult smokers. *J Natl Cancer Inst Monogr.* (2013) 2013:209–15. doi: 10.1093/jncimonographs/lgt021
32. Abdullah AS, Gaehde S, Bickmore T. A tablet based embodied conversational agent to promote smoking cessation among veterans: a feasibility study. *J Epidemiol Glob Health.* (2018) 8:225–30. doi: 10.2991/j.jegh.2018.08.104
33. Connelly J, Kirk A, Masthoff J, MacRury S. A website to promote physical activity in people with type 2 diabetes living in remote or rural locations: feasibility pilot randomized controlled trial. *JMIR Diabetes.* (2017) 2:e26. doi: 10.2196/diabetes.6669
34. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. *J Med Internet Res.* (2015) 17:e240. doi: 10.2196/jmir.4897
35. Ellis T, Latham NK, DeAngelis TR, Thomas CA, Saint-Hilaire M, Bickmore TW. Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease. *Am J Phys Med Rehabil.* (2013) 92:472–481; quiz 482–475. doi: 10.1097/PHM.0b013e31828cd466
36. van der Kolk N, de Vries MNM, Kessels R, Joosten PCH, Zwinderman AH, Post B, Bloem BR. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. *Lancet Neurol.* (2019) 18:998–1008. doi: 10.1016/S1474-4422(19)30285-6
37. Andrade AD, Anam R, Karanam C, Downey P, Ruiz JG. An overactive bladder online self-management program with embedded avatars: a randomized controlled trial of efficacy. *Urology.* (2015) 85:561–7. doi: 10.1016/j.urology.2014.11.017
38. Hartanto D, Brinkman WP, Kampmann IL, Morina N, Emmelkamp, P. *Home-Based Virtual Reality Exposure Therapy with Virtual Health Agent Support*. Cham:Springer International Publishing. (2016). doi: 10.1007/978-3-319-32270-4_9
39. WuYK, Liu Y, Kelleher A, Pearlman J, Cooper RA. Evaluating the usability of a smartphone virtual seating coach application for powered wheelchair users. *Med Engin Phys.* (2016) 38:569–75. doi: 10.1016/j.medengphy.2016.03.001
40. Pirovano M, Mainetti R G, Baud-Bovy, Lanzi PL, Borghese NA. Intelligent Game Engine for Rehabilitation (IGER). *IEEE Trans Comp Intel Games*. (2016) 8:43–55. doi: 10.1109/TCIAIG.2014.2368392
41. Michel P, Richard P, Yamaguchi T, Verhulst A, Verhulst E, Dinomais M. Virtual reality system for rehabilitation of children with cerebral palsy: a preliminary study. In: *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Vol. 1. Rome: SciTePress (2016), 306–313. doi: 10.5220/0005720503040311
42. Lupu RG, Irimia DC, Ungureanu F, Poboroniuc MS, Moldoveanu BC. FES based therapy for stroke rehabilitation using VR facilities. *Wireless Commun Mobile Comput.* (2018) 2018:4798359. doi: 10.1155/2018/4798359
43. Pham TN, Wong JN, Terken T, Gibran NS, Carrougher GJ, Bunnell A. Feasibility of a Kinect(®)-based rehabilitation strategy after burn injury. *Burns.* (2018) 44:2080–6. doi: 10.1016/j.burns.2018.08.032
44. Bell MD, Weinstein A. Simulated job interview skill training for people with psychiatric disability: feasibility and tolerability of virtual reality training. *Schizophr Bull.* (2011) 37 Suppl 2(Suppl 2):S91–97. doi: 10.1093/schbul/sbr061
45. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna MS. *Assessing the Usability of a Chatbot for Mental Health Care*. Cham:Springer International Publishing. (2019). doi: 10.1007/978-3-030-17705-8_11
46. Fok SC, Tan F, Lau M. Supervising Cognitive Tele-Rehabilitation with a Virtual Therapist. *Appl Mech Mater.* (2013) 344:213–8. doi: 10.4028/www.scientific.net/AMM.344.213

47. Wirzberger M, Schmidt R, Georgi M, Hardt W, Brunnett G, Rey GD. Effects of system response delays on elderly humans' cognitive performance in a virtual training scenario. *Sci Rep.* (2019) 9:8291. doi: 10.1038/s41598-019-44718-x
48. Freeman D, Lister R, Waite F, Yu LM, Slater M, Dunn G, et al. Automated psychological therapy using virtual reality (VR) for patients with persecutory delusions: study protocol for a single-blind parallel-group randomised controlled trial (THRIVE). *Trials.* (2019) 20:87. doi: 10.1186/s13063-019-3198-6
49. Liu R, Salisbury JP, Vahabzadeh A, Sahin NT. Feasibility of an autism-focused augmented reality smartglasses system for social communication and behavioral coaching. *Front Pediatr.* (2017) 5:145. doi: 10.3389/fped.2017.00145
50. Vahabzadeh A, Keshav NU, Salisbury JP, Sahin NT. Improvement of attention-deficit/hyperactivity disorder symptoms in school-aged children, adolescents, and young adults with autism via a digital smartglasses-based socioemotional coaching aid: short-term, uncontrolled pilot study. *JMIR Ment Health.* (2018) 5:e25. doi: 10.2196/mental.9631
51. Rosenfield N, Lamkin K, Re J, Day K, Boyd L, Linstead E. A virtual reality system for practicing conversation skills for children with autism. *Multimod Technol Interact.* (2019) 3:28. doi: 10.3390/mti3020028
52. Miloff A, Lindner P P, Dafgård, Deak S, Garke M, et al. Automated virtual reality exposure therapy for spider phobia vs. *in-vivo* one-session treatment: A randomized non-inferiority trial. *Behav Res Ther.* (2019) 118:130–40. doi: 10.1016/j.brat.2019.04.004
53. Kang Y, Tan AH, Miao C. An adaptive computational model for personalized persuasion. In: *Proceedings of the 24th International Conference on Artificial Intelligence*. Buenos Aires:AAAI Press (2015). p. 61-67.
54. Shamekhi A, Trinh H, Bickmore TW, DeAngelis TR, Ellis T, Houlihan BV, et al. A virtual self-care coach for individuals with spinal cord injury. In: *Proceedings of the 18th International ACM SIGACCESS Conference on Computers Accessibility*. Reno, NV:Association for Computing Machinery(2016).p. 327-328. doi: 10.1145/2982142.2982199
55. Fan J, Bian D, Zheng Z, Beuscher L, Newhouse PA, Mion LC, et al. A Robotic Coach Architecture for Elder Care (ROCARE) Based on Multi-User Engagement Models. *IEEE Tran.* (2017) 25:1153–63. doi: 10.1109/TNSRE.2016.2608791
56. Chiang VC, Lo KH, Choi KS. Rehabilitation of activities of daily living in virtual environments with intuitive user interface and force feedback. *Disabil Rehabil Assist Technol.* (2017) 12:672–80. doi: 10.1080/17483107.2016.1218554
57. Grimm F, Naros G, Gharabaghi A. Closed-loop task difficulty adaptation during virtual reality reach-to-grasp training assisted with an exoskeleton for stroke rehabilitation. *Front Neurosci.* (2016) 10:518. doi: 10.3389/fnins.2016.00518
58. Klamroth-Marganska V, Blanco J, Campen K, Curt A, Dietz V, Ettlin T, et al. Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial. *Lancet Neurol.* (2014) 13:159–66. doi: 10.1016/S1474-4422(13)70305-3
59. De Luca R, Lo Buono V, Leo A, Russo M, Aragona B, Leonardi S, et al. Use of virtual reality in improving poststroke neglect: Promising neuropsychological and neurophysiological findings from a case study. *Appl Neuropsychol.* (2019) 26:96–100. doi: 10.1080/23279095.2017.1363040
60. Tsiouris KM, Gatsios D, Tsakanikas V, Pardalis AA, Kouris I, Androutsou T, et al. Designing interoperable telehealth platforms: bridging IoT devices with cloud infrastructures. *Enter Inform Syst.* (2020) 2020:1-25. doi: 10.1080/17517575.2020.1759146
61. Hurmuz MZM, Jansen-Kosterink SM, den Akker HO, Hermens HJ. User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. *JMIR Res Prot.* (2020) 9:4. doi: 10.2196/16641
62. Manea V, Wac K. mQoL: mobile quality of life lab: from behavior change to QoL. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. Singapore: Association for Computing Machinery (2018).p. 642–7. doi: 10.1145/3267305.3267549
63. Petsani D, Mantziari D, Zilidou V, Konstantinidis EI, Billis A, Timoleon M, et al. Co-design the future CAPTAIN system with older adults: focusing on the e-coaching dimensions. In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. Rhodes:Association for Computing Machinery(2019).p. 639-644. doi: 10.1145/3316782.3322765
64. Philipp P, Merkle N, Gand K, Gißke C. Continuous support for rehabilitation using machine learning. *Inform Tech.* (2019) 61:273. doi: 10.1515/itit-2019-0022
65. Dimitrov Y, Gospodinova Z, Wheeler R, Žnidaršič M, Ženko B, Veleva V, Miteva N. Social activity modelling and multimodal coaching for active aging. In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. Rhodes:Association for Computing Machinery (2019).p. 608–15. doi: 10.1145/3316782.3322761
66. Justo R, Ben Letaifa L, Palmero C, Gonzalez-Fraile E, Torp Johansen A, Vázquez A, et al. Analysis of the interaction between elderly people and a simulated virtual coach. *J Ambient Intell Human Comput.* (2020) doi: 10.1007/s12652-020-01983-3
67. Guarneri MR, Mastropietro A, Caon M, Maldonado LF, Furfari F, Andreoni G, et al. *A Personalised Virtual Coach to Counteract Ageing Decline: The H2020 NESTORE Project*, Cham:Springer International Publishing (2019). doi: 10.1007/978-3-030-05921-7_1

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 08 October 2020
doi: 10.3389/fdgth.2020.567158

Detection of Mild Cognitive Impairment Through Natural Language and Touchscreen Typing Processing

*Anastasia Ntracha[1], Dimitrios Iakovakis[1], Stelios Hadjidimitriou[1], Vasileios S. Charisis[1], Magda Tsolaki[2] and Leontios J. Hadjileontiadis[1,3]**

[1] Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, [2] Third Department of Neurology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece, [3] Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Mild cognitive impairment (MCI), an identified prodromal stage of Alzheimer's Disease (AD), often evades detection in the early stages of the condition, when existing diagnostic methods are employed in the clinical setting. From an alternative perspective, smartphone interaction behavioral data, unobtrusively acquired in a non-clinical setting, can assist the screening and monitoring of MCI and its symptoms' progression. In this vein, the diagnostic ability of digital biomarkers, drawn from Fine Motor Impairment (FMI)- and Spontaneous Written Speech (SWS)-related data analysis, are examined here. In particular, keystroke dynamics derived from touchscreen typing activities, using Convolutional Neural Networks, along with linguistic features of SWS through Natural Language Processing (NLP), were used to distinguish amongst MCI patients and healthy controls (HC). Analytically, three indices of FMI (rigidity, bradykinesia and alternate finger tapping) and nine NLP features, related with lexical richness, grammatical, syntactical complexity, and word deficits, formed the feature space. The proposed approach was tested on two demographically matched groups of 11 MCI patients and 12 HC, having undergone the same neuropsychological tests, producing 4,930 typing sessions and 78 short texts, within 6 months, for analysis. A cascaded-classifier scheme was realized under three different feature combinations and validated via a Leave-One-Subject-Out cross-validation scheme. The acquired results have shown: (a) keystroke features with a k-NN classifier achieved an Area Under Curve (AUC) of 0.78 [95% confidence interval (CI):0.68–0.88; specificity/sensitivity (SP/SE): 0.64/0.92], (b) NLP features with a Logistic regression classifier achieved an AUC of 0.76 (95% CI: 0.65–0.85; SP/SE: 0.80/0.71), and (c) an ensemble model with the fusion of keystroke and NLP features resulted in AUC of 0.75 (95% CI:0.63–0.86; SP/SE 0.90/0.60). The current findings indicate the potentiality of new digital biomarkers to capture early stages of cognitive decline, providing a highly specific remote screening tool in-the-wild.

Keywords: Alzheimer's disease, natural language processing, keystroke dynamics, fine motor impairment, machine learning, remote screening, deep learning, smartphone

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Martin Dyrba, Helmholtz Association of German Research Centers (HZ), Germany
Anake Pomprapa, RWTH Aachen University, Germany

***Correspondence:**
Leontios J. Hadjileontiadis
leontios.hadjileontiadis@ku.ac.ae

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *29 May 2020*
Accepted: *27 August 2020*
Published: *08 October 2020*

Citation:
Ntracha A, Iakovakis D, Hadjidimitriou S, Charisis VS, Tsolaki M and Hadjileontiadis LJ (2020) Detection of Mild Cognitive Impairment Through Natural Language and Touchscreen Typing Processing. Front. Digit. Health 2:567158. doi: 10.3389/fdgth.2020.567158

1. INTRODUCTION

Mild Cognitive Impairment (MCI) (1, 2) affects 20% of people over 65 years old worldwide, causing cognitive decline, beyond normal aging, especially in the areas of memory and executive functions, while increasing the probability for the manifestation of the neurodegenerative Alzheimer's disease (AD) (3). People diagnosed with MCI have to be constantly tested over time to ensure that they have not transitioned from mild to severe dementia (4). Although there is currently no treatment for AD, early diagnosis (5) is essential so that the patients, their family and caretakers are better prepared, by making the necessary financial and legal decisions and testing different lifestyle interventions and medication in the MCI or another prodromal stage, to potentially delay or even prevent AD development (6). Studies have indicated that cognitive, behavioral, sensory and motor changes may precede clinical manifestations of AD by several years (7). Current diagnostic methods include a suite of neuropsychological and physiological tests conducted in a clinical setting, estimating key cognitive functions like memory, comprehension, and coordination, alongside laboratory and brain-imaging tests, estimating nerve degeneration and amyloid protein concentration levels (8). However, these validated AD diagnostic methods usually fail to efficiently distinguish early MCI from the normal cognitive trajectory.

The production and comprehension of speech are correlated with the activation and coordination of diverse sensory and cognitive processes in regions of the cerebral cortex, including semantic storage and retrieval, executive functions and working memory (9). Therefore, multiple aspects of the language content, correlated among others with lexical processing, grammatical, and syntactic complexity and word finding, are degrading sharply and rapidly with the progression of AD or the transition from an asymptomatic phase to MCI, compared to healthy ageing (10). Further alarming signs include word class deficits, with noun rates usually decreasing and verb and pronoun rates increasing, as well as limited lexical richness and vocabulary size (11). Efforts to identify and assess linguistic deficits in AD and MCI have mainly been focused on oral speech and its transcripts, either in the form of conversational speech or in narrations, naming tests and picture description tasks. Bucks et al. (12) using measures dependant on word frequencies of lexical items during spontaneous oral speech, managed to discriminate between healthy controls (HC) and people diagnosed with probable dementia, validating the decline in lexical richness and a lower noun rate in the latter group. In parallel, in the various studies analysing connected speech relevant to picture description tasks, reviewed by (13), verbal fluency, semantic processing, and pragmatic language use can be assessed even in prodromal stages, while automatic speech analysis emphasizing on vocal features in the work of (14) reaches an accuracy over 80%. Recent work of (15), using multimodal language data and cascaded classifiers, reveals that the language features alone reach an AUC of 0.70 and their combination with other feature types greatly enhances the system's performance. Tools, like the computer-based software of (16), collecting both textual and acoustic linguistic features from different tasks, facilitate the research toward such direction. As far as written speech is concerned, a longitudinal study of (17) on the texts of three novelists, who developed dementia, indicates the progressive lexical and syntactic changes associated with AD. However, despite the rich research in the area of oral speech, a lack of research in the field of Spontaneous Written Speech (SWS) interactions is evident.

Apart from the speech-related area, there are several indications that cognitive decline in MCI patients is associated, to a certain degree, with motor dysfunction in both lower (18) and upper (19) extremity level functions. Dual-task gait tests, that involve walking while doing a cognitively demanding task, have revealed poor gait performance for amnesic MCI patients (20), while (21) managed to discriminate MCI and HC, with AUC 0.83 and sensitivity/specificity 0.82/0.72, using dual-tasks in a clinical setting and involving a sensor-based upper extremity function motor task instead of walking. The search for useful MCI markers and especially digital biomarkers (22), taking advantage of the mobile and wearable consumer device-derived data and their passive collection (23), is a promising new research field, along the increasing plurality and sensitivity of smart sensors for unobtrusive data acquisition. In this vein, focusing on Fine Motor Control (FMC), variability in typing and finger tapping speed in smartphone screens and computer mouses have been used for early dementia detection, as they present a sharper decline compared to their trajectory in healthy ageing (24, 25). Computer-use behaviors are significantly associated with performance on cognitive and functional assessments, with the temporal characteristics of typing, number of pauses and inter-keystroke intervals having been tested as potential markers for cognitive decline (26). Keystroke dynamics, while typing in computer keyboards and smartphone screens, can be captured in a non-clinical setting and have thus been used for the early detection of multiple conditions, such as Parkinson's disease (27), depression (28), and AD with guided copy tasks (29), noticing again the absence of results in spontaneous unprompted keyboard interactions, that better reflect the natural state of the patient.

Having reviewed separately studies regarding linguistic characteristics and keystroke dynamics, the potential diagnostic properties of the cognitive load, associated and partly overlapping with both written speech production and motor dysfunction remain to be examined. Vizer and Sears (30) use a statistical model of keystroke and linguistic features, extracted from ordinary text-typing activities on a computer keyboard, to monitor signs of early cognitive decline in a PreMCI stage, showcasing the efficacy of a combined feature set, reaching an AUC of 0.80, but using as keystroke features time per key, pause rates and duration and, therefore, not correlating with motor dysfunction symptoms. These findings show the potential of multimodal features and equivalent approaches are needed in an MCI stage, reflecting the dysfunctions in multiple brain regions of the pre-frontal cortex, existent in the early stages of dementia. An overview of selected studies that attempted to detect MCI or AD based on speech and motor deficiency-related data or combinations of the above and their results can be found in **Table 1**.

TABLE 1 | References, feature sets, and results of studies regarding MCI and AD detection based on speech or motor deficiency-related features and multimodal data.

Feature Set	*Results*
(12)	
Noun- (N), pronoun- (P), adjective- (A), verb- (V) rate, type token ratio (TTR), Brunet's index (W), Honore's statistic (R), clause-like semantic unit rate (CSU) in transcribed conversational speech	87.5% correct classification, between individuals with probable dementia of Alzheimer type (DAT) and healthy controls, DAT participants had higher mean P, A, V- rate, lower N-rate, higher mean W, lower mean R and TTR, mean CSU did not differ
(14)	
Mean, median, ratio mean, standard deviation of voice and silence duration, periodic vs. aperiodic speech, vocal reaction time, amount of insertions-deletions, irregularity, semantic verbal fluency during short recordings of vocal tasks in-the-clinic	79% accuracy with 20% equal error rate (EER) in classifying MCI vs. HC, 87% accuracy with 13% EER in classifying AD vs. HC with equal SP-SE
(17)	
TTR, lexical repetition, N and V specificity, word class deficit, fillers, syntactic complexity with mean length of utterance, mean number of clauses per utterance, parse tree depth, Yngve depths, D-level scale, use of passive voice in fully parsed texts of three British novelists healthy or with AD	Vocabulary size, repetition and specificity measures validated pronounced decline for authors with AD, syntactic and passive voice analysis did not yield linear results
(13)	
Semantic content, information conciseness efficiency, lexical diversity, total number of words, syntax, V over N rates, coherence prosody, fluency, speech rate in connected speech studies with picture description tasks	Semantic content and conciseness of information yield the best results in detecting MCI and mild AD with picture description tasks
(24)	
Touch and off phase during a finger tapping task of 15 s	SP-SE 0.91–0.52 for ruling out cognitive impairment
(26)	
Mouse operations (amount and time of clicks), keystrokes (amount and timing of text and operational keystrokes), total duration of computer use and pauses during semi-directed computer tasks within a 2-h single testing session	AUC 0.8–0.92 for different features and task combinations with SP 0.62–0.91 and SE 0.8–0.95
(21)	
Motor function speed and variability as measured with two gyroscopes attached to the wrist and upper-arm of the dominant hand during dual tasks (move hand-count numbers) in-the-clinic	AUC 0.83 with SP-SE 0.72-0.82 in predicting MCI and AD
(15)	
Language (26), speech (12), eye movement (22), comprehension (11) features from audio recordings, text transcripts, comprehension questions, and eye tracking during reading silently, aloud and picture description tasks in-the clinic	Combined features: AUC 0.71 and SP-SE 0.79–0.55, Picture description task: AUC 0.72 and SP-SE 0.67–0.63, Verbal reading task: AUC 0.79–0.82 and SP-SE 0.72–0.65, Silent reading task: AUC 0.88 with SP-SE 0.85–0.78 in task fusion
(31)	
370 linguistic (syntactic complexity, grammatical constituents, vocabulary richness, repetitions, information content) and acoustic features (MFCCs) from short narrative samples of the DementiaBank	Top 35 features: 81% accuracy in distinguishing people with AD from HC All the features: 58% accuracy
(30)	
Keystroke (timing, pauses, rates of words, sentence lengths) and linguistic (unique words rate, word class rates, words indicating emotions and cognitive complexity) features from computer-typed texts from older adults with and without PreMCI	Linguistic features: 60% accuracy Keystroke timing features: 68.6% accuracy Combined features: 77.1% accuracy with SP-SE 0.83–0.7 and AUC 0.8

MCI, Mild Cognitive Impairment; AD, Alzheimer's Disease; HC, Healthy Controls; AUC, Area Under the Curve; SP-SE, Specificity–Sensitivity.

Motivated by the aforementioned, the aim of this study is to provide an automated method that can identify MCI individuals and distinguish them from HC, based on digital biomarkers, through their routine interaction with a smartphone keyboard in a non-clinical setting. Specifically, it is firstly assessed how cognitive decline is linked with the natural language processing (NLP) of linguistic features of spontaneous written speech (SWS) production and sequentially with the keystroke information regarding specific motor impairment symptoms. Then, their combination is examined with ensemble models, aiming to capture the interlinked dysfunctions in the respective brain regions associated with language and motor skills; thus, improving the overall classification performance. The promising results, when the proposed approach was tested on data from HC and MCI patients, show that such an analysis could provide with predictive analytics of early stages of cognitive impairment, taking into consideration the pragmatic conditions of everyday living. This will contribute in automatic, unobtrusive,

TABLE 2 | Demographics of participants (MCI, Mild Cognitive Impairment; HC, Healthy Controls) including number of participants, gender (F, female; M, male), age and education level (2: secondary education, 3: higher education, 4: masters-PhD).

Demographics	MCI	HC	Statistical significance (*p*-value)
n	11	12	N.A
Gender F:M (%)	9:2 (81.8:18.2%)	7:5 (58.3:41.6%)	n.s (*p*)
Avg. Age (std)	67.2 (5.96)	66.2 (4.72)	n.s.($p = 0.41$)
Avg. Education level (std)	2.63 (0.67)	2.75 (0.62)	n.s.(*p*)

N.A., not applicable; n.s., not significant with $p > 0.05$.

remote monitoring, and recommendation of MCI and AD diagnosis.

2. MATERIALS AND METHODS

2.1. Data Collection

2.1.1. Participants

The participants of this 6-month long study were recruited in the Day Center "Saint John" of the Greek Alzheimer Association and comprised two groups consisting of in total 12 HC, with the official diagnosis of Subjective Cognitive Impairment (SCI) related to the effects of normal aging and a group of 11 patients diagnosed with MCI. During the duration of the study, there was no clinical progression of the participants' condition. All participants had undergone the same clinical assessment for the official diagnosis within the last 3 years, that included as measures the Mini Mental State Examination-MMSE (32), the Functional Rating Scale for Symptoms of Dementia-FRSSD (33), and the Functional Cognitive Assessment Scale-FUCAS (34), all translated in the participants' native language and scored by medical professionals. The participants were also evaluated in the two main scales related to anxiety and depression, the Beck Depression Inventory-BDI (35) and the Geriatric Depression Scale-GDS (36), with 50% of MCI patients having being identified with minimal to mild depression, while the HC scored negatively. Lastly, blood and neuroimaging tests were conducted and all the study subjects were found negative to Parkinson's Disease. All the participants were 60–75 years old, had finished secondary or higher education and had all been using a smartphone for more than 12 months prior to the study. The groups did not differ in age, gender, and education levels, based on the two-sided Mann–Whitney *U*-test for the age and Chi-squared test for the gender and the education levels ($p > 0.05$). The demographic information of all participants is also presented in **Table 2**.

2.1.2. Study Protocol

Participants downloaded in their own smartphones the "Type of Mood" mobile application from Google Play store (37) and created an account, providing information about their gender, date of birth, education, and income levels, smartphone usage, as well as filling in the digitized version of the PHQ-9 questionnaire on depression (38), while giving their consent in their data processing within the app. The study was reviewed and approved by Greece, Bioethics Committee of the Aristotle University of Thessaloniki, Medical School, Thessaloniki, Greece (359/3.4.17) and all participants provided their written informed consent to participate in this study. Sequentially, they activated the app's custom keyboard, which replaced the default typing input method across all aspects of the Android Operating System. The keyboard recorded keystroke timing information, i.e., sequences of timestamps of key presses and releases, as well as typing metadata (delete rate, pauses, number of characters typed, and typing sessions' duration), in the background, without interfering with participants' routine typing and without capturing the characters typed, rendering the process privacy-aware. For each typing session (keyboard shown and afterwards hidden, with at least one key tap in the meantime), the above-mentioned keystroke timing information was stored in a JSON format and indexed as database entries in a local SQLite database, available only to the application. The application would periodically transmit database entries to a remote cloud server (Microsoft Azure), when the user's device was connected to Wi-Fi and charging, accompanied by the uniquely coded ID of the user. Within 6 months, 4,930 typing sessions were collected, 3,000 from 11 MCI patients and 1,930 from 12 HC (more than 100 sessions per individual smartphone), while 3,139 of them, with more than 40 key presses per session, were eventually used for the analysis.

Simultaneously, participants were asked to type down on their phones up to 4 short texts, around a paragraph in length, to be used for the NLP analysis. To simulate SWS production the suggested topics included: (1) a message to a loved one or a good friend, (2) a description of one's day, (3) giving advice on someone, and (4) narrating a short story from a happy memory, as those were the things they usually text their social circle about. The participants typed the texts on their phones at home, based on their own availability and without a time limit, to simulate a non-clinical setting. They could not use the auto-correction feature of their phones and they sent afterwards the texts via email. In total, 10 MCI patients contributed 40 texts and 7 HC contributed 28 texts. From all the participants, 10 MCI patients and 5 HC, respectively contributed for the fused features analysis both texts and keystroke information through their typing activities with the custom keyboard.

2.2. Models and Experiments

2.2.1. Pre-processing

The texts were delivered by the participants as emails and saved as txt files with the equivalent user ID. The spelling mistakes and accidental punctuation marks were corrected, to avoid interference with the Part-Of-Speech (POS) Tagging process (words receive tags based on their word class type i.e., noun, verb etc.), given they do not represent studied measures in the following experiments. Using a POS-Tagger trained for the Greek language (39), each text was also stored with a CONNL-U dependency parse tree format (40). Dependency parsing is based on the notion that linguistic units, e.g. words, are interconnected with directional links within a sentence. Therefore, the parsing

tree reflects the syntactical relation of the words within the sentence, with the "children" words being dependent from the "root" word, while its depth corresponds to the overall syntactical complexity of the sentence.

The timestamped sequences of the typing data, contained in the stored JSON files, were used to extract keystroke dynamics variables, namely the "hold time" (HT-time interval between pressing and releasing a key) and "flight time" (FT-time interval between releasing a key and pressing the next one), as shown:

$$HT_n = t_n^r - t_n^p \quad \text{with} \quad n = 1, 2, \ldots, N, \tag{1}$$

$$FT_n = t_{n+1}^p - t_n^r \quad \text{with} \quad n = 1, 2, \ldots, N\text{-}1, \tag{2}$$

where t_n^p and t_n^r refer to the pressing and releasing times, respectively. With appropriate filtering, by keeping the HT values smaller than 700 μs and the FT values smaller than 3 s, the long pauses were excluded and the keystroke timing information was less susceptible to noise. Zero-padding was used to achieve common dimensionality of 100 values in HT and FT arrays for each typing session.

2.2.2. Feature Extraction

2.2.2.1. NLP features

For each text, a set of nine features was extracted, using the Natural Language Toolkit suite of tools and libraries (41), along custom made functions and the mean value of each feature across the different texts of each user was used in the final features array. Specifically, each text was tokenized by breaking it up into words and punctuation marks and the set of unique words, consisting its vocabulary, was further extracted. The lexical diversity and richness of the texts are expressed with three features. The ratio of unique words over the total number of words is calculated as shown below:

$$dvrst = V/W, \tag{3}$$

where W denotes the total number of words and V the total number of unique words, namely the vocabulary. The Brunet Index—"BI" indicates richer language with lower values, independently from the text's length (42) and is calculated as shown below:

$$BI = W^{V^{-0.165}}. \tag{4}$$

Honore's Statistic—"HS," indicating richer language with higher values (43) is calculated as seen below:

$$HS = \frac{100 * logW}{1 - \frac{hap}{V}}, \tag{5}$$

where *hap* denotes the number of hapaxes legomena (number of words appearing only once within a text). The average number of words per sentence is calculated as seen below:

$$wrd\_sent = \text{sum(w\_sents)/len(w\_sents)}, \tag{6}$$

where *w_sents* denotes a list with the number of words per sentence as elements. The ratio of the meaningful vocabulary (words that are not part of the Greek stopwords from the NTLK corpus) is expressed as shown below:

$$nonstop = \text{len(content)}/V, \tag{7}$$

where *content* refers to the vocabulary words that are not considered stopwords based on the NLTK corpus. Features related to word class deficits were extracted from the POS-Tagged texts. The nouns over verbs ratio is calculated as shown below:

$$NnVrb = nn/vrb, \tag{8}$$

where *nn* denotes the number of nouns and *vrb* the number of verbs within the text. The noun ratio and word finding difficulties are expressed as seen below:

$$Nn = nn/(vrb + nn). \tag{9}$$

The pronoun ratio, which quantifies indirect referencing, is presented below:

$$Prn = prn/(nn + prn), \tag{10}$$

where *prn* denotes the number of pronouns within the text. Lastly, a form of the Mean Dependency Distance—"MDD" was used (44) to express the syntactical complexity. Specifically, for each text the MDD of each sentence was calculated separately and then the sum of those MDDs was divided by the number of sentences, getting a mean estimation of this feature for each text. The MDD of each sentence is defined as the sum of the distances of each "child"/dependent word from its "root"/head word over the number of tokens of the sentence as shown below:

$$MDD_{(\text{sentence})} = \frac{1}{N-1} \sum_{i=1}^{N} |DD_i|, \tag{11}$$

where N denotes the number of tokens and DD_i the dependency distance from the $i - th$ link. The latter was calculated with Pre-order Tree Traversal (traverse the root, the left sub-tree, and then the right sub-tree) of the dependency parsed trees format of the sentences in the CONLL-U files. In **Figure 1**, there is an example of a sentence with dependency parsing.

2.2.2.2. Keystroke features

As far as the keystroke dynamics are concerned, the extracted features are associated with three indices Bradykinesia (B), Rigidity (R), and Alternate Finger Tapping (AFT), the symptoms describing Fine Motor Impairment (FMI). Specifically, one dimensional CNN-based autoencoders, consisting of two sequentially connected convolutional layers (kernel size of 5, 16 filters) without max-pooling layers, took as inputs the HT and FT sequences of the typing data and were used to learn a neural

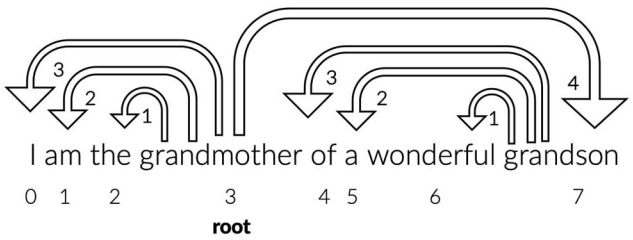

FIGURE 1 | The dependency parsing of the sentence "I am the grandmother of a wonderful grandson" is depicted. The ids of the words, representing the position of the words in the sentence, are indicated with numbers 0–7 and the word "grandmother" is identified by the parser as the root word. The arrows have the direction from the head/root word to the dependent/children, the dependency distance between the pair is indicated on the arrow head and equals the difference of their id values. The words "the," "am," "I," and "grandson" are dependent from the word "grandmother" and the words "wonderful," "a," and "of" are dependent from the word "grandson." The MDD of this sentence based on the Equation (11) equals (3+2+1+4+3+2+1)/7= 2.28.

network's encoding mechanism for representing efficiently the keystroke dynamics and reject the noise. The CNN was trained in an unsupervised manner (80–20% train-test split on 34,000 typing sessions, back-propagation for 50 epochs with mini-batches of size 64, RMSprop optimizer with learning rate of 10^{-3} for the mean squared error loss function) on a dataset of a relevant study of Parkinsonian screening (45), to learn the inherent structure of keystroke dynamics and represent them with a limited number of features. Indicatively, the in-the-wild development dataset, used for unsupervised pre-training of the neural network parameters consisted of 34,000 typing sessions with keystroke dynamics drawn from subjects with age $\leq$ 40 years (self-reported). A two-layer fully-connected network with 50 hidden nodes was added to the already trained networks and the final network was fine-tuned on an in-the-clinic dataset of 33 subjects, by optimizing regression models with leave-one-subject-out (LOSO) cross validation (50 epochs with mini-batches of size 32) to each estimate the severity of the symptoms Rigidity/Alternate Finger/Bradykinesia in relation with the ground-truth Unified Parkinson's Disease Rating Scale (UPDRS) Part III single-item scores 22/23/31 (46). The number of parameters reaches 10,500 altogether and the overall model layout can be found in **Figure 2**.

The in-the-clinic development dataset consists of 274 typing sessions in total (up to 10 text-excerpts for each user) with keystroke dynamics from 33 demographically matched subjects [18 early Parkinson Diseases (PD) patients and 15 HC], whom underwent clinical examination by neurologists and their FMI was evaluated by their UPDRS Part III single-item scores 22/23/31, expressing rigidity of upper extremity/alternate finger tapping/general body bradykinesia-hypokinesia, respectively. Before being tested to our dataset, consisting of more than 3,000 typing sessions of MCI patients and HC, the models have also been tested in 4 in-the-wild datasets: (1) a dataset with 216,000 typing sessions with keystroke dynamics from clinically examined 214 subjects (PD vs. HC), with self-reported

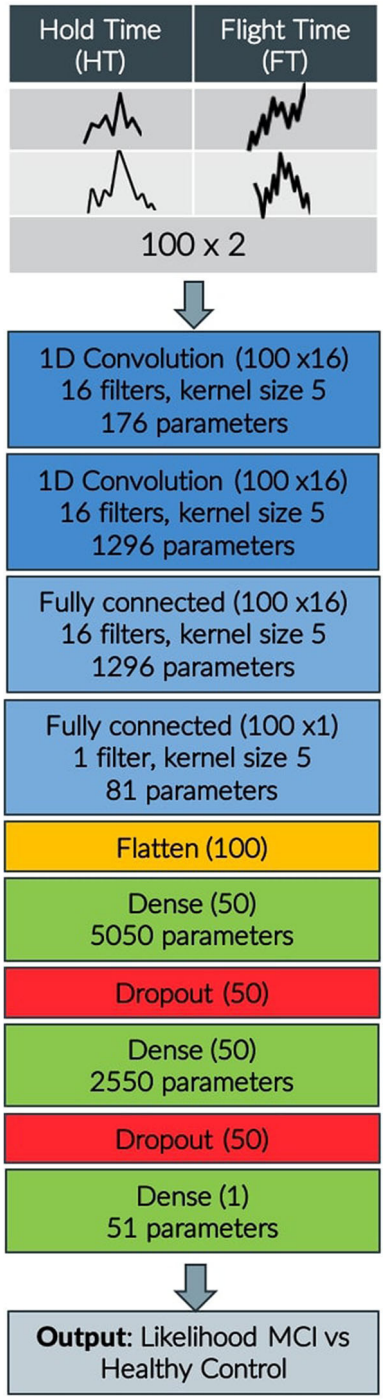

FIGURE 2 | CNN model layout indicating input, convolution, flattening, dense, and dropout layers, output shape of each layer and number of parameters.

demographics, through the iPROGNOSIS app, (2) a dataset with 36,000 typing sessions with keystroke dynamics from 39 subjects (PD/HC:22/17), (3) a subset of the previous dataset with 7,600 typing sessions, drawn from *de novo* PD patients and the same HC (*de novo* PD/HC: 9/17), (4) the union of the first two datasets with 252,000 typing sessions with keystroke dynamics from

TABLE 3 | Features from the natural language and typing processing.

Features	Descriptions
NLP features	
dvrst	Lexical diversity
nonstop	Meaningful content
wrd_sent	Avg. No. words per sentence
BI	Brunet index
HS	Honore's Statistic
NnVrb	Noun per Verbs ratio
Nn	Noun ratio
Prn	Pronouns ratio
MDD	Syntactical complexity
Keystroke features	
R indices	Rigidity
B indices	Bradykinesia
AFT indices	Alternate finger tapping

253 subjects (PD/HC: 67/186) The optimized scores produced indicators that can be used in-the-wild prediction of UPDRS scores 22/23/31, yielding correlation 0.66/0.73/0.58, respectively, in the validation set of 36,000 typing sessions. The trained models were used to infer based on the typing data from 11 MCI patients and 12 HC, and estimations for each symptom (R/B/AFT 0-4) were extracted for each typing session of each user (47). All the features from the natural language and typing processing can be found in **Table 3**.

2.2.3. Experiments and Classification Models

In order to address the study research questions, that is, the exploration of the classification performance between MCI and HC, using features drawn from the NLP, keystroke dynamics, and combined feature spaces, three experiments, i.e., EXP1, EXP2, and EXP3, with different feature sets and demographically matched sub-cohorts were conducted, accordingly. In each experiment, three different classifiers, Logistic Regression, Random Forest, and k-Nearest Neighbors are evaluated based on their accuracy and the receiver operating characteristic (ROC) analysis, after multiple rounds of LOSO cross-validation. ROC analysis is an iterative process of varying the discrimination threshold of a binary classifier and outputting the (Sensitivity, Specificity) pair for each threshold. The ROC curve is then formed by plotting the output pairs of (1—Specificity, Sensitivity). This analysis provides reliable insights into the performance of a classification model even when datasets are not completely balanced. To assess the statistical significance of classification results, sampling with replacement (1,000 bootstraps) is further used here to define a ROC curve distribution, by obtaining the average value (solid line in figures) and the confidence intervals (shadowed areas in the figures) of the area under the ROC curve (AUC). Where reported, specificity/sensitivity values correspond to the optimal ROC-based cut-off point (decision threshold), estimated by maximizing the Youden Index (48), for equal cost of misclassifying MCI patients and HC. A Univariate Feature Selection process was used, where each feature was individually assessed with regards to its statistical significance with the label and combinations of the features with the highest correlation were chosen for the optimal feature set in each experiment. Standardization and scaling was performed in the feature sets when needed, by removing the mean and scaling to unit variance. An overview of the pipeline and the different experiments and cohorts can be seen in **Figure 3**.

In particular, at EXP1, 78 texts from 10 MCI patients and 7 HC were used to extract NLP features and test the predictive potential of SWS production. In EXP2, 3,139 typing sessions from 11 MCI patients and 12 HC were used to extract the keystroke related features that quantify FMI symptoms and test their predictive potential linked to cognitive decline. In EXP3, NLP and keystroke feature sets, combined differently in 3 models (A, B, C), were used to assess the performance of fused feature sets against models with separate NLP and keystroke features ("Just NLP," "Just Keys"), from 10 MCI patients and 5 HC contributing both texts and typing sessions. Model "A" concatenates the probabilities/predictions of the "Just NLP" model with keystroke features and feeds them into another Random Forest classifier, while model "B" concatenates the probabilities/predictions of the "Just Keys" model with NLP features and feeds them into another k-Nearest Neighbors classifier. Lastly, model "C" concatenates the probabilities/predictions of both the "Just Keys" and "Just NLP" models and feeds them into another Random Forest classifier. The cohorts and the chosen optimal feature sets of each experiment can be found in **Table 4**.

3. RESULTS

Table 5 along with **Figure 4** summarize the acquired results from all three experiments. In the following subsections, the specific results per experiment are presented.

3.1. First Experiment

Combinations of the features "nonstop," "dvrst," "HS," and "MDD," comprising the optimal feature set, were used with the different classifiers. The Logistic Regression classifier has an AUC of 0.76, accuracy 76% and specificity/sensitivity of 0.80/0.71, respectively. The Random Forest Classifier has an AUC of 0.73, accuracy 71% and specificity/sensitivity of 0.83/0.43 respectively. The k-nn classifier with seven nearest neighbors has an AUC of 0.69, accuracy 71% and specificity/sensitivity of 0.80/0.57, respectively. The Logistic Regression classifier appears to have the best performance, whereas the Random Forest one, due to the small number of data wasn't able to perform that well. We observe higher specificity than sensitivity levels. The three ROC curve distributions of the first experiment can be found in **Figure 4A**.

3.2. Second Experiment

Combinations of the "B" and "R" indices, comprising the optimal feature set, were used with the different classifiers. The Logistic Regression classifier has an AUC of 0.69, accuracy

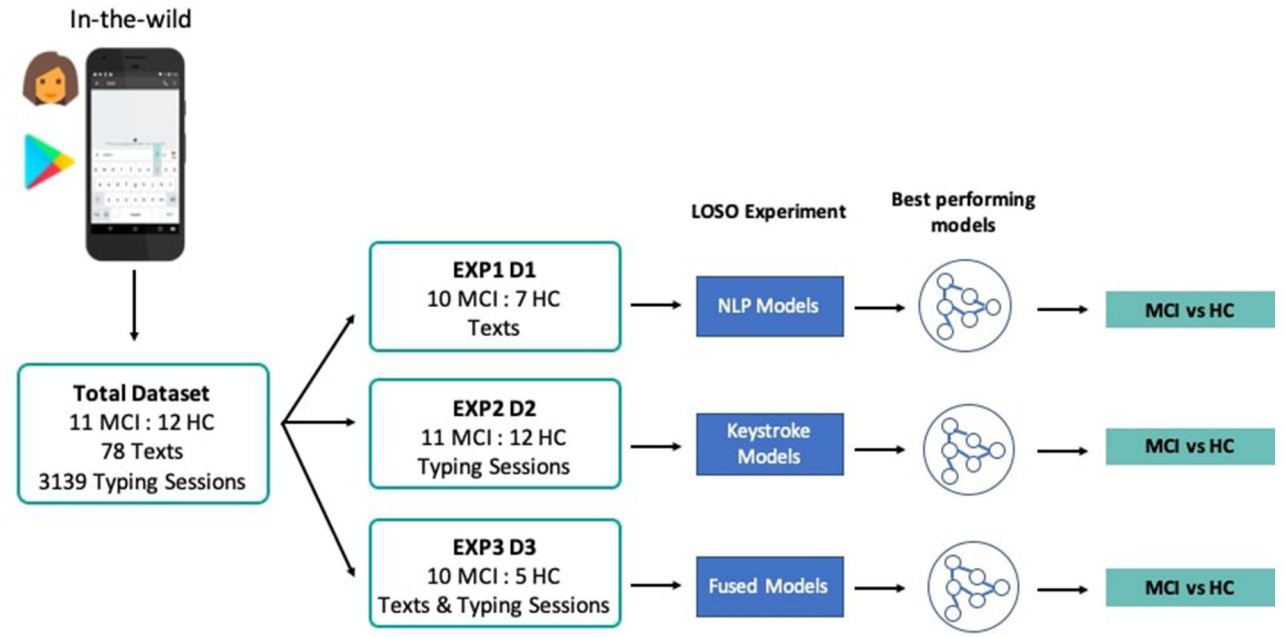

FIGURE 3 | System pipeline and experiments' overview. Texts and typing sessions were collected from MCI patients and HC through the custom keyboard of the TypeOfMood application in a non-clinical setting. The total dataset was split accordingly (D1, D2, D3) to the needs of the three experiments that tested NLP models (EXP1), keystroke models (EXP2), and models with fused feature sets (EXP3). Multiple rounds of LOSO experiments yielded the best performing models to distinguish among MCI and HC.

TABLE 4 | Cohorts (number of MCI:HC), optimal feature sets and models of the experiments EXP1, EXP2, EXP3.

Cohort (MCI:HC)	Feature set	Model
EXP1: NLP features		
10:7	nonstop, dvrst, HS, MDD	N.A
EXP2: Keystroke features		
11:12	B and R indices	N.A
EXP3: Fused features		
10:5	Nonstop, dvrst, MDD	Just NLP
	B and R indices	Just Keys
	Probabilities of Just NLP & B indices	A
	Probabilities of Just Keys & dvrst, nonstop, MDD	B
	Probabilities of Just NLP & probabilities of Just Keys	C

N.A., not applicable.

TABLE 5 | Results of the three experiments with NLP features, keystroke features and their combination in cascaded classifiers.

Classifier	Accuracy	AUC	Specificity	Sensitivity	Model
EXP1					
LR	**0.76**	**0.76**	**0.80**	**0.71**	**N.A**
RF	0.71	0.73	0.83	0.43	N.A
k-nn	0.71	0.69	0.80	0.57	N.A
EXP2					
LR	0.70	0.69	0.54	0.83	N.A
RF	0.66	0.65	0.55	0.75	N.A
k-nn	**0.77**	**0.78**	**0.64**	**0.92**	**N.A**
EXP3					
LR	0.67	0.60	0.80	0.40	Just NLP
k-nn	0.80	0.75	0.90	0.60	Just Keys
RF	0.75	0.68	0.90	0.40	A
k-nn	**0.80**	**0.75**	**0.90**	**0.60**	**B**
RF	0.73	0.65	0.90	0.40	C

Bold values denote the highest performing models. N.A., not applicable.

70% and specificity/sensitivity of 0.54/0.83, respectively. The Random Forest Classifier has an AUC of 0.65, accuracy 66% and specificity/sensitivity of 0.55/0.75, respectively. The k-Nearest Neighbors classifier with seven nearest neighbors has an AUC of 0.78, accuracy 77% and specificity/sensitivity of 0.64/0.92, respectively. The k-Nearest Neighbors classifier appears to have the best performance and we observe higher sensitivity than specificity levels. The three ROC curve distributions of the second experiment can be found in **Figure 4B**.

3.3. Third Experiment

The "Just NLP" model resulted in an AUC of 0.60, accuracy 67% and specificity/sensitivity of 0.80/0.40, respectively. The "Just Keys" model resulted in an AUC of 0.75, accuracy 80% and specificity/sensitivity of 0.90/0.60, respectively. Model "A," resulted in an AUC of 0.68, accuracy 75% and

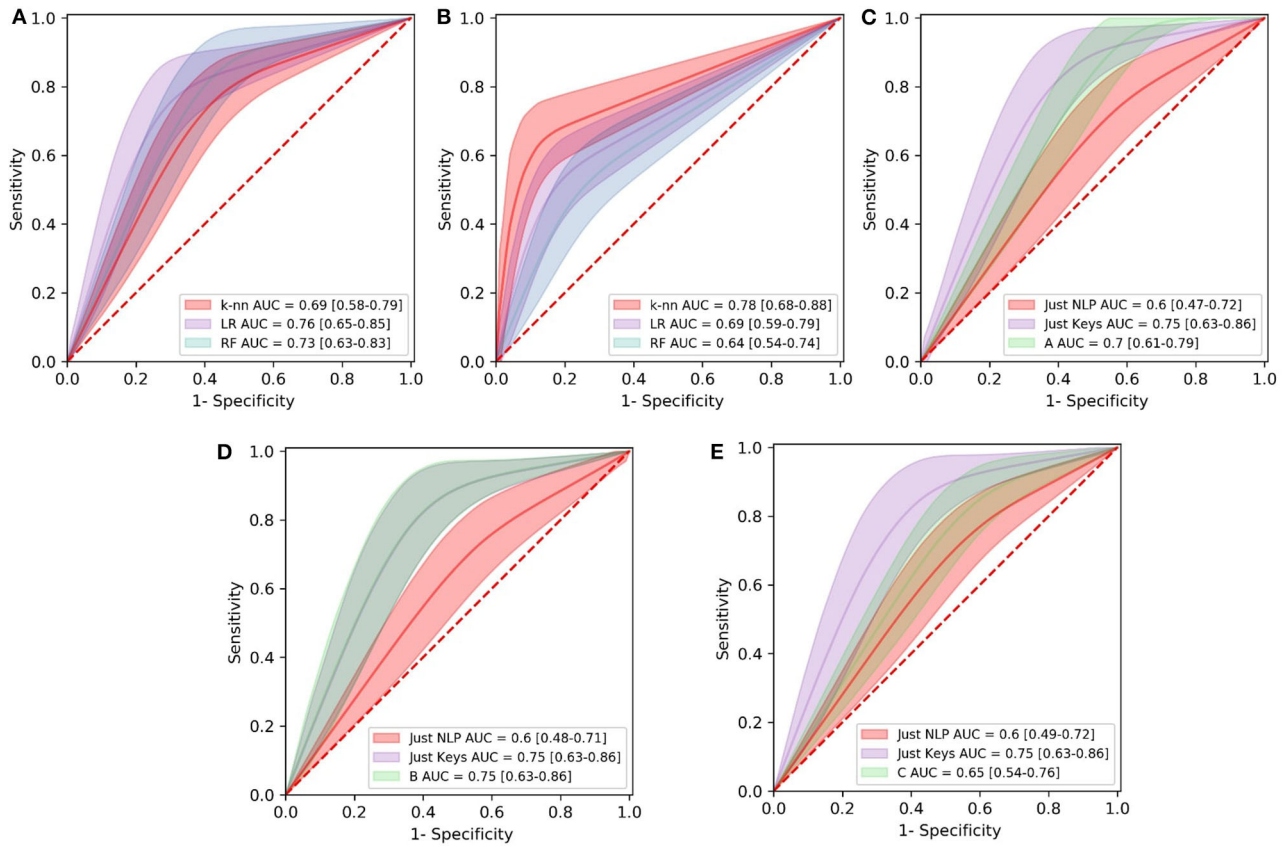

FIGURE 4 | Comparison of Receiver Operating Characteristics (ROC) curves of all the models in the first **(A)**, second **(B)**, and third **(C–E)** experiment. Area Under the Curve (AUC) values for models are shown with 95% Confidence Intervals. LR, Logistic Regression; RF, Random Forest; knn, k-Nearest Neighbors.

specificity/sensitivity of 0.90/0.40, respectively (**Figure 4C**). Model "B" resulted in an AUC of 0.75, accuracy 80% and specificity/sensitivity of 0.90/0.60, respectively (**Figure 4D**). Model "C" resulted in an AUC of 0.65, accuracy 73% and specificity/sensitivity of 0.90/0.40, respectively (**Figure 4E**). It is validated that the feature sets that have combined the potential of both NLP and keystroke dynamics have better performance, even though the dataset tested and the number of data subjects was quite limited and we observe, once more, higher specificity than sensitivity levels. The highest performance in the current dataset is accredited to model "B," with similar results with the "Just Keys" model, while model "A" also has very high specificity levels.

4. DISCUSSION

Digital and Connected Health is an emerging field, encompassing novel, efficient, effective, and accessible technological tools, that could contribute greatly to early disease diagnosis and management. The aim of this study has been the development of an objective tool for the detection of MCI patients against HC, by exploiting linguistic characteristics and keystroke dynamics, during routine typing on mobile touchscreens. The adopted design reflects the natural spontaneous state of the users and its unobtrusive data collection fashion meets the need of long-adherence. Simultaneously, such a tool assists the longitudinal passive monitoring and diagnosis of the condition, in an interpretable way, facilitating physicians and medical interpretation, through high-frequency sampled data streams. Overall, the study further verifies the relationship between cognitive functions and motor dexterity, being consistent with the overlapping cognitive and motor neural circuitry in the brain, and paving the way toward new approaches for dementia risk screening.

A number of studies have shown that speech and language, being ubiquitous in everyday communication, can provide early signs of MCI and other prodromal stages of AD (49), while being correlated with a lapse in episodic and semantic memory. Episodic memory refers to the multifaceted process that enables the retrieval of detailed evocative memories from the past, while semantic memory is linked with the retrieval of general conceptual knowledge divested of specific spatiotemporal contexts (9). Word finding difficulties and pronounced word class deficits are considered as some of the earliest manifestations of language breakdown in MCI and AD and implicate loss of semantic knowledge and difficulties in encoding new information (12). Our study further validates these impairment patterns per MCI subject, with lower mean number of nouns [both overall (*Nn*) and when compared to verbs (*NnVrb*)], and higher mean

number of pronouns (*Prn*) in MCI patients compared to HC, despite the fluctuation noticed in the number of words per sentence within subjects over successive text sessions (*wrd_sent*). The measures of the vocabulary size and lexical diversity (*dvrst*, *BI*, *HS*) were decreased within the group of MCI patients as expected (11), with slightly higher mean values of the *BI* and greatly lower values of the *HS*, indicating pronounced lexical repetitions with less hapaxes and lexical richness deficits for the MCI patients. Regarding the *nonstop* metric, that was part of all optimal feature sets in all experiments, the MCI patients appeared to use less connecting words/stopwords, reflecting potentially a decreased syntactical complexity, as an expected remedy of their memory loss capacity (50), that was further validated with the lower mean values of the *MDD* metric. Although the *MDD* metric has been primarily used to assess the general complexity of different languages (44), we are using it for the first time on a micro-level and it managed, in our context, to efficiently reflect the syntactical complexity of written speech per subject.

Regarding motor dysfunction in the upper extremities in early stages of dementia, several studies suggest that mild Parkinsonian signs are associated with MCI patients (51) and the degree of motor impairment may help identify those at risk for AD (18). Our machine learning-based estimation of dominant hand bradykinesia (slowness of movement) and rigidity (muscle stiffness) indices managed to efficiently distinguish between MCI patients and HC, validating the connection between FMI and cognitive decline, given that none of our participants was diagnosed with PD. The diagnostic potential of keystroke dynamics has been proven in previous studies (52) and our results of longer, more variant pressing of the keyboard keys and slower finger coordination across the screen for the MCI patients, further strengthens this potential. These findings are also aligned with studies experimenting with finger tapping speed (24), associated with short-memory lapses and other fine motor dexterity (53) and upper extremity function (21) tests conducted in a clinical setting (data in-the-clinic). Nevertheless, our approach goes several steps further by encapsulating the natural state of the users, capturing non-invasively and longitudinally their typing sessions, without the need for special technical equipment and during their pragmatic real-life activities (data in-the-wild), thus enabling a continuous monitoring of the very early stages of the condition.

Taking it a step further, the study evaluated the combined diagnostic potential of the NLP and keystroke related features, in various models, to validate the clinical suggestions that hand dexterity and FMI co-exist with MCI, possibly sharing similar pathogeneses (54). The combined feature sets in EXP3 had indeed an increased performance in detecting MCI patients from HC, especially model *B* that combined cascadedly the predictions of the *JustKeys* model of solely keystroke features with NLP features, being aligned with other studies evaluating multimodal data combinations (15, 30). The rationale of this association of features resides in the fact that hand dexterity requires complex cognitive processes, beyond sensorimotor coordination of the limbs with the eyes, linked with executive functions, such as attention, judgement, planning, and memory (20). Our results further reinforce the relationship between cognitive and motor function in other functional activities, such as written speech production in our context, beyond the typical mobility tasks and self-reporting as suggested by (19), with a novel FMI detection tool. Moreover, SWS production activates primarily the prefrontal cortex of the brain, associated with both executive cognitive functions and hand dexterity, while being consistent with the overlapping and reciprocal neural circuitry of motion and cognition in the cerebellum and the subcortical structures (55). Therefore, our fused models reflect the overall intricacies related to SWS production, as the thought and cognition process, linked with motor deficiencies, materialize to the actual speech production, providing a highly valuable patient phenotype. The high specificity levels of the NLP models are further enhanced with the highly sensitive and granular information coming from the greater amount of typing sessions against the fewer text samples, resulting in a robust system. In a correlation analysis (using the Pearson correlation coefficient) between the clinical scores of MMSE, FUCAS, and FRSSD scales and the predictions of our models and individual features, the bradykinesia indices significantly correlated ($r = -0.56$, $p < 0.01$) with the MMSE clinical scores and the predictions of the NLP models significantly correlated ($r = -0.55$, $p < 0.01$) with the FUCAS clinical scores. These correlations tangibly showcase the connection of the FMI symptoms, like bradykinesia, with the clinically verified cognitive decline, aligned with the relevant literature (19, 51, 53), and showcase how linguistic deficiencies, captured with the NLP analysis, reflect clinically measured functional deficiencies in everyday life tasks (9), as those used for the FUCAS measure. The estimated correlation levels (around 0.5) between the overall predictive models and the chosen measures were anticipated, since these scales have low sensitivity and variation within the MCI spectrum and the diagnosis at such an early stage is greatly dependant from additional neuroimaging and physiological tests beyond these individual scores.

The novelty of the current study is that it sets up an interpretable framework of unobtrusive assessment of individual symptoms of the early stages of dementia and MCI, harvesting data in-the-wild and thus reflecting the natural state of the user, while still reaching for high correlation with the equivalent clinical scores. The latter is of high importance toward personalized monitoring of different risk factors preceding clinical diagnosis, while the detection of the severity level of FMI and cognition-related symptoms could also facilitate personalized interventions for better management of the patient's condition and increased quality of life. In parallel, such analyses are linked with the use of smartphones and virtual keyboards, assisting further the mobile health booming. The high specificity levels of the proposed models are aligned with the requirements of a remote monitoring system that needs to limit "false alarms" for the HC and the CNN architecture paves the way for new methodologies in digital diagnostic systems. Future work could examine whether interventions targeting neuromuscular traits, such as hand and motor dexterity, may also benefit higher cognitive and functional outcomes. Furthermore, the combination of other data sources, e.g., acoustic features of oral speech (31), gait performance (56), behavioral and social metrics (57), along

the further sophistication of the linguistic features extraction, can yield greater performance. Moreover, scaling the study to a larger pool of clinically validated subjects, along a continuous data stream acquisition, will lead to even more robust values in diagnostic performance and thus assist physicians in their clinical estimations and treatment responses of the condition. Therefore, beyond the micro-level approach of this tool, targeted to the user's needs, the benefits extend to the realm of precision medicine toward more efficient clinical decision making.

4.1. Limitations and Implications

Despite the promising results presented here, there are some limitations to be considered. Firstly, the overall size of the cohort was small as the study demanded specific educational levels, technological familiarization, and an extensive suite of neuropsychological and physiological tests for both MCI patients and HC. Clearly, this resulted in a difficult recruitment process at the specific age group. Nevertheless, the models have been designed with scalability in mind and re-analyzing data collected from a larger cohort will lead to a more accurate and robust performance. The validity of the self-reported demographics of the study's participants could also be considered as another limitation, but these characteristics did not yield any statistical significance and therefore could not greatly affect the end result. As far as the NLP analysis is concerned, large population observational studies with a greater amount of text samples are required to account for inter- and intra- subject variability in written speech production, given the expected heterogeneity in linguistic changes among individuals in both normal ageing and dementia. Although linguistic decline is accelerated in the presence of MCI, these changes are also highly correlated with the educational and literacy level and familiarization with written speech production, demanding a longitudinal monitoring of speech characteristics of the participants for more generalized results. Moreover, syntax and word choice is also dependent on the stylistic choice of each participant's writing style and thus may account for differences in the metrics that are linked with syntactical complexity. Nevertheless, here we focused on metrics that could yield diagnostic potential despite the limited number of texts and reflect the most common linguistic changes linked to cognitive decline and analyzed the data, bearing in mind these constraints. Furthermore, errors of the Greek POS-Tagger in the part of speech tagging process, although rare, may occur and thus can affect the accuracy of the syntactical parsing process and other measures like the noun, pronoun and verb ratios. As far as the keystroke dynamics analysis is concerned, the algorithms identifying early FMI symptoms were trained on an extensive PD and HC cohort of a relevant study (45), as the MCI and HC cohort of this study was relatively small. However, this does not interfere with the validity of the process, as the algorithms detect the motor symptoms' severity and not Parkinson's condition itself, even though clinically these symptoms do have similarities in these two neurodegenerative diseases (51). Moreover, the results have been calculated with a confidence interval to take into account potential deviations and the longitudinal nature of the data, collected in a non-clinical setting during the routine typing of participants, gave us an ecologically valid and more realistic picture of the FMI symptoms. Lastly, although this study can be materialized as a diagnostic tool to be used at a clinical setting, privacy, and security issues related to the written speech content and the medical data of the participants have to be catered properly, while abiding with all the guidelines for digital health tools (58).

5. CONCLUSION

In this work, a new perspective in the detection of MCI, based on natural language and touchscreen typing processing during SWS production, was presented. Linguistic features, keystroke dynamics and their fusion, reflecting both cognitive and motor deficiencies and their reciprocal expression in MCI patients, are assessed as digital biomarkers. Experimental results in demographically matched cohorts and machine learning models have justified an efficient discrimination performance for the fused feature sets ($AUC \geq 0.75$), that provide a complete phenotype of MCI-related symptoms. The promising results presented here pave the way toward a holistic, objective patient-centric AD detection tool to be even successfully deployed in a non-clinical setting.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they cannot be publicly uploaded and accessed without a reasonable request. Requests to access the datasets should be directed to Leontios J. Hadjileontiadis, leontios@auth.gr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Bioethics Committee of the Aristotle University of Thessaloniki, Medical School, Thessaloniki, Greece (359/3.4.17). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AN, DI, and LH conceived the study protocol and contributed to the manuscript. AN, DI, SH, and VC developed the keyboard and the keystroke dynamics processing algorithms. AN developed the NLP feature extraction and processing algorithms and conducted all experiments. MT and Alzheimer Hellas center conducted the clinical evaluations. AN and DI analyzed the data, developed the participants demographic information, participant information and consent process, and handled the data governance procedures as well as the corresponding ethic approvals. All authors discussed the results. All authors contributed to the article and approved the submitted version.

FUNDING

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 690494—i-PROGNOSIS: Intelligent Parkinson early detection guiding novel supportive interventions.

ACKNOWLEDGMENTS

The authors wish to acknowledge the donations of time and data from all of the study participants and the personnel, especially psychologist Moses Gialaouzides, from the day center Saint John of the Alzheimer Hellas Association.

REFERENCES

1. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. (1999) 56:303–8. doi: 10.1001/archneur.56.3.303
2. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. *J Intern Med*. (2014) 275:214–28. doi: 10.1111/joim.12190
3. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. (2004) 256:183–94. doi: 10.1111/j.1365-2796.2004.01388.x
4. Janoutová J, Sery O, Hosák L, Janout V. Is mild cognitive impairment a precursor of Alzheimer's disease? Short review. *Cent Eur J Public Health*. (2015) 23:365–7. doi: 10.21101/cejph.a4414
5. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol*. (2011) 10:819–28. doi: 10.1016/S1474-4422(11)70072-2
6. Chang CY, Silverman DH. Accuracy of early diagnosis and its impact on the management and course of Alzheimer's disease. *Expert Rev Mol Diagnost*. (2004) 4:63–9. doi: 10.1586/14737159.4.1.63
7. Sperling R, Mormino E, Johnson K. The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron*. (2014) 84:608–22. doi: 10.1016/j.neuron.2014.10.038
8. Waldemar G, Dubois B, Emre M, Georges J, McKeith I, Rossor M, et al. Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. *Eur J Neurol*. (2007) 14:e1–26. doi: 10.1111/j.1468-1331.2006.01605.x
9. Szatloczki G, Hoffmann I, Vincze V, Kalman J, Pakaski M. Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Front Aging Neurosci*. (2015) 7:195. doi: 10.3389/fnagi.2015.00195
10. Levy R, et al. Aging-associated cognitive decline. *Int Psychogeriatr*. (1994) 6:63–8. doi: 10.1017/S1041610294001626
11. Burke DM, Shafto MA. Language and aging. In: Craik FIM, Salthouse TA, editors. *The Handbook of Aging and Cognition*. New York, NY: Psychology Press (2011). p. 381–451.
12. Bucks RS, Singh S, Cuerden JM, Wilcock GK. Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology*. (2000) 14:71–91. doi: 10.1080/026870300401603
13. Mueller KD, Hermann B, Mecollari J, Turkstra LS. Connected speech and language in mild cognitive impairment and Alzheimer's disease: a review of picture description tasks. *J Clin Exp Neuropsychol*. (2018) 40:917–39. doi: 10.1080/13803395.2018.1446513
14. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's Dement*. (2015) 1:112–24. doi: 10.1016/j.dadm.2014.11.012
15. Fraser KC, Lundholm Fors K, Eckerström M, Öhman F, Kokkinakis D. Predicting MCI status from multimodal language data using cascaded classifiers. *Front Aging Neurosci*. (2019) 11:205. doi: 10.3389/fnagi.2019.00205
16. Komeili M, Pou-Prom C, Liaqat D, Fraser KC, Yancheva M, Rudzicz F. Talk2Me: Automated linguistic data collection for personal assessment. *PLoS ONE*. (2019) 14:e212342. doi: 10.1371/journal.pone.0212342
17. Le X, Lancashire I, Hirst G, Jokel R. Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Literary Linguist Comput*. (2011) 26:435–61. doi: 10.1093/llc/fqr013
18. Aggarwal NT, Wilson RS, Beck TL, Bienias JL, Bennett DA. Motor dysfunction in mild cognitive impairment and the risk of incident Alzheimer disease. *Arch Neurol*. (2006) 63:1763–9. doi: 10.1001/archneur.63.12.1763
19. Rycroft SS, Quach LT, Ward RE, Pedersen MM, Grande L, Bean JF. The relationship between cognitive impairment and upper extremity function in older primary care patients. *J Gerontol*. (2019) 74:568–74. doi: 10.1093/gerona/gly246
20. Montero-Odasso M, Oteng-Amoako A, Speechley M, Gopaul K, Beauchet O, Annweiler C, et al. The motor signature of mild cognitive impairment: results from the gait and brain study. *J Gerontol Ser A*. (2014) 69:1415–21. doi: 10.1093/gerona/glu155
21. Toosizadeh N, Ehsani H, Wendel C, Zamrini E, O'Connor K, Mohler J. Screening older adults for amnestic mild cognitive impairment and early-stage Alzheimer's disease using upper-extremity dual-tasking. *Sci Rep*. (2019) 9:1–11. doi: 10.1038/s41598-019-46925-y
22. Dorsey ER, Papapetropoulos S, Xiong M, Kieburtz K. The first frontier: digital biomarkers for neurodegenerative disorders. *Digital Biomark*. (2017) 1:6–13. doi: 10.1159/000477383
23. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *NPJ Digital Med*. (2019) 2:1–9. doi: 10.1038/s41746-019-0084-2
24. Rabinowitz I, Lavner Y. Association between finger tapping, attention, memory, and cognitive diagnosis in elderly patients. *Percept Motor Skills*. (2014) 119:259–78. doi: 10.2466/10.22.PMS.119c12z3
25. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The trajectory of gait speed preceding mild cognitive impairment. *Arch Neurol*. (2010) 67:980–6. doi: 10.1001/archneurol.2010.159
26. Stringer G, Couth S, Brown L, Montaldi D, Gledson A, Mellor J, et al. Can you detect early dementia from an email? A proof of principle study of daily computer use to detect cognitive and functional decline. *Int J Geriatr Psychiatry*. (2018) 33:867–74. doi: 10.1002/gps.4863
27. Iakovakis D, Hadjidimitriou S, Charisis V, Bostantzopoulou S, Katsarou Z, Hadjileontiadis LJ. Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease. *Sci Rep*. (2018) 8:1–13. doi: 10.1038/s41598-018-25999-0
28. Mastoras RE, Iakovakis D, Hadjidimitriou S, Charisis V, Kassie S, Alsaadi T, et al. Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Sci Rep*. (2019) 9:1–12. doi: 10.1038/s41598-019-50002-9
29. Van Waes L, Leijten M, Mariën P, Engelborghs S. Typing competencies in Alzheimer's disease: an exploration of copy tasks. *Comput Hum Behav*. (2017) 73:311–9. doi: 10.1016/j.chb.2017.03.050
30. Vizer LM, Sears A. Classifying text-based computer interactions for health monitoring. *IEEE Pervas Comput*. (2015) 14:64–71. doi: 10.1109/MPRV.2015.85
31. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimer's Dis*. (2016) 49:407–22. doi: 10.3233/JAD-150520
32. Folstein MF, Folstein SE, McHugh PR. '*Mini-Mental State*': *A Practical Method for Grading the Cognitive State of Patients for the Clinician*. Pergamon (1975). doi: 10.1037/t07757-000
33. Hutton T. *Alzheimer's Disease in Conn's Current Therapy*. Philadelphia, PA: Psychopharmacol Bull (1990).
34. Kounti F, Tsolaki M, Kiosseoglou G. Functional cognitive assessment scale (FUCAS): a new scale to assess executive cognitive function in daily life activities in patients with dementia and mild cognitive impairment. *Hum Psychopharmacol*. (2006) 21:305–11. doi: 10.1002/hup.772

35. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. (1961) 4:561–71. doi: 10.1001/archpsyc.1961.01710120031004
36. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res*. (1982) 17:37–49. doi: 10.1016/0022-3956(82)90033-4
37. TypeOfMood. *i-PROGNOSIS Eu H2020 Funded Project* (2019). Available online at: https://play.google.com/store/apps/details?id=typeofmood.ime&hl=el
38. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Int Med*. (2001) 16.9:606–13
39. Prokopidis P, Georgantopoulos B, Papageorgiou H. A suite of natural language processing tools for Greek. In: *The 10th International Conference of Greek Linguistics*. Komotini (2011).
40. CoNLL-U. *CoNLL-U Format* (2019). Available online at: https://universaldependencies.org/format.html
41. NLTK. *Natural Language Toolkit* (2020). Available online at: https://www.nltk.org
42. Brunet E. Le Vocabulaire de Jean Giraudoux. In: *Structure et Evolution*. Vol. 1. Geneve: Slatkine (1978).
43. Honoré A. Some simple measures of richness of vocabulary. *Assoc Liter Linguist Comput Bull*. (1979) 7:172–7.
44. Liu H. Dependency distance as a metric of language comprehension difficulty. *J Cogn Sci*. (2008) 9:159–91. doi: 10.17791/jcs.2008.9.2.159
45. Iakovakis D, Hadjidimitriou S, Charisis V, Bostanjopoulou S, Hadjileontiadis L. *Keystroke Timing and Pressure Data Captured During Touchscreen Typing by Early Parkinson's Disease Patients and Healthy Controls* (2019). Available online at: http://zenodo.figshare.com/articles/Keystroke_timing_and_pressure_data_captured_during_touchscreen_typing_by_early_Parkinson_s_disease_patients_and_healthy_controls/7854650/1
46. Fahn S. Members of the UPDRS development committee. Unified Parkinson's disease rating scale. *Recent Dev Parkinson's Dis*. (1987) 2:293–304.
47. Iakovakis D, Ray C, Klingelhoefer L, Bostantzopoulou S, Katsarou Z, Trivedi D, et al. Screening of Parkinsonian subtle fine-motor impairment from touchscreen typing via deep learning. *Sci Rep*. (2020) 10. doi: 10.1038/s41598-020-69369-1
48. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biometr J*. (2005) 47:458–72. doi: 10.1002/bimj.200410135
49. Ahmed S, Haigh AMF, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*. (2013) 136:3727–37. doi: 10.1093/brain/awt269
50. Cheung H, Kemper S. Competing complexity metrics and adults' production of complex sentences. *Appl Psycholinguist*. (1992) 13:53–76. doi: 10.1017/S0142716400005427
51. Louis E, Schupf N, Manly J, Marder K, Tang M, Mayeux R. Association between mild parkinsonian signs and mild cognitive impairment in a community. *Neurology*. (2005) 64:1157–61. doi: 10.1212/01.WNL.0000156157.97411.5E
52. Iakovakis D, Hadjidimitriou S, Charisis V, Bostantjopoulou S, Katsarou Z, Klingelhoefer L, et al. Motor impairment estimates via touchscreen typing dynamics toward Parkinson's disease detection from data harvested in-the-wild. *Front ICT*. (2018) 5:28. doi: 10.3389/fict.2018.00028
53. de Paula JJ, Albuquerque MR, Lage GM, Bicalho MA, Romano-Silva MA, Malloy-Diniz LF. Impairment of fine motor dexterity in mild cognitive impairment and Alzheimer's disease dementia: association with activities of daily living. *Braz J Psychiatry*. (2016) 38:235–8. doi: 10.1590/1516-4446-2015-1874
54. Kobayashi-Cuya KE, Sakurai R, Sakuma N, Suzuki H, Yasunaga M, Ogawa S, et al. Hand dexterity, not handgrip strength, is associated with executive function in Japanese community-dwelling older adults: a cross-sectional study. *BMC Geriatr*. (2018) 18:192. doi: 10.1186/s12877-018-0907-z
55. Middleton FA, Strick P. Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res Rev*. (2000) 31:236–50. doi: 10.1016/S0165-0173(99)00040-5
56. IJmker T, Lamoth CJ. Gait and cognition: the relationship between gait stability and variability with executive function in persons with and without dementia. *Gait Posture*. (2012) 35:126–30. doi: 10.1016/j.gaitpost.2011.08.022
57. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess Alzheimer's disease progression and intervention efficacy. *Front Aging Neurosci*. (2015) 7:102. doi: 10.3389/fnagi.2015.00102
58. FDA. Digital Health Innovation Action Plan. In: *FDA Center for Devices and Radiological Health* (2017).

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 19 November 2020
doi: 10.3389/fdgth.2020.571839

Knee Acoustic Emissions as a Digital Biomarker of Disease Status in Juvenile Idiopathic Arthritis

Daniel C. Whittingslow[1,2]*, *Jonathan Zia*[2,3], *Sevda Gharehbaghi*[3], *Talia Gergely*[2], *Lori A. Ponder*[2], *Sampath Prahalad*[2†] *and Omer T. Inan*[3†]

[1] Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Emory University, Atlanta, GA, United States, [2] Emory University School of Medicine, Atlanta, GA, United States, [3] Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Hemant Ghayvat, Fudan University, China
Theodora Chaspari, Texas A&M University, United States

***Correspondence:**
Daniel C. Whittingslow
dwhittingslow@gmail.com

[†] These authors share senior authorship

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: 12 June 2020
Accepted: 22 October 2020
Published: 19 November 2020

Citation:
Whittingslow DC, Zia J, Gharehbaghi S, Gergely T, Ponder LA, Prahalad S and Inan OT (2020) Knee Acoustic Emissions as a Digital Biomarker of Disease Status in Juvenile Idiopathic Arthritis. Front. Digit. Health 2:571839. doi: 10.3389/fdgth.2020.571839

In this paper, we quantify the joint acoustic emissions (JAEs) from the knees of children with juvenile idiopathic arthritis (JIA) and support their use as a novel biomarker of the disease. JIA is the most common rheumatic disease of childhood; it has a highly variable presentation, and few reliable biomarkers which makes diagnosis and personalization of care difficult. The knee is the most commonly affected joint with hallmark synovitis and inflammation that can extend to damage the underlying cartilage and bone. During movement of the knee, internal friction creates JAEs that can be non-invasively measured. We hypothesize that these JAEs contain clinically relevant information that could be used for the diagnosis and personalization of treatment of JIA. In this study, we record and compare the JAEs from 25 patients with JIA—10 of whom were recorded a second time 3–6 months later—and 18 healthy age- and sex-matched controls. We compute signal features from each of those record cycles of flexion/extension and train a logistic regression classification model. The model classified each cycle as having JIA or being healthy with 84.4% accuracy using leave-one-subject-out cross validation (LOSO-CV). When assessing the full JAE recording of a subject (which contained at least 8 cycles of flexion/extension), a majority vote of the cycle labels accurately classified the subjects as having JIA or being healthy 100% of the time. Using the output probabilities of a JIA class as a basis for a joint health score and test it on the follow-up patient recordings. In all 10 of our 6-week follow-up recordings, the score accurately tracked with successful treatment of the condition. Our proposed JAE-based classification model of JIA presents a compelling case for incorporating this novel joint health assessment technique into the clinical work-up and monitoring of JIA.

Keywords: wearable sensors, machine learning, juvenile idiopathic arthiritis, acoustic sensing, signal processing

INTRODUCTION

Juvenile idiopathic arthritis (JIA) describes a heterogeneous group of arthritides that present in children. JIA encompasses all forms of arthritis that begin before a patient is 16 years old, lasts for at least 6 weeks, and are of an unknown origin. It is a leading cause of disability and the most common chronic rheumatic disease of childhood with a prevalence of 150 cases per 100,000 (1). It is an autoimmune disorder with a complex etiology thought to be related to a combination of pre-disposing genetic factors and environmental influence (2, 3).

The heterogeneity of presentation sometimes makes diagnosing JIA difficult. This difficulty is exacerbated by the lack of conclusive, diagnostic laboratory tests. Diagnosis currently relies on taking a thorough history, physical exam, and several laboratory and imaging studies (4). Once diagnosed, to select the most suitable treatment for JIA, the disease should be classified into its subtype. JIA is divided into seven subtypes based on laboratory and clinically observed features (5, 6). To determine the most appropriate subtype, and thus the most effective therapy, an extensive workup must be performed on each patient. This is a time and resource intensive process. These workups include the history and physical exam, as well as a full blood exam. Imaging studies are also commonly used to grade the disease. After diagnosis, the goal is to enable the child to resume normal childhood activities with normal growth and development (4). Managing JIA requires a combination of pharmacological interventions, physical and occupational therapy, and psychosocial support. The pharmacological treatment may involve corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDS), or disease-modifying anti-rheumatic drugs (DMARDs) including biological response modifiers (7–9). This treatment protocol is largely reactive with decisions made based on subjective and qualitative measures of response to therapy.

Early diagnosis with effective treatment is necessary for preventing the long-term sequela of JIA (4). However, JIA's highly variable presentation, symptomatology and course make diagnosis and selection of the most suitable treatment difficult. Pediatric rheumatologists are most well-suited for diagnosing and treating JIA; however, there is currently a severe shortage of pediatric rheumatologists. As of 2019, there are fewer than 400 board-certified and practicing pediatric rheumatologists in the United States. This shortage contributes to only one in four children with JIA being able to regularly see a pediatric rheumatologist (10, 11). To address the difficulty of diagnosis, subjectivity of treatment, and severe lack of access to pediatric rheumatologists, more research must be performed in to develop objective biomarkers of JIA. A suitable biomarker could help more effectively diagnose patients, identify risk profiles, and predict/track an individual's response to treatment. Additionally, the development of such a biomarker could allow for more effective translation of the many genetic and immunological mechanistic studies of the disease to further improve clinical outcomes. Ideally, this biomarker would also be readily measurable with affordable technologies, so that JIA could be easily diagnosed and monitored by non-specialist healthcare workers.

The use of acoustics—recording the sounds that the joints make during movement—could provide a basis for developing such a biomarker (12). These sounds, or joint acoustic emissions (JAEs), can be readily measured on the surface of the skin and have shown promise in diagnosing joint pathologies and injuries. Most existing research into JAEs has focused on developing diagnostic techniques to differentiate "healthy" vs. "unhealthy" joints (13, 14). In one study, osteoarthritic knees were found to produce more frequent, louder, and longer duration acoustic emissions when compared against healthy knees (15). In the case of a chronic condition—such as JIA—JAEs could serve as a means of not only diagnosing but also longitudinally monitoring the conditioning. If JAEs show a correlation with disease status in JIA, they could regularly be measured to help personalize the management of JIA. Until recently, longitudinal assessment using JAEs in healthcare was not feasible due to a lack of technologies for recording JAEs outside of a laboratory or clinical setting. However, the development and application of piezoelectric accelerometers to JAE assessment has substantially advanced the field. This type of sensor is sensitive to physical vibrations (such as those seen on the skin during joint articulation), but does not substantially record external noises (16). JAE assessment technologies if properly applied to JIA, could lead to earlier diagnosis, improved and personalized care, and could serve as an objective measure in the next generation of clinical trials.

In this paper, we explore the potential of using JAE analysis to diagnose and longitudinally track JIA. In this work, JAEs were recorded from the knees - one of the most commonly affected joints in JIA (17, 18). Our team recently showed that by damaging the meniscus in a cadaver model of the knee, the resulting JAEs were substantially altered (19). In the case of JIA, affected joints are characterized by persistent joint swelling caused by an accumulation of synovial fluid and thickening of the synovial lining (3) (**Figure 1A**). We hypothesize that these pathologic changes in the knee will similarly alter JAE profile of the knee. If that hypothesis is supported, the JAEs of the knee could then be correlated with disease status.

To test this hypothesis, first, we built a custom hardware and software setup for recording JAEs and designed a novel signal analysis algorithm that windows the JAE recording based on the cycles of flexion/extension (**Figures 1B–D**). We placed two piezoelectric accelerometers medial and lateral to the distal patellar tendon, and an inertial measurement unit (IMU) around the ankle. With the hardware in place, the subject performs 10 flexion/extension cycles. The JAEs from the knees of two groups of children are recorded: one group had active JIA and the other was an age- and sex-matched healthy control group. To assess the effectiveness of JAEs for tracking therapeutic efficacy and changes in disease status, we also recorded the JAEs from the children with JIA 6 weeks after successful treatment. Our

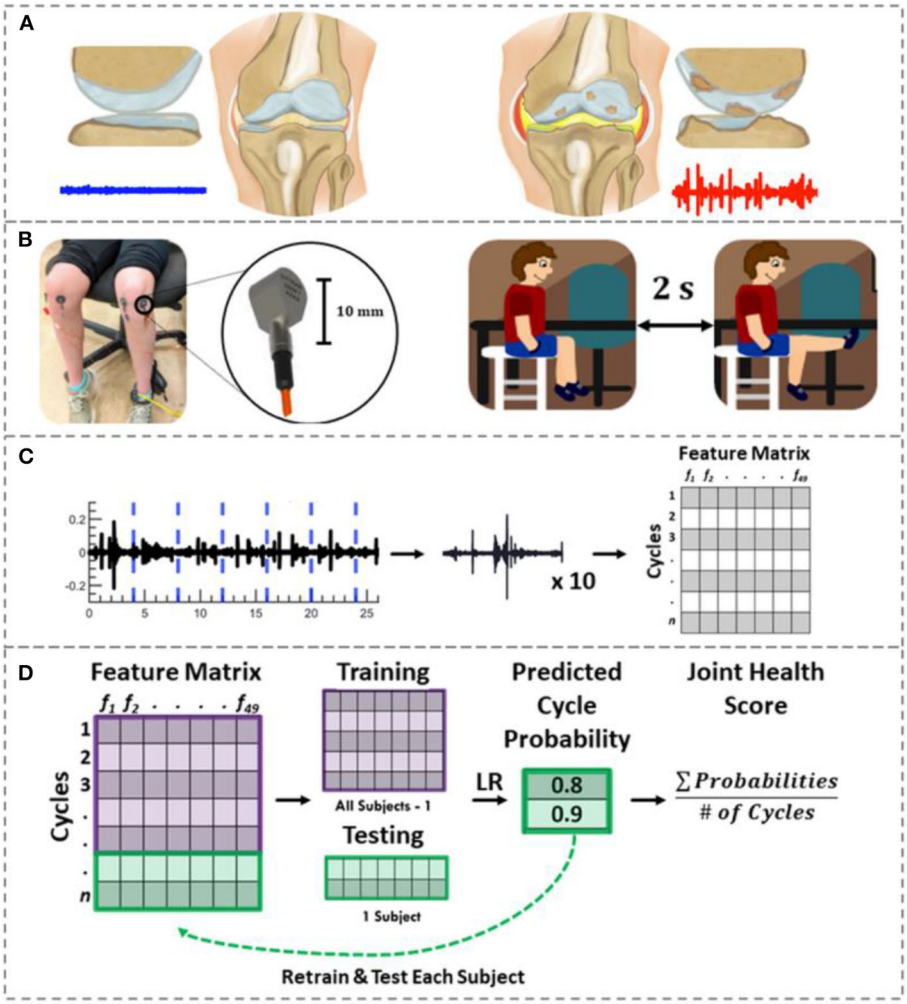

FIGURE 1 | Joint acoustic emission overview. **(A)** A healthy knee articulates smoothly due to its smooth cartilage and appropriate amount/constituency of synovial fluid. This smooth articulation creates a noise-like JAE (blue). In JIA, thickened/inflamed synovium with excessive joint effusions, cartilage loss and/or bone erosions may be observed. These changes are hypothesized to create a JAE with several large spikes (red). **(B)** To record the knee JAEs, two contact accelerometers were placed on each child's knees. They viewed and replicated the movements in an instructional cartoon during JAE recording such that their movement speed and range of motion was controlled. **(C)** The resulting JAEs were split into their approximately ten component cycles. Forty-nine features were calculated to describe these cycles. The features, subject numbers, and clinically determined disease status were fit to a feature matrix. **(D)** Using logistic regression and LOSO-CV, the probability of each cycle belonging to JIA were calculated. The average of those cycle probabilities is used as a "joint health score" to indicate the severity of JIA. If the majority of cycles for a given subject had a probability of JIA $\geq$ 0.5, that subject was classified as having JIA.

proposed algorithm, powered by logistic regression, analyses 49 signal features (summarized in **Supplementary Table 1**) of each individual cycle of flexion/extension and outputs the probability that a cycle belongs to a patient with JIA. This output probability forms the basis for our proposed JIA digital biomarker. Finally, we assess the importance of each signal feature in the algorithm as well as the accuracy and generalizability of the model using leave-one-subject-out cross-validation (LOSO-CV).

RESULTS

Qualitative Comparison of Knee JAEs

The JAEs were recorded from the knees of two groups of children. One group had actively inflamed knees with either newly diagnosed or poorly controlled JIA as diagnosed by their treating pediatric rheumatologist; the other group was composed of age- and sex-matched health controls with no JIA or known injuries to the knee. There are several notable differences in the time-domain patterns of the JAEs between these groups. The 18 healthy controls had no noticeable peaks in their audio signals and upon listening the recorded JAEs resembled white noise (**Figure 2A**). The 25 subjects with JIA consistently exhibited periodic, high-energy clicks in each flexion-extension cycle. These "clicks" have a spike-like appearance in the time-domain plot which correspond to the high power content in the higher frequency components in spectrogram (**Figure 2B**). Ten of these patients with JIA had a second recording after 6 weeks of treatment as prescribed by their treating pediatric

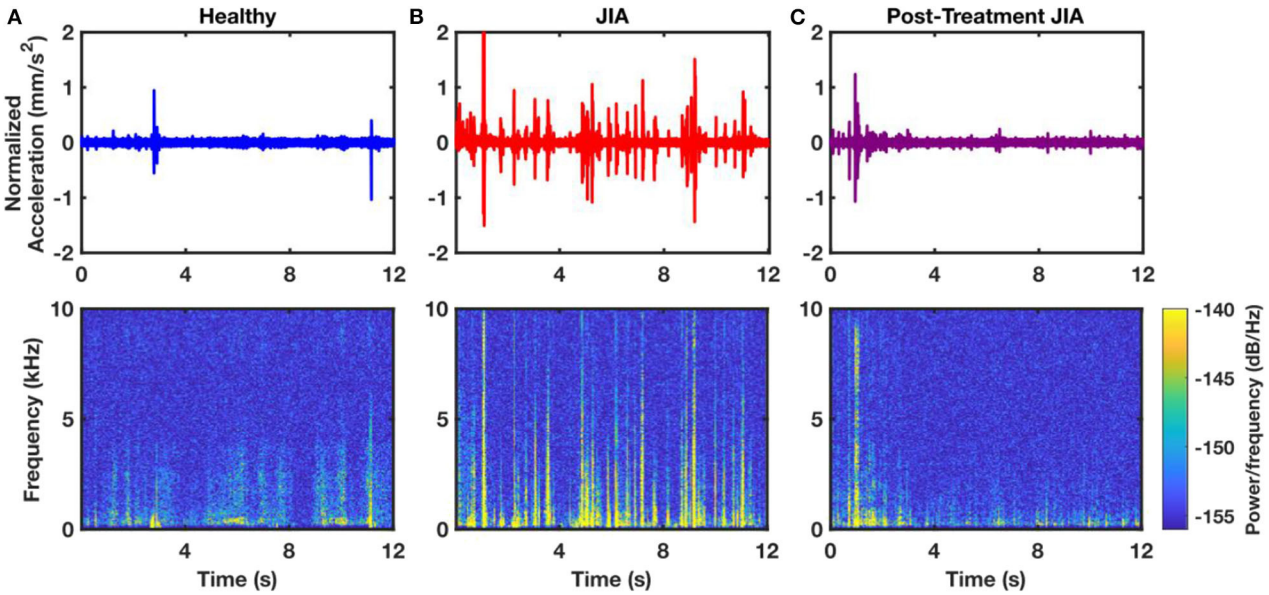

FIGURE 2 | Representative time-domain and spectrogram plots of JAEs from a sample healthy control **(A)**, a subject with active JIA **(B)**, and that same subject after 6-weeks of successful treatment **(C)**. The 12 s of the JAEs represent approximately 4 flexion/extension cycles. Spectrogram of the subject with JIA contains more high power and high frequency components compared to that of healthy and post-treatment subjects.

rheumatologist. The JAEs of this follow-up group showed a large reduction in the amplitude and frequency of the clicks noted during their actively inflamed stage (**Figure 2C**). The post-treatment JAEs more closely resembled the healthy controls both in the time-domain and spectrogram plots of the JAEs as well as in audibly listening to the recordings. A representative subject's JAE recording from each of these groups is presented in **Figure 2.**

Knee Audio Score Classification

The knee audio score for each subject was defined as the probability of a cycle belonging to a subject with JIA. In this manner, a knee score of 0 indicates 0 probability of having JIA, and a score of 1 indicates an actively inflamed joint with JIA. A threshold was set at a score of 0.5 to delineate the classification of the two groups. A threshold cutoff of 0.5 was chosen heuristically but could theoretically be changed to place an emphasis on sensitivity vs. specificity as desired. Subjects' joint scores were calculated by averaging all the computed cycle probabilities of each individual subject's flexion/extension cycles. The subject-level joint scores are presented as a histogram in **Figure 3A**. Notice the heavy overlap between the healthy (blue) and post-treatment, follow-up subjects (purple). This was expected based on the success of the treatment as reported by the treating pediatric rheumatologist. The JIA distribution is centered around a score of 0.82 with clean separation from the other two distributions. The overall cycle-based logistic regression analysis had an accuracy of 82.7% for classifying individual cycles. The receiver operating characteristic (ROC) curve and confusion matrix are presented in **Figures 3B,C**. The ROC curve had an area under the curve (AUC) of 0.899. The cycle classification had a specificity of 80.4%, a sensitivity of 84.5%, an error rate of 20.1%, a positive predictive value (PPV) of 84.7%, and a negative predictive value (NPV) of 90.2%.

Feature Importance Ranking and Model Performance

Logistic regression is a binary classification algorithm that finds the best hyperplane in the feature space which separates the two classes: healthy and JIA (20). The absolute values of the individual feature weights describing that hyperplane are used to quantify the impact that each feature has on the model and thus its importance. **Figure 4A** shows the relative importance of the top 20 features used in computing the knee health score. Of note, the majority of these features for classifying the two classes are in the spectral domain which agrees with the results from our earlier pilot work on the topic (12).

Next, the number of features and cycles were varied to quantify the change in accuracy that the inclusion of each consecutively less important feature and each recorded cycle had on the classification accuracy of each subject. The output of this testing is visualized as an accuracy heatmap in **Figure 4B** where the color represents the average accuracy from testing on each subject in the dataset using LOSO-CV using the depicted number of features and cycles of movement. At the bottom left of this plot is the accuracy of the model when only trained on the most important feature—the mean spectral spread—and tested on just one randomly selected cycle of flexion/extension from the subject. All permutations of possible cycle selection were performed and averaged to yield the accuracy under these conditions. In the case of just one cycle and one feature, the average cycle classification accuracy was only 11.1%. Ascending along the y-axis, one feature is consecutively added based on

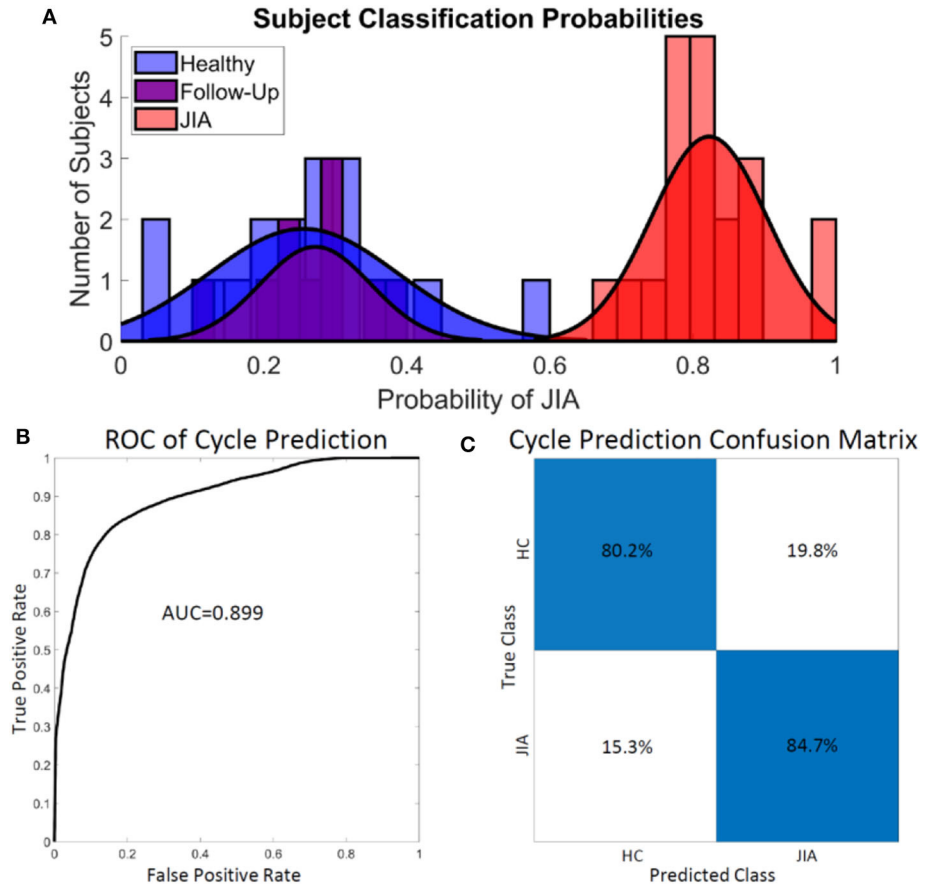

FIGURE 3 | Assessing the performance of the logistic regression classifier on subjects **(A)** and cycles **(B,C)**. **(A)** There was little overlap in the computed joint health score of the healthy control group and the group with JIA. A sub-group from the JIA group after effective treatment had JIA scores heavily overlapping with the healthy control group at follow-up. **(B,C)** The logistic regression model overall classified the individual cycles accurately 82.7% of the time. The model achieved adequately high sensitivity (84.5%) and specificity (80.4%). HC, healthy control.

its relative importance, such that at the top left corner of the heatmap the model has been trained on the top 20 most important features. Still, when tested with only one cycle from a subject, the accuracy remains low at 25.0%. From left to right, the algorithm is tested on an increasing number of cycles recorded from a subject. The model has an accuracy of 42.8% in the bottom right corner, where it was trained on just the mean spectral spread and tested using all recorded cycles of a subject from all four microphones. The algorithm had the highest accuracy of 80.6% when trained on the top 20 most important features and tested using all recorded cycles. This is slightly <82.7% observed in **Figure 3**. This discrepancy is because the model in **Figure 3** had the added benefit to the classification of all 49 features (**Supplementary Table 1**), not only the top 20 most important (21).

Knee Audio Score's Longitudinal Health Tracking Capability

The knee audio scores were calculated for 10 of the subjects with JIA before and after 3–6 months of treatment. At first visit, these subjects were either newly diagnosed with JIA, or having a resurgent flare of arthritis. Their treatments were prescribed according to the current clinical standards by their treating pediatric rheumatologist and were recorded but not controlled for in this study. Every subject at follow-up reported a reduction in symptoms and the treating physician reported an overall improvement of the arthritis. In **Figure 5**, the calculated joint health scores are shown before and after treatment for this cohort. The average joint health score at initial visit was 0.84 ± 0.08. At follow-up, the scores dropped to an average of 0.19 ± 0.09. This drop in joint health scores is statistically significant with a p-value $= 5.3 \times 10^{-8}$ (tstat $= 16.4$, 9° of freedom), when tested with a one-tailed t-test. The scores from individual subjects are represented with dashed lines in **Figure 5** and in all cases mirror the clinical assessment of their improvement.

DISCUSSION

There is a compelling need for the development of a non-invasively measurable biomarker that can both diagnose and

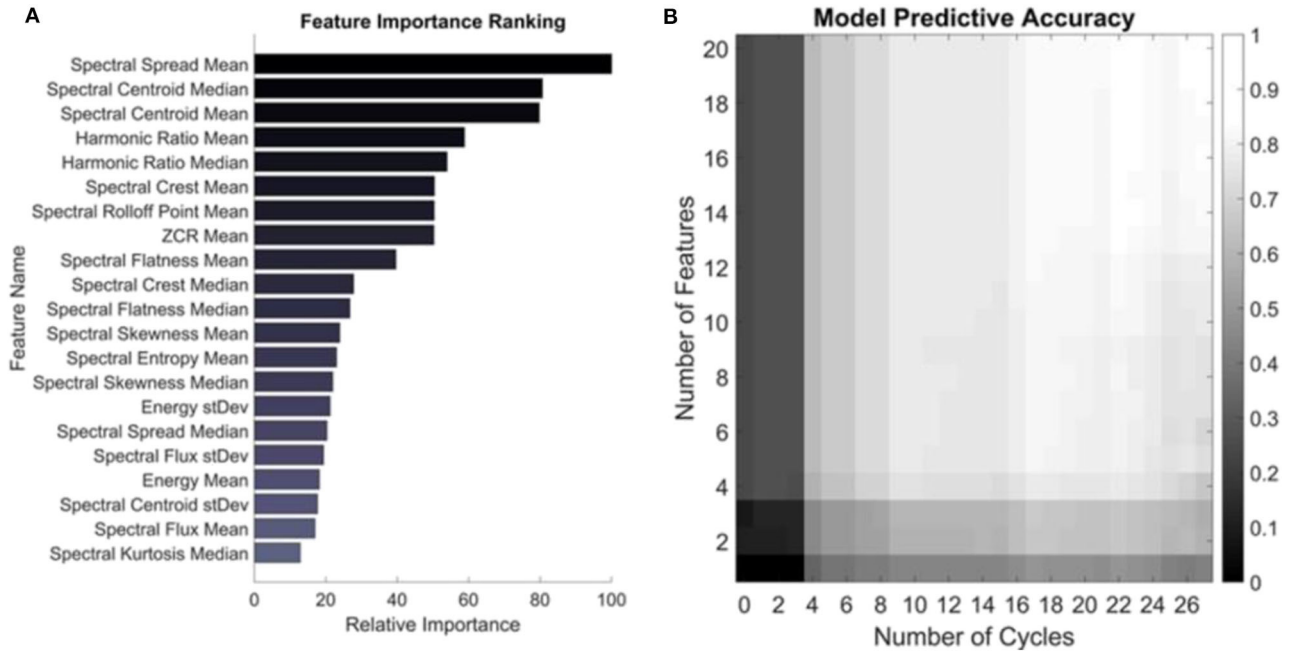

FIGURE 4 | Feature importance and model performance based on number of features and cycles. **(A)** Features are ranked based on their weighted coefficients as output by the trained logistic regression model. The most important feature was the mean spectral spread. **(B)** The model was trained on a feature set containing just one and up to 20 of the top features and the accuracy was assessed based on including those features and number of cycles recorded from a subject. The colors represent the average accuracy across all subjects for all permutations of cycle selection for a given set of testing parameters. The maximum accuracy of 80.6% is seen in the top right corner when trained on the 20 most important features and tested on all cycles of a given subject.

track the status of affected joints in JIA. JIA is a chronic, autoimmune disease of childhood with a highly variable presentation, an etiology linked to genetics and environment, and a complex treatment strategy (9). Assuming a child is properly diagnosed, determining which treatment regimen will work best for them is largely reactive. A certain course of treatment is prescribed and adjusted based on patient-reported feedback and infrequent clinical assessments. In this work, we explore the impact that JAE monitoring could have on the diagnosis and treatment of JIA. If JAEs were found to contain clinically relevant information, they could potentially be used as an initial screening tool by primary care medical professionals – reducing the burden on the healthcare system of unnecessary referrals to specialists. Furthermore, this could help diagnose patients earlier, which may prevent the long-term sequelae of JIA (17). After diagnosis, if joint sounds were found to closely track with treatment efficacy and joint health longitudinally, they could be used as an objective biomarker to decide or even predict the most effective course of treatment. This would reduce the burden of frequent JIA flare-ups on patients and allow for a tightening of the treatment feedback loop leading to overall better management of the condition.

In this study, the effects of JIA on the JAEs produced by articulation of the knee were explored. The study population was made up of 43 subjects, 25 of whom had JIA, and 10 of these 25 subjects had repeat recordings 6 weeks after the initial visit. The JAEs from a pediatric population with JIA of this size have never before been compiled and analyzed. These JAEs were first compared qualitatively to better visualize the differences in the recordings as seen in **Supplementary Table 1**. It was noted that there are characteristic high frequency clicks in the JAEs of subjects with JIA, that fade away with successful treatment and are not present in matched healthy controls' JAEs. More work is needed to determine the precise mechanistic origin of these high frequency clicks, but we hypothesized that they occur due to increased internal friction in the joint, caused by the characteristic inflammation of the synovial membrane, breakdown of cartilage, and reduced joint space in JIA (3, 22). Of note, similar clicks are apparent in the case of acute injury as was recently discovered by our work in a cadaver model of knee injury (19) and a similar study in an injured athlete model (23). Rather than relying strictly on one or even a few characteristics of these JAEs as was done in previous work, in this study we attempt to more thoroughly quantify the differences between the recorded JAEs. We do this by splitting the joint sound recordings from each subject into their component flexion/extension cycles. On each cycle, 49 features (from the spectral and time-domain) were calculated to describe the observed JAEs. These features and cycles were organized into a feature matrix which was used to train a machine learning, classification model using logistic regression. This technique should provide a more exhaustive analysis of the features of the JAEs, and overall be more generalizable than past efforts to interpret JAEs. The results of this model are described below.

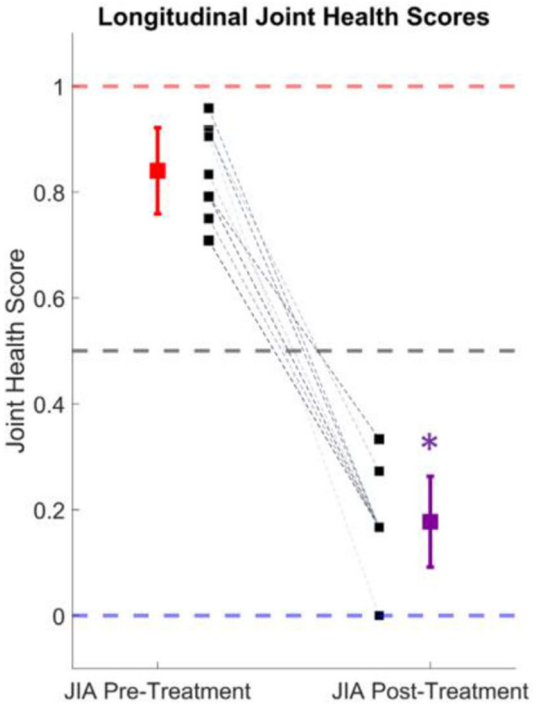

FIGURE 5 | Longitudinal joint health score tracking. The average joint health score, which describes the probability of having JIA, dropped from 0.84 ± 0.08 to 0.19 ± 0.09 after successful treatment of the condition in 10 subjects. The individual subject scores are denoted by the black squares and dashed lines. The mean and standard deviation of the actively inflamed subjects with JIA is shown in red, and the purple marker indicates the mean and standard deviation at follow-up. This drop in joint health score was statistically significant ($p < 0.001$).

Knee Audio Score Classification

Logistic Regression and linear discriminant analysis are two of the most widely used statistical methods for analyzing categorical outcome variables. Because logistic regression is more flexible and robust than linear discriminant analysis when considering the assumptions made about underlying data, it is commonly used in medical data binary classification tasks (24). When compared against more complex machine learning models, the modeling parameters in logistic regression are generally easier to interpret rather than a "black-box" approach. This flexibility, robustness, and interpretability should encourage more widespread acceptance of the conclusions provided in this work by the medical research community (20). Logistic regression is a binary classification algorithm that attempts to find the best hyperplane in k-dimensional space for separating the two classes (e.g., healthy and JIA), while minimizing logistic loss (20).

In our application, the logistic regression outputs the probability that a given test cycle belongs to the healthy or JIA class. We have also proposed that the output JIA probability could be used as a basis for quantifying knee joint health. In this paradigm, a probability of 0 indicates a healthy knee with no signs of JIA, whereas a score of 1 indicates a knee clearly affected with JIA. The classification accuracy of the model is presented in *B*. First, the subject-level classification histogram showed clear separation of the joint health scores when the 0.5 classification threshold was applied to the output probabilities (**Figure 3A**). This finding helps support the idea that knee JAEs could be used as part of the screening and diagnosis of JIA. The accuracy of labeling each cycle is then quantified to better understand the performance of the logistic regression model (**Figures 3B,C**). The overall accuracy of the cycle labeling was 82.7%, which corresponds to a sensitivity of 84.5% and a specificity of 80.4%. As discussed, JIA is challenging to diagnose not only due to the highly variable nature of the condition and presentation, but also because of the shortage of pediatric rheumatologists who are specially trained to identify the disease. One potential use of JAE-based assessment in JIA is to allow for better screening of the condition by healthcare providers that are less trained to identify it. JAE based assessment is entirely non-invasive and achievable with affordable hardware. The high sensitivity of this technique means that few false positive test results will occur. The technique may be slow to be adopted for final diagnosis, but in the near-future JAEs could at least be used as a preliminary screening tool that gates whether a patient should pursue a specialist consult for further diagnostic workup (i.e., point-of-care screening).

Feature Importance Ranking and Model Performance

To understand the effects of feature selection and length of recording on JIA JAE assessment, we presented our findings on which signal features are most important for the algorithm, and how it performs with less cycles to classify using a subset of features. In our model, there were 49 features describing each cycle of movement from each subject. A feature weights vector of length 49 was output from the model describing the hyperplane that best separates the JIA from healthy labeled cycles. The absolute values of the individual feature weights were used to quantify the importance of a given feature for the model. The relative importance of the top 20 features in the algorithm are presented in **Figure 4A**. Each subject had two microphones on each of their knees recording the JAEs during 10 cycles of flexion/extension at a rate of 1 cycle every 4 s. These four audio files are subdivided into the individual cycles of movement based on the simultaneously recorded motion data captured by the inertial measurement unit (IMU) attached to the subjects' ankles. The resulting data structure thus had approximately 40 segments of data describing one subject's movement. **Figure 4** graphically depicts the results of varying the number of those segments included in the testing dataset. Each square in **Figure 4** describes the average accuracy when each subject was tested with the described parameters as a part of LOSO-CV on the trained model. Along the y-axis, features were sequentially added in order of descending importance, such that at the bottom of the plot, only the most important feature—the mean spectral spread—was used to classify the cycles. Upon ascending the y-axis, each of the 20 features as described in **Figure 4A** are consecutively included in training the logistic regression model. This figure thus depicts the impact that feature selection has on

the accuracy of the classification. There is a clear benefit on the accuracy of the model by including more features, and this should help with the generalization of the model to novel data. In the past, attempts have been made to describe knee JAEs using only one or a few different signal features (14, 25, 26). These attempts generally have success on a small data set, but when applied to a data set of this size were suboptimal when compared to the accuracy of the model proposed in this work.

The impact of the length of the JAE recording is also demonstrated in **Figure 4B** from left to right. Each step to the right includes an additional, and randomly selected, flexion-extension cycle, and the color of the square indicates the accuracy of classifying a subject with that many cycles. On the left, we test the model with only one cycle recorded from one microphone on each subject. On the far right, every cycle recorded for every microphone is used to test any given subject. The impact is similar to increasing the number of features in the trained model – as the number of cycles increases the classification accuracy similarly increases. Note that there is some possible redundancy in having two microphones recording the JAEs from each knee. In this case, the impact of having similar recordings in two of the microphones can be noted by the relative plateau of the accuracies around the 18th recorded cycle (accuracy is no longer substantially increasing with each added cycle). Overall, this analysis demonstrates the impact that the feature selection and length of JAE recording has on the accuracy of the model. In our case, the accuracy was at its lowest with one feature and one cycle at 11.1% and achieved a high of 80.6% with the top 20 most important features and every recorded cycle from a subject. This analysis also demonstrates why past approaches have had only limited success in generalizing their findings. If only a subset of these features were used to describe JAEs, the accuracy would significantly diminish. Many features are needed to fully describe the nature of these sounds and separate the differences between populations. Later work comparing a different clinical scenario, or a larger dataset may find that a different feature is more important for delineating two study groups, but the approach applied in this paper should hopefully provide guiding influence on future assessments of JAEs.

Longitudinal Joint Health Tracking

To discover if knee JAEs had the potential for quantifying joint health longitudinally, 10 subjects with JIA had their JAEs recorded during an active flare-up of the condition and 3–6 months later at their follow-up visit. In this particular cohort, every subject showed clinical improvement and reported a lessening of symptoms. To calculate these subjects' knee scores, the logistic regression model was trained on all subjects not in this cohort. The recordings before and after treatment were tested on the trained model and the knee audio scores computed as described in the section "Knee Audio Score Classification Using Logistic Regression". The hypothesis was that as a child's knees healed from effective treatment, their knee scores would decrease from the JIA range (0.5–1.0) toward the healthy range (0.0–0.5). In all subjects, this hypothesis was shown to be valid. There was a statistically significant drop in the average scores of 0.65, or a 77.4% improvement in the joint health score. This closely tracked with the reported clinical workup of the subjects indicating that joint health scores based on JAEs may be clinically applicable for not only diagnosing JIA (as discussed in section Knee audio score classification), but also monitoring the condition over time.

In this study, these 10 patients represent a subset of the overall JIA population and before claiming how consistently joint sounds track with knee health status in an individual the sample size of those studied should be further increased. However, these findings represent the first time that a population large enough to adequately power a study of children with JIA has been assessed longitudinally. The close correlation between the change in joint sounds and the observed clinical status supports further research into this relationship. Overall, this study represents an early, but important step toward understanding the nature of JAEs. The strong separation of the classes alongside the close tracking of disease activity make it clear that JAEs contain clinically relevant information. This information if properly leveraged could 1 day enable better more personalized treatment of JIA.

Limitations and Steps to Clinical Adoption

JIA is a chronic condition that affects multiple joints in the body. The knee is one of the most commonly affected joints and made for a viable target for this attempt at analyzing JAEs. To better understand the clinical utility of this sensing modality, JAEs should be studied in other commonly affected joints in JIA. Additionally, the sensitivity of this method should be compared against the performance of the current clinical standard procedure for diagnosing and staging the condition, as well as against other modalities such as magnetic resonance imaging or ultrasound, which typically are time consuming and expensive. Treatment of JIA seeks to reduce the frequency of acute, symptomatic flare-ups, and to ultimately achieve clinical remission. In this study, the treatments our subjects underwent were not controlled for due to the small sample size. In the future, the effectiveness of therapy should be quantified using a prospective study design. Additionally, in this cohort all subjects improved with treatment and we observed a corresponding drop in the joint health score. Since no patients got worse at follow-up, we were unable to discover if JAE assessment could track worsening of the condition. The sensitivity of joint sounds for detecting not only different severities of the condition but also the course of the condition should also be assessed. JAEs would be significant clinically if they were able to determine the difference between an acutely inflamed joint and a more chronic, undiagnosed state. Determining that duration of disease activity would help with selecting the ideal treatment for a patient. Classifying subjects into the different subtypes of JIA and delineating joint sounds caused by JIA vs. all other causes would also offer clinical merit. This study was performed on a fairly large sample size of subjects to date, and enrollment is ongoing to support future work. Increasing the number of subjects would better support the generalizability as well as mitigate possible overfitting of the discussed results. Overall, in this paper we present JAEs as a novel technique for analyzing the health of a joint in JIA. The findings in this paper present significant clinical merit to this type of analysis, but there is still much to be discovered.

MATERIALS AND METHODS

Human Subject Protocol and Subject Demographics

The study was conducted under a protocol approved by the Georgia Institute of Technology and Emory University Institutional Review Boards. Forty-three subjects participated in this study after completing a written informed consent. Twenty-five of the subjects were diagnosed with JIA by a pediatric rheumatologist and 18 of the subjects were healthy controls with no history of JIA or acute knee injuries. The group with JIA consisted of 20 females and five males (12.2 ± 3.1 years old, BMI 20.1 ± 4.1 kg/m^2). The healthy control group consisted of 15 females and three males (12.9 ± 2.7 years old, BMI 22.3 ± 2.8 kg/m^2) with no history of joint disease, surgery or significant joint injury. To capture longitudinal changes in the knee JAEs during the course of treatment, data were acquired from 10 of the subjects (1 male, 9 female, 12.5 ± 3.3 years old, BMI 20.8 ± 3.5 kg/m^2) with JIA a second time, 3–6 months after initial measurements (follow-up group). Note, that JIA is more prevalent in females with estimates ranging from 65–78% of all cases occurring in females, thus the demographics of this study were selected accordingly to match this distribution as closely as possible (27, 28).

The data acquisition set up for each subject is shown in **Figure 1B.** To record the sounds produced by the joints, two uniaxial analog accelerometers (3225F7, Dytran Instruments Inc. Chatsworth, CA) were attached 2 cm medial and lateral to the distal patellar tendon using double-sided adhesive pads (Rycote Microphone Windshields Ltd, Stroud, Gloucestershire, GL5 1RN, United Kingdom) on both knees. These professional-grade pads tightly coupled the accelerometer to the subject's knee. This accelerometer has a broad bandwidth (2 Hz−10 kHz), high sensitivity (100 mV/g), low noise floor (700 μgrms), miniature size and low weight (1 gram). This accelerometer placement location has been shown to allow for the of capture high-fidelity signals capable of differentiating meniscus injury status in an JAE cadaver model (19).

To record the knee JAEs, each subject performed 10 unloaded knee flexion/extension exercises, while seated on a height-adjustable stool to prevent foot contact with the ground. The subjects repeated the movement as seen on an instructional cartoon that encouraged a cycle to be completed every 4 s through the full range of motion (RoM) of each subject. The signals from the accelerometer were sampled at 100 kHz and recorded using a data acquisition module (USB-4432, National Instruments Corporation, Austin, TX). An inertial measurement unit (IMU) attached around the ankle of the subject recorded synchronous positional data during JAE recording at 50 Hz to allow for analysis on a cycle-by-cycle basis, as well as to ensure the subject maintained an appropriate speed and RoM. The ideal speed and angles to move through have previously been explored using a cadaver model of JAEs (19). The exercise and recording protocol were repeated for both knees for all subjects. The recorded signals were analyzed using Matlab (MathWorks, Natick, MA).

Signal Processing and Feature Extraction

The JAEs were analyzed in the time and frequency domains. **Figure 2** shows a representative plot of the time domain signal after bandpass filtering from one subject with JIA, that subject's JAEs at their 3-months follow-up visit, and a healthy, matched control's JAE recording. It is notable that the number of spikes in the time domain of the patient with JIA went down with effective treatment as seen at follow-up to more closely resemble the JAE recording from the healthy control. The JAEs from these subjects have high bandwidth frequency content as expected from earlier pilot work (25, 29, 30). **Figure 1C** graphically depicts the signal analysis workflow for knee JAEs. The signals are pre-processed using a digital finite impulse response (FIR) band-pass filter with 250 Hz−10 kHz bandwidth. The bandwidth employed in this filtering is based on prior work: at the low end, the cutoff of 250 Hz is selected to reduce low frequency artifacts and muscle sounds (<100 Hz) while preserving the sub kHz friction-generated components of the sounds; at the high end, the cutoff of 10 kHz is selected to remove high frequency artifacts while still preserving the kHz range of frequencies responsible for the acoustic emissions that are observed from the joint. To segment the JAE data into individual flexion/extension cycles, an FIR low-pass filter (5 Hz) is applied to the raw JAE signals to visualize the movement of the knee through its RoM. This motion data is compared against the synchronized IMU data and the proper indices for the beginning and end of each flexion/extension cycle were selected. These individual cycles were separated and subdivided into 400 ms long frames. This frame (or window) length was selected to provide sufficient width to capture lower frequency information while still providing multiple frames per flexion/extension cycle. A total of 49 signal features are extracted from each frame for each microphone, comprising features that—in our group's prior work, and in audio processing and classification work in other domains—have been found to contain salient information. Feature descriptions are available in **Supplementary Table 1**. The 10 frames corresponding to one cycle are averaged to give 49 descriptors of each cycle of flexion/extension. This process was repeated for all four microphones – two on each knee. These feature sets were stored in the row-matrix, $\boldsymbol{X}$. The rows of $\boldsymbol{X}$ each represent a single cycle of movement as recorded from each microphone, and the columns represent each of the 49 features extracted. The matrix $\boldsymbol{X}$ was standardized to zero mean and unit variance by subtracting the mean of each column and dividing by its standard deviation (see Feature Matrix in **Figure 1C**).

The features extracted can be categorized into two groups: either time domain or spectral features. The time domain features include the zero-crossing rate (ZCR), energy, root-mean-square (RMS) amplitude, and entropy. The frequency characteristics of the joint sounds are described by the spectral features including the spectral centroid, spectral flux, spectral density, spectral roll-off, spectral spread, and spectral entropy (A full list of the features is available in **Supplementary Table 1**.)The mean, standard deviation, and coefficient of variance are all computed for the set of 400 ms windows on each cycle to better classify these features. This approach using these particular features to classify

joint JAEs is based on the appearance and sound of the signals, and their selection was supported by previous pilot work on this topic (12, 31).

Knee Audio Score Classification Using Logistic Regression

With the data appropriately organized, we trained a logistic regression classification model. Logistic regression is a common statistical machine learning technique for binary classification problems (e.g., healthy vs. JIA). At the core of this algorithm is the logistic function, which was originally developed by ecologists to describe population growth – it is a sigmoidal curve that rises quickly and levels off at a given environment's carrying capacity (32, 33). The algorithm uses this function to map any real number input to a value between 0 and 1.

$$\frac{1}{(1+e^{-1})} \tag{1}$$

Logistic Function

In logistic regression, the input values ($x_1 \ldots x_n$) are combined linearly to predict an output value (y) using weighted coefficients ($b_0 \ldots b_n$). However, unlike linear regression, in logistic regression the output values being predicted are binary (0 or 1, or in our case healthy or JIA). The logistic regression equation thus takes on the following format:

$$y = \frac{e^{b_0+b_1x_1+\ldots+b_nx_n}}{1+\ e^{(b_0+b_1x_1+\ldots+b_nx_n)}} \tag{2}$$

Logistic Regression Mapping Function

Where y is the predicted output, b_0 is the intercept, b_1 – b_n are the coefficients for the input feature values (x_1 – x_n). In our use case, x corresponds to a row of matrix $\boldsymbol{X}$, which contains the values of each of the 49 computed signal features for an individual cycle from one accelerometer. Once trained, each column of the input matrix $\boldsymbol{X}$ (i.e., each feature) has an associated coefficient learned through training (b_1 – b_n). The vector of b_1 – b_n is stored in the coefficient vector (β). β is found using a maximum-likelihood estimation (MLE), specifically the quasi-Newton method, that minimizes the error of the predicted probabilities (20, 34, 35).

The predicted output (y in **Equation 2**) is the probability that a given input belongs to the default class, selected in our case as JIA. These probability predictions are transformed into binary values (0 or 1) in order to create the final probability-based predicted label for each feature using the threshold in **Equation 3**.

$$\begin{aligned} &If\ mean(p(\mathbf{x})) \leq 0.5,\ y\ =\ Healthy \\ &If\ mean(p(\mathbf{x})) > 0.5,\ y\ =\ JIA \end{aligned} \tag{3}$$

Threshold for Healthy Control vs. JIA Classification

As mentioned, each cycle of flexing/extension (each row of $\boldsymbol{X}$) is classified on as a 0 or 1, with 0 representing a healthy, unaffected knee and 1 representing a knee with active JIA. To calculate this score, we train a logistic regression classification model. All rows for a subject can be removed from $\boldsymbol{X}$ to leave behind $\boldsymbol{X'}$ and $\mathbf{X}_{\text{subject}}$. Each row in these matrices corresponded to one accelerometer's output for one cycle of movement. Each subject had two accelerometers on each leg and was asked to perform 10 cycles of flexion/extension. The average number of rows in these submatrices was 36 ± 3 rows, with the average number of rows per accelerometer being 8 ± 1 rows. If the majority of the predicted labels for an individual row were classified as 0, the cycle was labeled as healthy. If the majority of the predicted labels were predicted as 1, the cycle was considered to be JIA. In this way, the median of the predicted labels of each row determines the classification of that cycle of that microphone. The median of the rows in any given $\boldsymbol{X}_{\text{subject}}$ is taken to be the subject classification. If the majority of the rows was predicted to be 1's, the subject was labeled as having JIA. Inversely, if the majority of rows was predicted as 0's, the subject was labeled as healthy.

The logistic regression model's performance was assessed using LOSO-CV (36). In each fold of this validation, the logistic regression classifier was trained using the data in $\boldsymbol{X'}$ with one subject omitted - $\boldsymbol{X}_{\text{subject}}$. The trained model then classified the signal of the excluded subject's knee JAEs. During LOSO-CV, the matrix $\boldsymbol{X'}$ was standardized after the removal of $\boldsymbol{X}_{\text{subject}}$. The mean and standard deviation of $\boldsymbol{X'}$ were then subtracted and divided, respectively, from the columns in $\boldsymbol{X}_{\text{subject}}$. By doing this, the calculated features for $\boldsymbol{X}_{\text{subject}}$ were not prematurely included in the standardization of $\boldsymbol{X}$. The model estimates the probability of JIA for each row (cycle) in $\boldsymbol{X}_{\text{subject}}$. These probabilities were stored in the vector, $\boldsymbol{p}_{\boldsymbol{predicted}}$. The overall subject's audio scores were calculated by averaging the contents of $\boldsymbol{p}_{\boldsymbol{predicted}}$ (**Figure 1D**). The 0.5 threshold was applied to this average probability to assign the predicted label of healthy (0) or JIA (1). The cross-validation was completed by calculating knee audio scores for all 43 subjects, excluding one subject per fold. The follow-up recordings were not included in this model accuracy calculation because the treating physician stated they were along a spectrum of convalescence, and thus their ground truth label was unknown. The generalizability of the model is assessed by calculating the accuracy of our algorithm in labeling each cycle, as well as in labeling each.

The average probabilities were used not only for predicting labels, but also as an indicator of knee health. In this way, as the average probability of a subject trends toward 0, the signal more greatly resembles a healthy knee. For subjects with JIA that have follow-up recordings, this process was repeated to calculate the change in the probability of JIA between the first recording and second. Importantly, the follow-up recordings are never used as part of the training set, since at the time of recording those subjects the ground-truth of their disease status is unknown.

Feature Importance Ranking

The relative weighting of each of the features in the model needs to be explored to understand which features most relate to differentiating JAEs from patients with JIA compared to healthy controls. To quantify the importance of each feature, the standardized data from every subject with JIA (excluding the follow-up data due to it lacking a ground truth classification) is used to train the classifier. The resulting model is used to

generate relative feature importance scores. In this case, no testing set is required to quantify feature importance since we are not assessing the generalizability of the model. In the case of logistic regression, the model computes a coefficient for each input feature that describes the k-dimensional hyperplane that best separates the two input classes. When the input matrix $\boldsymbol{X}$ is standardized to zero mean and unity variance, the absolute value of each of the coefficients output from the model can be directly compared to assess relative importance to the model. In this way, a coefficient with a large absolute value has a larger effect on the model than one with a smaller absolute value. All 49 features are ranked in order from most to least important as seen in **Figure 4A**.

Effect of Number of Features and Cycles of Movement on Model Performance

After ranking the 49 features, we further assessed the impact on the accuracy of the model's predictive capabilities by training the model on one to forty-nine features in order of their relative importance. We first trained a model on only the most important feature, and assessed the accuracy of the model as detailed above using LOSO-CV. Next, we iteratively added each new feature in order of descending relative importance to observe how that accuracy improved with the addition of each new feature. We simultaneously assessed the importance of the number of flexion/extension cycles by testing each iteration of the model on a subset of all of the cycles. For example, we first trained the model on the most important feature, and tested the model using one cycle from the subject left out, next two cycles, then three cycles, all the way up to the full number of recorded cycles. In doing so, we calculated how the model responded for each feature input and for each additional cycle of movement input. Of note, when choosing the subsets of cycles to test we iteratively tested up to 1,000 unique permutations on any given sized subset of cycles and the average of those cycles was reported. A heatmap of these results was generated and can be seen in **Figure 4B**.

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 human participants were reviewed and approved by Emory University School of Medicine Institutional Review Board Georgia Institute of Technology Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DW served as the project lead and was involved in every part of its design, execution, analysis, and reporting. JZ provided his machine learning expertise and helped design the JAE algorithm. SG helped organize the data and performed the IMU assessment. TG and LP were lead clinical coordinators that helped devise an appropriate protocol for consenting, assenting, and recording JAEs in a clinical setting. OI served as the principal investigator for the project, and was integral in the funding, managing, planning, and execution of all aspects. SP closely collaborated with OI and sponsored this project: specifically, SP provided access to patients, clinic space, and medical expertise on the current state of JIA management. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Science Foundation, CAREER Grant No. 1749677.

ACKNOWLEDGMENTS

The authors thank the many clinical coordinators and families that contributed their time to the collection of this dataset.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2020.571839/full#supplementary-material

REFERENCES

1. Ravelli A, Martini A. Juvenile idiopathic arthritis. *Lancet.* (2007) 369:767–78. doi: 10.1016/S0140-6736(07)60363-8
2. Murray K, Thompson SD, Glass DN. Pathogenesis of juvenile chronic arthritis: genetic and environmental factors. *Arch Dis Child.* (1997) 77:530–4. doi: 10.1136/adc.77.6.530
3. Prahalad S, Glass DN. A comprehensive review of the genetics of juvenile idiopathic arthritis. *Pediatr Rheumatol.* (2008) 6:11. doi: 10.1186/1546-0096-6-11
4. Wallace CA. Current management of juvenile idiopathic arthritis. *Best Pract Res Clin Rheumatol.* (2006) 20:279–300. doi: 10.1016/j.berh.2005.11.008
5. Woo P. Systemic juvenile idiopathic arthritis: diagnosis, management, and outcome. *Nat Clin Pract Rheumatol.* (2006) 2:28–34. doi: 10.1038/ncprheum0084
6. Prince FHM, Otten MH, van Suijlekom-Smit LWA. Diagnosis and management of juvenile idiopathic arthritis. *BMJ.* (2010) 341:c6434. doi: 10.1136/bmj.c6434
7. John M. Eisenberg Center for Clinical Decisions and Communications Science JMEC for CD and C. *DMARDs for Juvenile Idiopathic Arthritis: A Review of the Research for Parents and Caregivers*. Rockville, MD: Agency for Healthcare Research and Quality (US) (2005).
8. Gartlehner G, Hansen RA, Jonas BL, Thieda P, Lohr KN. Biologics for the treatment of juvenile idiopathic arthritis: a systematic review and critical analysis of the evidence. *Clin Rheumatol.* (2007) 27:67–76. doi: 10.1007/s10067-007-0654-6
9. Abramowicz S, Kim S, Prahalad S, Chouinard AF, Kaban LB. Juvenile arthritis: current concepts in terminology, etiopathogenesis, diagnosis, and management. *Int J Oral Maxillofac Surg.* (2016) 45:801–12. doi: 10.1016/j.ijom.2016.03.013

10. Deal CL, Hooker R, Harrington T, Birnbaum N, Hogan P, Bouchery E, et al. The United States rheumatology workforce: supply and demand, 2005–2025. *Arthritis Rheum.* (2007) 56:722–9. doi: 10.1002/art.22437
11. Arthritis Foundation. *Increase Access to Pediatric Rheumatologists*. Available online at: https://www.arthritis.org/advocate/our-policy-priorities/access-to-care/increase-access-to-pediatric-rheumatologists/ (accessed September 12, 2019).
12. Semiz B, Hersek S, Whittingslow DC, Ponder LA, Prahalad S, Inan OT. Using knee acoustical emissions for sensing joint health in patients with juvenile idiopathic arthritis: a pilot study. *IEEE Sens J.* (2018) 18:9128–36. doi: 10.1109/JSEN.2018.2869990
13. Chu ML, Gradisar IA, Railey MR, Bowling GF. Detection of knee joint diseases using acoustical pattern recognition technique. *J Biomech.* (1976) 9:111–4. doi: 10.1016/0021-9290(76)90149-4
14. Frank CB, Rangayyan RM, Bell GD. Analysis of knee joint sound signals for non-invasive diagnosis of cartilage pathology. *IEEE Eng Med Biol Mag.* (1990) 9:65–8. doi: 10.1109/51.62910
15. Shark LK, Chen H, Goodacre J. Discovering differences in acoustic emission between healthy and osteoarthritic knees using a four-phase model of sit-stand-sit movements. *Open Med Inform J.* (2010) 4:116–25. doi: 10.2174/1876396001004010116
16. Winter DA. *Biomechanics and Motor Control of Human Movement*. Waterloo, ON: Wiley (2009).
17. Ording Muller LS, Humphries P, Rosendahl K. The joints in juvenile idiopathic arthritis. *Insights Imaging.* (2015) 6:275–84. doi: 10.1007/s13244-015-0406-0
18. Palazzo C, Ravaud JF, Papelard A, Ravaud P, Poiraudeau S. The burden of musculoskeletal conditions. *PLoS ONE.* (2014) 9:e90633. doi: 10.1371/journal.pone.0090633
19. Whittingslow DC, Jeong HK, Ganti VG, Kirkpatrick NJ, Kogler GF, Inan OT. Acoustic emissions as a non-invasive biomarker of the structural health of the knee. *Ann Biomed Eng.* (2019) 48:1—11. doi: 10.1007/s10439-019-02333-x
20. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform.* (2002) 35:352–9. doi: 10.1016/S1532-0464(03)00034-0
21. Peeters G. *A Large Set of Audio Features for Sound Description, vol. 54* (2004).
22. Ravelli A. Toward an understanding of the long-term outcome of juvenile idiopathic arthritis. *Clin Exp Rheumatol.* (2004) 22:271—5.
23. Jeong HK, Whittingslow D, Inan OT. b-Value: a potential biomarker for assessing knee-joint health using acoustical emission sensing. *IEEE Sensors Lett.* (2018) 2:1–4. doi: 10.1109/LSENS.2018.2871981
24. Pohar M, Blas M, Turk S. Comparison of Logistic regression and linear discriminant analysis: a simulation study. *Metodoloski Zvezki*. (2004) 1:143–61.
25. Jeong HK, Pouyan MB, Whittingslow DC, Ganti V, Inan OT. Quantifying the effects of increasing mechanical stress on knee acoustical emissions using unsupervised graph mining. *IEEE Trans Neural Syst Rehabil Eng.* (2018) 26:594–601. doi: 10.1109/TNSRE.2018.2800702
26. Wu Y, Krishnan S, Rangayyan RM. Computer-aided diagnosis of knee-joint disorders via vibroarthrographic signal analysis: a review. *Crit Rev Biomed Eng.* (2010) 38:201–24. doi: 10.1615/CritRevBiomedEng.v38.i2.60
27. Saurenmann RK, Levin AV, Feldman BM, Rose JB, Laxer RM, Schneider R, et al. Prevalence, risk factors, and outcome of uveitis in juvenile idiopathic arthritis: a long-term followup study. *Arthritis Rheum.* (2007) 56:647–57. doi: 10.1002/art.22381
28. Saurenmann RK, Rose JB, Tyrrell P, Feldman BM, Laxer RM, Schneider R, et al. Epidemiology of juvenile idiopathic arthritis in a multiethnic cohort: ethnicity as a risk factor. *Arthritis Rheum.* (2007) 56:1974–84. doi: 10.1002/art.22709
29. Chu ML, Gradisar IA, Railey MR, Bowling GF. An electro-acoustical technique for the detection of knee joint noise. *Med Res Eng.* (1976) 12:18–20.
30. Hersek S, Pouyan MB, Teague CN, Sawka MN, Millard-Stafford ML, Kogler GF, et al. Acoustical emission analysis by unsupervised graph mining: a novel biomarker of knee health status. *IEEE Trans Biomed Eng.* (2018) 65:1291–300. doi: 10.1109/TBME.2017.2743562
31. Giannakopoulos T, Pikrakis A. *Introduction to Audio Analysis : A MATLAB Approach* Waltham, MD: Academic Press – Elsevier (n.d).
32. Ricklefs RE. A graphical method of fitting equations to growth curves. *Ecology*. (1967) 48:978–83. doi: 10.2307/1934545
33. Gabriel JP, Saucy F, Bersier LF. Paradoxes in the logistic equation? *Ecol Modell.* (2005) 185:147–51. doi: 10.1016/j.ecolmodel.2004.10.009
34. Broyden CG. Quasi-Newton methods and their application to function minimisation. *Math Comput.* (1967) 21:368. doi: 10.1090/S0025-5718-1967-0224273-2
35. Jansche M. Maximum expected F-measure training of logistic regression models. In: In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*. Morristown, NJ: Association for Computational Linguistics (2005). p. 692–9.
36. Hersek S, Toreyin H, Teague CN, Millard-Stafford ML, Jeong HK, Bavare MM, et al. Wearable vector electrical bioimpedance system to assess knee joint health. *IEEE Trans Biomed Eng.* (2017) 64:2353–60. doi: 10.1109/TBME.2016.2641958

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 25 November 2020
doi: 10.3389/fdgth.2020.587452

Queensland Telepaediatric Service: A Review of the First 15 Years of Service

Anthony C. Smith[1,2,3]*, *Nigel R. Armfield*[1,2,4], *Mark G. Coulthard*[5,6], *Michael L. Williams*[7] and *Liam J. Caffery*[1,2]

[1] Centre for Online Health, The University of Queensland, Brisbane, QLD, Australia, [2] Centre for Health Services Research, The University of Queensland, Brisbane, QLD, Australia, [3] Centre for Innovative Medical Technology, University of Southern Denmark, Odense, Denmark, [4] Recover Injury Research Centre, The University of Queensland, Brisbane, QLD, Australia, [5] Pediatric Intensive Care Unit, Queensland Children's Hospital, Brisbane, QLD, Australia, [6] Mayne Academy of Pediatrics, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia, [7] Mackay Base Hospital, Queensland Health, Brisbane, QLD, Australia

In November 2000, the Queensland Telepaediatric Service (QTS) was established in Brisbane, Australia, to support the delivery of telehealth services to patients and clinicians in regional and remote locations. The QTS was built on a centralized coordination model, where telehealth services could be effectively managed by a dedicated telehealth coordinator. In doing so, telehealth referral and consultation processes were efficient and clinicians felt better supported as they adjusted to new processes for engaging with patients. We have conducted a retrospective review of activity associated with the QTS and summarized key activities which have arisen from this extensive program of work. Telehealth service records and associated publications were used to describe the evolution of the QTS over a 15-year period. From November 2000 to March 2016, 23,054 telehealth consultations were delivered for 37 pediatric clinical specialties. The most common service areas included child and youth mental health, neurology, burns care, surgery, and ear nose and throat services. A range of different telehealth service models were developed to align with different clinical service needs and location of services. Whilst most work involved video consultation between hospitals, some services involved the delivery of telehealth services into the home, schools or community health centres. Despite its longevity, the QTS was not immune to the usual challenges associated with telehealth implementation, service redesign and sustainability. Experience reported from the QTS will be useful for other health services seeking to develop comprehensive telehealth services in a rapidly changing healthcare environment.

Keywords: telehealth, telemedicine, telepaediatrics, digital health, indigenous, specialist health care, models of care, regional and remote health services

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Venkat Vangaveti, James Cook University, Australia
Anthony Herbert, Queensland University of Technology, Australia

***Correspondence:**
Anthony C. Smith
asmith@uq.edu.au

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *26 July 2020*
Accepted: *12 October 2020*
Published: *25 November 2020*

Citation:
Smith AC, Armfield NR, Coulthard MG, Williams ML and Caffery LJ (2020) Queensland Telepaediatric Service: A Review of the First 15 Years of Service. Front. Digit. Health 2:587452. doi: 10.3389/fdgth.2020.587452

INTRODUCTION

Conventional models of health care in Australia require patients to travel (often great distances) to receive specialist care. Occasionally specialist teams travel to remote communities to deliver health care services; but these tend to occur on an intermittent basis. For logistical reasons, some patients do not receive the care they require because of the difficulties of having to leave their community for an appointment and/or treatment. Telehealth can be used to improve access to health services

for people living in distant locations; this is important in Australia where the majority of specialist health services are based in metropolitan areas, and the distances between these hospitals and small rural hospitals may be considerable.

Over the last two decades, the use of telehealth to deliver pediatric telehealth services (telepaediatrics) has been reported by many countries (1, 2). Telepaediatric service models have included multidisciplinary services operated from a centralized coordination centre; discipline-specific telehealth services for children and young people; and services in different settings (such as hospitals, community health settings, schools and in the home). The idea of providing telehealth services particularly for children and their families makes sense because of the centralization of pediatric specialist services, imposition of travel away from home, and the requirement for a child to be accompanied by a parent or caregiver for travel to and from their specialist appointment.

Despite the clear benefits of telehealth, a long-term effort was required to address the challenges of telehealth implementation and uptake in the Queensland public health service. In November 2000, a pediatric telehealth service model was established at the Royal Children's Hospital (RCH) in Queensland, Australia (3). The Queensland Telepaediatric Service (QTS) offered a convenient referral process (single point of contact) for telehealth referrals and coordination of telehealth consultations (4). The majority of telehealth consultations involved a videoconference appointment between the specialist hospital in Brisbane and a referring hospital. Other communication methods included correspondence by email or telephone. The telehealth service used the videoconferencing network operated by the state health department—comprising both hardware and software systems. In some cases, customized telehealth systems were deployed to improve child-friendliness, or where standard systems did not meet the clinical requirements, see **Figure 1** (5).

The University of Queensland's Centre for Online Health (COH) was responsible for establishing and operating the QTS in partnership with the Queensland health department. Operational responsibilities were funded by a service level agreement; and an integrated research program was funded by community and corporate organizations. The aim of this review is to summarize patterns of service activity, outline specific service models, and describe the key enablers and challenges associated with the service.

METHODS

This study presents a retrospective review of QTS activity reported over a 15-year period from November 2000 to March 2016. Service activity was obtained from an operational database, which was owned and maintained by the COH. This database contained information about each consultation including specialty, duration, location and modality. This review also summaries published studies undertaken during the course of this program. All published studies reported in this review received ethical approval from the appropriate committees. Further permission and exemption from ethical review was obtained from Children's Health Queensland Hospital and Health Service Human Research Ethics Committee to publish overall service activity according to service records managed by the COH (dated 13 June 2019).

RESULTS

Service Activity

From November 2000—April 2015, a total of 23,054 telehealth consultations were coordinated through the QTS. The majority of these (95%) involved consultations by videoconference, whereas the remaining involved email (3%) or telephone consultations (<1%). A total of 37 clinical specialties were actively involved in the QTS, delivering services to 110 sites throughout Queensland and Northern New South Wales, see **Figure 2**.

The most common specialties were child psychiatry (35%), neurology (10%), burns (9%), surgery (6%), and Ear, Nose, and Throat (6%), see **Figure 3**. "Other" services included metabolic, cardiology, neurosurgery, palliative care, ophthalmology, immunology and allergy, neonatology, plastic surgery, dietetics, rehabilitation, speech pathology, pain management, audiology, occupational therapy, physiotherapy, infectious diseases, child development, and social work.

The volume of telehealth activity gradually increased during the first 10 years (2000–2010), with the introduction of new specialties and expansion of services within certain clinical disciplines, see **Figure 4**. In 2006 (A), our mobile videoconferencing systems were used to provide pediatric support to regional hospitals (child-friendly robot ward rounds); and in 2008 (B), the mobile ear, nose and throat (ENT) surveillance service for Indigenous children was established in Cherbourg, resulting in additional ENT consultations at the RCH. From 2010 onwards, activity levels remained static or fell slightly. This mainly coincided with a significant reduction in the QTS operational budget (C); staged closure of the RCH (D); and the transfer of the QTS (E) over to the new children's hospital in Brisbane.

Referrals for telehealth consultations were made from over 270 health services, mainly hospitals throughout Queensland and northern New South Wales. The top five referring sites were Mackay (20%), Atherton (14%), Hervey Bay (7%), Mt Isa (5%), and Innisfail (5%). Almost all referrals originated from a regional hospital—or from the specialist hospital (provider site) for patient follow-up.

The QTS was primarily a clinical service; 97% of its use involved providing advice about a patient, reviewing a case, initial assessment before transfer to the specialist hospital or handover of a patient before return to regional hospital. The remaining activity (<3%) concerned the delivery of education or administrative services.

Service Models

The QTS was flexible and responsive to clinical needs; a variety of different service models were developed amongst the specialist areas, reflecting the needs of the patient and the purpose of the consultation. Some were developed for general outpatient appointments with specialists, emergency

FIGURE 1 | Wireless (robot) videoconference system used for bedside consultations in regional pediatric wards.

advice for the assessment of infants with a cardiac condition, follow-up of patients receiving specialist burns care, case conferencing with regional health care teams involved in the care of children and families with mental health conditions, handover of patients to regional hospitals and home care, and community-based assessments of children with chronic ear disease. Other applications were primarily developed for education and training purposes.

General Outpatients

The most common application within the telehealth service was the delivery of outpatient appointments for children and families who would normally travel to Brisbane. Most clinical specialties were actively engaged and provided telehealth clinics on either a weekly or a monthly basis depending on demand. Common examples included clinics for diabetes, neurology, orthopedics, nephrology, rheumatology, and pediatric surgery (6–11). Often these clinics would be run in parallel with the in-person clinics in Brisbane, and the specialist time was allocated as required to the telehealth service. For certain specialties, a telehealth clinic list was established prior to the session, and connections involved multiple patients at the same (referring site) or multiple patients in a range of different sites. Telehealth clinics all required careful coordination to ensure site preparation and collection of the necessary clinic information in advance of the consultation.

Ad-hoc and Urgent Consultations

The assessment of newborn children with suspected cardiac defects was one of the services offered by the QTS. A pediatric cardiologist was able to assess infants remotely, by instructing the remote sonographer and viewing the echocardiogram in real-time (12). This meant a timely diagnosis and management plan could be discussed with the remote pediatrician caring for the child, and an informed decision could be made whether to transfer the infant to the specialist hospital or not. In the majority of cases, transfer of infants was then avoided and the infants continued to be managed locally with remote specialist support/advice as required. This service also provided the sonographer conducting the scan with valuable training experience whilst working with the cardiologist (13).

Post-acute Burns Care

In Queensland, specialist burns care is provided by one hospital. Referral guidelines for children with a serious burn injury indicate that referral to the specialist is necessary. Once care is received, follow-up care may last for many months or years. Prior to the use of telehealth, some outpatient appointments in Brisbane lasted for only minutes, despite some travel to the hospital taking many hours. The use of telehealth for outpatient burns care has revolutionized the support for children throughout the state (14). The burns team regularly provide videoconference appointments to all throughout Queensland and northern New South Wales. Appointments often involve occupational therapists (OTs) and nurses in regional hospitals, and the specialist burns staff in Brisbane (a medical consultant, OT and Nurse). In addition to the general follow-up appointments, telehealth has also been very useful for interim advice for a burn injury—to assist with immediate treatment at the remote hospital and planning for the transfer of the patient (15).

Mental Health Services

The delivery of telehealth by the e-child and youth mental health services (e-CYMHS) demonstrates a very effective model,

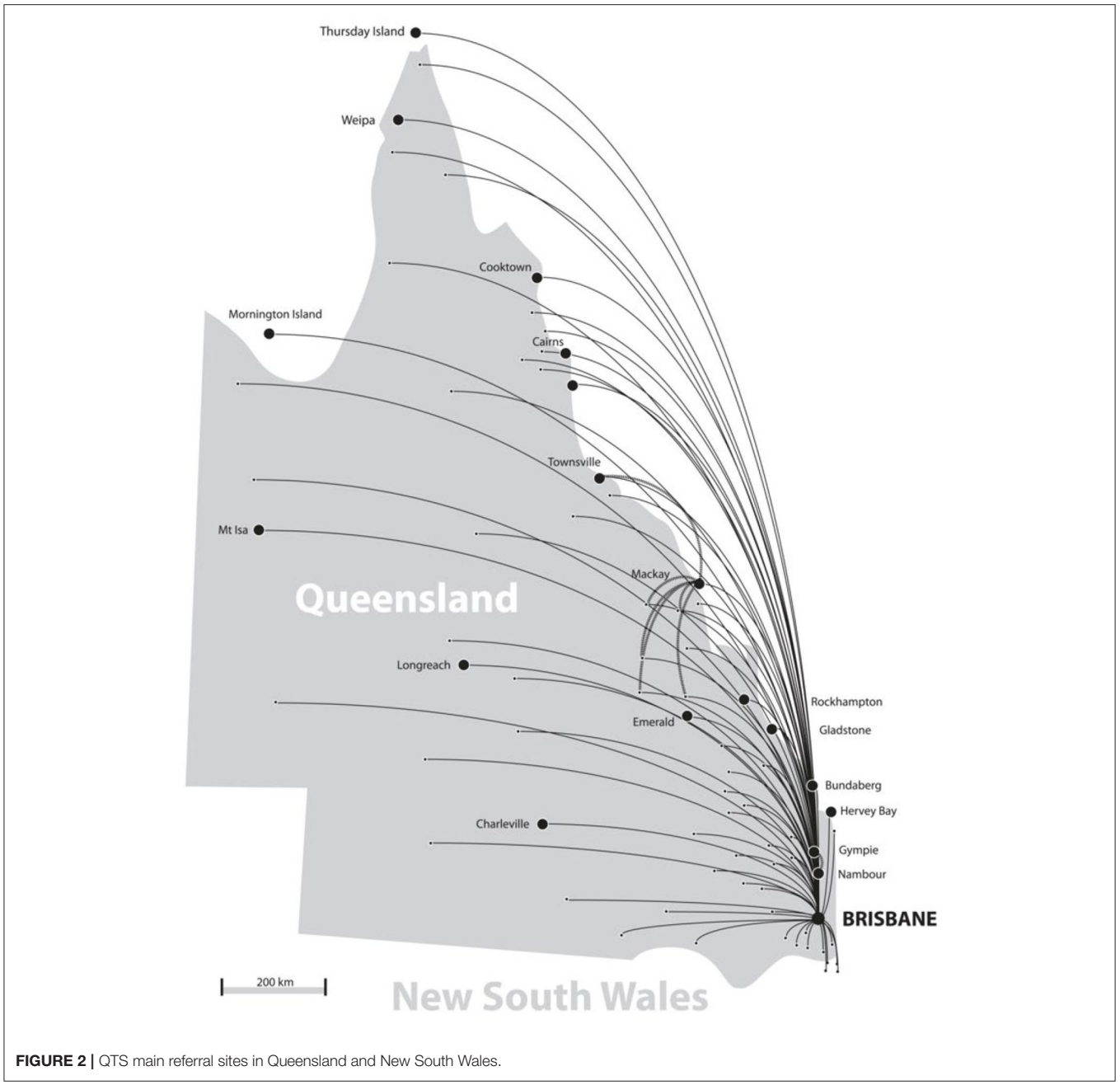

FIGURE 2 | QTS main referral sites in Queensland and New South Wales.

combining conventional outreach services (where the specialist team travel to the regional towns) with telehealth support (16, 17). Most telehealth clinics involve case conferencing, where a series of cases are presented to the specialist team (psychiatrist and other mental health clinicians) via videoconference. In some cases, the patient's family would also participate in the session. As with the overall QTS, the success of the e-CYMHS was attributed to the role of the dedicated e-CYMHS telehealth coordinator. The telehealth component of the service was a very cost-effective way of keeping in contact with regional sites (see section on cost savings) (18).

Discharge Planning and Home Care

For patients receiving specialist care in Brisbane, telehealth was used to support the process of back-transfer to regional hospitals or to home care. In these cases, specialist teams could hand over important information about the cases via videoconference and ensure that regional staff and families were prepared for and understood the clinical care requirements. Anecdotally, families reported that they felt at ease knowing that the regional staff were familiar with the treatment and follow-up care. This was fairly common practice for children referred from the oncology and palliative care unit (19, 20). In the case of home care, often

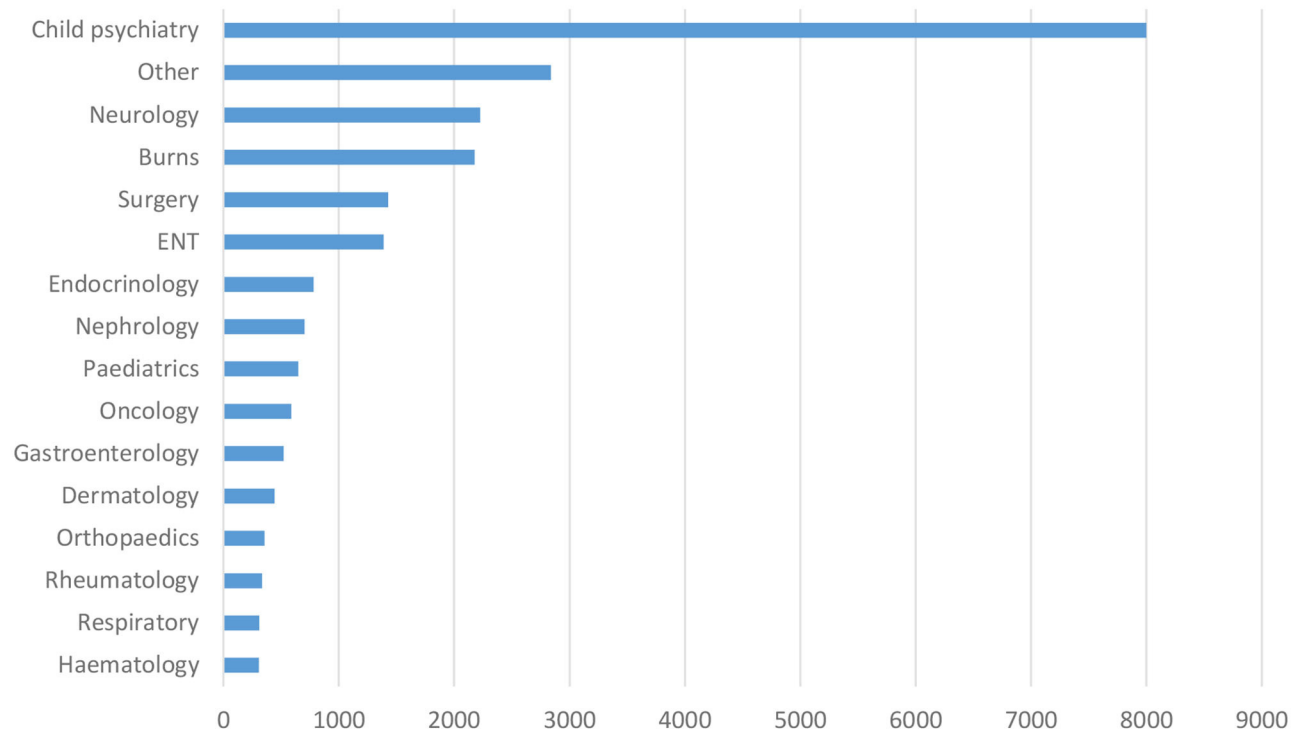

FIGURE 3 | Number of QTS consultations per specialty, from November 2000 to March 2016.

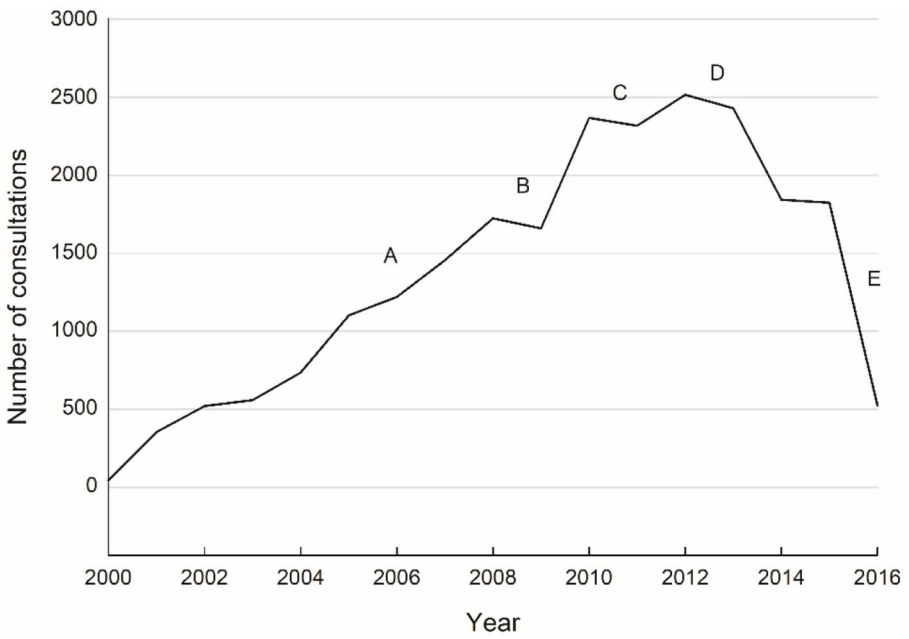

FIGURE 4 | All QTS consultations over time (November 2000 to March 2016). Milestones: **(A)** mobile videoconferencing systems deployed for regional paediatric ward rounds; **(B)** mobile ENT surveillance service established; **(C)** QTS operational budget reduced; **(D)** staged closure of the RCH; and **(E)** transfer of the QTS to the new Queensland Children's Hospital in Brisbane.

the home nursing services and local general practitioners were engaged in the telehealth service.

Community-Based Health Surveillance

Through consultation with the Cherbourg community health service (∼260 km from Brisbane) and specialists in Brisbane, we developed a surveillance program for Aboriginal and Torres Strait Islander children at risk of ear disease, to ensure early detection and referral for treatment. We developed a mobile telehealth-enabled ear screening service, see **Figure 5**, which was operated by an experienced local Aboriginal Health Worker (AHW) (21). The AHW used the mobile service to visit schools and routinely assess Indigenous children. The AHW assessments included pure-tone audiometry, tympanometry, and digital otoscopy. In cases where children failed a screening test or if the AHW had any concerns, the AHW assessments were shared asynchronously via a secure online database, and reviewed by an ENT specialist. Assessment and treatment planning would then be done by the specialist and/or referred to the local medical service. This screening service has resulted in improvements in overall screening rates and the emergence of a model of care, which is community led and culturally appropriate (22, 23).

Education and Training

Whilst the majority of services delivered through the QTS were of a clinical nature, the use of videoconferencing was also important for education and training purposes. All clinical consultations had an intrinsic educational benefit because of the interaction that occurred between specialists and clinicians at the referring sites. Anecdotally, clinicians appreciated the service because of the learning opportunities it offered. Specific services were also developed to support the training requirements of regional staff responsible for children with special care needs (such as burns care and child development). When patients were being directed back to primary and secondary centres, it was important that clinicians were supported with clinical education (24, 25). Students undertaking their clinical training in rural and remote hospitals were also supported by the QTS, with access to interactive lectures by videoconference—allowing participation irrespective of location (26).

Cost Savings

The majority of savings were associated with the reduced need for patient travel. Economic evaluations using cost-minimization analysis methods demonstrated the level of activity required to reach a threshold, whereby the costs of providing one service were the same as the other. The child and youth mental health service, which was responsible for almost one-third of all QTS activity, demonstrated that at the level of activity achieved in their service, it was less expensive to provide telehealth services than doing outreach (where the specialist team traveled to the regional town) or arranging for the patient and family to travel (18, 27). Similar studies showed potential savings to the health service for ENT services (28).

Key Enablers

A key factor in the success of this program was the centralization of support made available by the QTS, which made the referral, consultation and documentation process convenient for clinicians. Another key factor was the integration of telehealth services on a business as usual basis, which was reflected in clinic schedules, service delivery planning and staffing allocations for each specialty. Running these clinical activities alongside a robust research program gave clinical teams the opportunity to contribute to the evaluation process and also to the planning of innovative services within the department. Clinician availability and support for both the near (provider) and far (receiver) end was very important—as was the need to train clinicians in certain skills relevant to telehealth consultation processes. The telehealth process also required new referral processes—and once these were made clear, the coordination of appointments and clinics became more straightforward.

Access to high quality telehealth facilities in a central and easily accessible location was important. The COH provided dedicated telehealth studios, which were used for most clinics. Over time, with improvements in software-based videoconferencing systems, some clinical groups were able to conduct their own telehealth work within their own department. This still required an appropriate place, which was private, and had good lighting and suitable acoustics.

World Firsts

Academically, the COH published over 75 journal articles relating to "telepaediatrics" during the 15-year period; and pioneered a number of "world firsts" which are improving access to health and support services for regional families. These included the establishment of QTS—the first fully serviced multidisciplinary pediatric telehealth service (2, 3); the first child-friendly mobile telehealth service (robots) (5, 24, 29); the first use of telehealth for the delivery of clown doctor outreach services (30); and the first telehealth-supported Indigenous ear screening service with online links to pediatric specialists (21, 22).

Challenges

Funding to cover the cost of telehealth is a commonly reported challenge, and one faced by the QTS since establishment. Initially, QTS telehealth services were not funded, so unless the telehealth service was purely substitution of face-to-face clinic appointments, then clinicians were providing services without direct funding. In 2011–2012, new funding opportunities emerged when the Commonwealth Government introduced funding for specialist video consultations under the Medical Benefits Schedule (MBS) (31). Around the same time, the Queensland Government Statewide Telehealth Unit introduced incentive funding for telehealth, to promote the uptake of telehealth. This incentive funding was in addition to activity-based funding which includes all activity (telehealth and in-person consultations). In small rural hospitals, activity-based funding is typically not viable due to relatively lower activity, and therefore block-funding arrangements are supported by the health department. In 2020, new temporary funding was introduced by the Australian Government in response to the

FIGURE 5 | (A) Mobile screening van used for surveillance of Aboriginal and Torres Strait Islander children at risk of ear disease and **(B)** child being screened by an experienced Aboriginal Health Worker.

2019 coronavirus pandemic (COVID-19). Collectively, these funding developments have resulted in substantial growth in telehealth activity across Queensland and throughout Australia (32, 33).

Staff availability was also a challenge because telehealth sessions not only relied on the availability of the specialist, but also the availability of the referring clinicians and support staff at the regional hospital. We addressed this challenge by setting up clinics in advance, so that regular clinical days and times were available—either on a weekly, fortnightly, monthly or quarterly basis, depending on demand. The delivery of telehealth also changed from an *ad-hoc* arrangement to an appreciation that telehealth was integrated and routine. Like any telehealth operation, we did experience some staff resistance, but this was mainly related to the lack of clear processes, time constraints, and telehealth awareness. In this context, clinician acceptance and willingness to practice were important factors in the uptake of telehealth (34). Resistance transformed into interest over time as clinicians gained experience, and processes were put into place to ensure appropriateness of telehealth referrals and case preparation (case history and other relevant documentation) (35).

DISCUSSION

The QTS represents an extensive program of work conducted over a 15-year period. Examples of pediatric telehealth services have emerged as a result of different clinical requirements. The expansion of telehealth was sustained over an extensive period of time, and for a small number of mature services, we observed a willingness to conduct telehealth consultations outside of the telehealth centre and in the clinical departments. This worked particularly well when there was administrative assistance available in the department to help prepare cases, conduct test calls, send appointment details to families and help with the documentation (hospital records, investigative tests, referral notes etc.). The work done in Queensland also highlighted the importance of the role of a telehealth coordinator. This was a key requirement for the facilitation of services and an intended strategy to ensure that the referral and telehealth consultation processes were managed efficiently and without unreasonable burden on the clinician. Originally considered a superfluous resource by some health managers in Queensland, telehealth coordinator positions are now fully supported throughout the state—and recurrently funded by the health department, on a business as usual basis.

Over time, it was encouraging to see the number of specialties engaged in the QTS. It was clear that telehealth was and has continued to be used as a routine method of consultation for medical, nursing and allied health staff in the health service. The development of the service was also inspired by a variety of COH-led research projects, which helped to generate new ideas amongst clinicians when caring for children and families in remote locations. Funding for these projects was mainly derived from competitive research grants and philanthropic funding. Combining research and service delivery was a useful process because it meant that clinician engagement was strong and ideas were generated in direct response to clinical needs. The duration of the service development work also meant that information could be collected to demonstrate trends in activity and opportunities for service growth.

The work highlighted in Queensland is one of the most prominent examples of telepaediatrics reported worldwide, operating over a significant period and demonstrating a large volume of activity across many different specialties. Other successful examples exist in the USA and Canada where telehealth services have been established for emergency and intensive care support, hospital outpatients, primary care and home support (36–41). Work in California also demonstrated cost savings and significant environmental benefits due to reduced travel requirements for patients—hence another reason for doing telehealth (42).

Key Lessons Learned

1. The establishment of a successful telehealth service requires time, patience and close engagement with clinicians and health service managers.
2. Effective telehealth services require dedicated administrative support services (telehealth coordination) and strong clinical leadership.
3. Integrating telehealth into existing hospital systems (such as referral and triage management, scheduling and billing processes) is important for services to become routinely adopted.
4. Clinician involvement in the planning and delivery of new telehealth-supported models of care ensures that services fundamentally address clinical requirements and patient needs.
5. The broad nature of services established through the QTS demonstrates the value of telehealth for a diverse range of clinical specialities and also highlights the importance of different service models for different clinical areas (not a one-model-fits-all approach).
6. Most potential savings attributed to the QTS were related to reduced patient travel.
7. The unique partnership between the COH (university) and the health department (service) resulted in 15 years of pioneering work and the development of a rich evidence base for telepaediatrics.

CONCLUSION

The QTS work has laid the foundations for the provision of pediatric telehealth services in Queensland, and many of the service models have been replicated in other places throughout Australia. During the operational period reported in this review, the partnership between the service provider and the university was a unique opportunity to leverage research funding and to drive innovation within the service. We encourage future reviews of the QTS to monitor progress and to demonstrate the benefits for children and families living in remote locations. It is highly likely that new telehealth-enabled models of care will continue to evolve in response to the many challenges faced by the health department, new funding arrangements, advances in communications technology and the expectations of consumers due to increased experience and raised awareness of telehealth.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article is available upon request, subject to ethics approval requirements.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AS was responsible for establishing the QTS and for leading the establishment of most of the services described in the review, obtaining clearance from the appropriate ethics committee, leading the analysis of activity, and drafting the manuscript with input from all authors. AS, NA, and LC conceived and designed the review of the QTS and were responsible for data collection. All authors were involved in reviewing the manuscript and critically appraising the content and were responsible for final approval of the manuscript before submission for publication.

FUNDING

Operational telehealth services were funded by service level agreements with the public health service (Queensland Government) and research activities were funded by external government, industry, and community agencies.

ACKNOWLEDGMENTS

Thanks to the clinicians at the Queensland Children's Hospital (formerly the Royal Children's Hospital) in Brisbane for their role in providing telepaediatric services in Queensland. We also thank the following staff for their contribution to the planning and operational aspects of the telepaediatric service, and QTS data collection and analysis: Prof. Alan Isles, Prof. Richard Wootton, Dr. Bob McCrossin, Dr. Jasper Van der Westhuyzen, Dr. Natalie Bradford, Dr. Cathy Xu, Dr. Mark Bensink, Cecil Brown, Jennylee Wood, Lisa Garner, Karen Lucas, Vivien Harris, Megan White, Keshia De Guzman, and Dr. Centaine Snoswell.

REFERENCES

1. Olson CA, McSwain SD, Curfman AL, Chuo J. The current pediatric telehealth landscape. *Pediatrics.* (2018) 141:e20172334. doi: 10.1542/peds.2017-2334
2. Smith AC. Telepaediatrics. *J Telemed Telecare.* (2007) 13:163–6. doi: 10.1258/135763307780908021
3. Smith AC. Telepaediatrics in Queensland. In: Wootton R, Batch J, editors. *Telepediatrics: Telemedicine and Child Health*. London: Royal Society of Medicine Press (2004). p. 25–39.
4. Smith AC, Isles A, McCrossin R, Van der Westhuyzen J, Williams M, Woollett H, et al. The point-of-referral barrier—a factor in the success of telehealth. *J Telemed Telecare.* (2001) 7:75–8. doi: 10.1258/1357633011937227
5. Smith AC, Coulthard M, Clark R, Armfield N, Taylor S, Goffe R, et al. Wireless telemedicine for the delivery of specialist paediatric services to the bedside. *J Telemed Telecare.* (2005) 11:81–5. doi: 10.1258/135763305775124669
6. Trnka PN, White MM, Renton WD, McTaggart S, Burke J, Smith AC. A retrospective review of telehealth services for children referred to a paediatric nephrologist. *BMC Nephrol.* (2015) 16:1–7. doi: 10.1186/s12882-015-0127-0
7. Smith AC, Garner L, Caffery L, McBride C. A review of paediatric telehealth for pre-and post-operative surgical patients. *J Telemed Telecare.* (2014) 20:400–4. doi: 10.1177/1357633X14552373
8. Rowell PD, Pincus P, White M, Smith AC. Telehealth in paediatric orthopaedic surgery in Queensland: a 10-year review. *ANZ J Surg.* (2014) 84:955–9. doi: 10.1111/ans.12753

9. Brownlee GL, Caffery LC, McBride CA, Patel B, Smith AC. Telehealth in paediatric surgery: accuracy of clinical decisions made by videoconference. *J Paediatr Child Health.* (2017) 53:1220–5. doi: 10.1111/jpc.13599
10. Smith AC, Pelekanos JT. Telemedicine in paediatric neurology. In: Wootton R, Patterson V, editors. *Teleneurology*. London: Royal Society of Medicine Press (2005). p. 117–35.
11. Smith AC, Batch J, Lang E, Wootton R. The use of online health techniques to assist with the delivery of specialist paediatric diabetes services in Queensland. *J Telemed Telecare.* (2003) 9:54–7. doi: 10.1258/135763303322596273
12. Justo R, Smith AC, Williams M, Van der Westhuyzen J, Murray J, Sciuto G, et al. Paediatric telecardiology services in Queensland: a review of three years' experience. *J Telemed Telecare.* (2004) 10:57–60. doi: 10.1258/1357633042614258
13. Smith AC, Williams M, Justo R. The multidisciplinary management of a paediatric cardiac emergency. *J Telemed Telecare.* (2002) 8:112–4. doi: 10.1258/1357633021937578
14. Smith AC, Kimble RM, O'Brien A, Mill J, Wootton R. A telepaediatric burns service and the potential travel savings for families living in regional Australia. *J Telemed Telecare.* (2007) 13:76–9. doi: 10.1258/135763307783247130
15. Smith AC, Youngberry K, Mill J, Kimble R, Wootton R. A review of three years experience using email and videoconferencing for the delivery of post-acute burns care to children in Queensland. *Burns.* (2004) 30:248–52. doi: 10.1016/j.burns.2003.11.003
16. Wood J, Stathis S, Smith AC, Krause J. E-CYMHS: an expansion of a child and youth telepsychiatry model in Queensland. *Australas Psychiatry.* (2012) 20:333–7. doi: 10.1177/1039856212450756
17. Ryan V, Stathis S, Smith AC, Best D, Wootton R. Telemedicine for rural and remote child and youth mental health services. *J Telemed Telecare.* (2005) 11:76–8. doi: 10.1258/135763305775124902
18. Smith AC, Stathis S, Randell A, Best D, Ryan V, Bergwever E, et al. A cost-minimization analysis of a telepaediatric mental health service for patients in rural and remote Queensland. *J Telemed Telecare.* (2007) 13:79–83. doi: 10.1258/135763307783247239
19. Bradford N, Herbert A, Mott C, Armfield NR, Young J, Smith AC. Components and principles of a pediatric palliative care consultation: results of a Delphi study. *J Palliat Med.* (2014) 17:1206–13. doi: 10.1089/jpm.2014.0121
20. Bradford NK, Young J, Armfield NR, Herbert A, Smith AC. Home telehealth and paediatric palliative care: clinician perceptions of what is stopping us? *BMC Palliat Care.* (2014) 13:29. doi: 10.1186/1472-684X-13-29
21. Smith AC, Armfield NR, Wu W, Brown C, Perry C. A mobile telemedicine-enabled ear screening service for Indigenous children in Queensland: activity and outcomes in the first three years. *J Telemed Telecare.* (2012) 18:485–9. doi: 10.1258/jtt.2012.gth114
22. Smith AC, Armfield NR, Wu W, Brown CA, Mickan B, Perry C. Changes in paediatric hospital ENT service utilisation following the implementation of a mobile, indigenous health screening service. *J Telemed Telecare.* (2013) 19:397–400. doi: 10.1177/1357633X13506526
23. Nguyen K, Smith AC, Armfield NR, Bensink M, Scuffham PA. Cost-effectiveness analysis of a mobile ear screening and surveillance service versus an outreach screening, surveillance and surgical service for indigenous children in Australia. *PLoS ONE.* (2015) 10:e0138369. doi: 10.1371/journal.pone.0138369
24. Smith AC, Armfield NR, White MM, Williams M, Koh G, Hurley T, et al. Clinical services and professional support: a review of mobile telepaediatric services in Queensland. *Stud Health Technol Inform.* (2010) 161:149–58. doi: 10.3233/978-1-60750-659-1-149
25. Williams M, Smith AC. Paediatric outreach services. *J Paediatr Child Health.* (2004) 40:501–3. doi: 10.1111/j.1440-1754.2004.00450.x
26. Smith AC, White MM, McBride CA, Kimble RM, Armfield NR, Coulthard MG. Multi-site videoconference tutorials for medical students in Australia. *ANZ J Surg.* (2012) 82:714–9. doi: 10.1111/j.1445-2197.2012.06212.x
27. Smith AC, Scuffham P, Wootton R. The costs and potential savings of a novel telepaediatric service in Queensland. *BMC Health Serv Res.* (2007) 7:35. doi: 10.1186/1472-6963-7-35
28. Xu C, Smith AC, Scuffham PA, Wootton R. A cost minimisation analysis of a telepaediatric otolaryngology service. *BMC Health Serv Res.* (2008) 8:30. doi: 10.1186/1472-6963-8-30
29. Armfield N, Donovan T, Bensink M, Smith AC. The costs and potential savings of telemedicine for acute care neonatal consultation: preliminary findings. *J Telemed Telecare.* (2012) 18:429–33. doi: 10.1258/jtt.2012.gth101
30. Armfield NR, Bradford N, White MM, Spitzer P, Smith AC. Humour Sans Frontieres: the feasibility of providing clown care at a distance. *Telemed eHealth.* (2011) 17:316–8. doi: 10.1089/tmj.2010.0166
31. Bradford N, Caffery L, Smith A. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. *Rural Remote Health J.* (2016) 16:3808. doi: 10.22605/RRH3808
32. Wade V, Soar J, Gray L. Uptake of telehealth services funded by Medicare in Australia. *Aust Health Rev.* (2014) 38:528–32. doi: 10.1071/AH14090
33. Snoswell CL, Caffery LJ, Haydon HM, Thomas EE, Smith AC. Telehealth uptake in general practice as a result of the coronavirus (COVID-19) pandemic. *Aust Health Rev.* (2020) 44: 737–40. doi: 10.1071/AH20183
34. Wade V, Eliott JA, Hiller JE. Clinician Acceptance is the key factor for sustainable telehealth services. *Qual Health Res.* (2014) 25:682–94. doi: 10.1177/1049732314528809
35. Smith AC. Telemedicine: challenges and opportunities. *Expert Rev Med Devices.* (2007) 4:5–7. doi: 10.1586/17434440.4.1.5
36. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. *Pediatr Res.* (2016) 79:169–76. doi: 10.1038/pr.2015.192
37. Marcin JP. Telemedicine in the pediatric intensive care unit. *Pediatr Clin.* (2013) 60:581–92. doi: 10.1016/j.pcl.2013.02.002
38. Dharmar M, Smith AC, Armfield NR, Trujano J, Sadorra C, Marcin JP. Telemedicine for children in need of intensive care. *Pediatr Ann.* (2009) 38:562–6. doi: 10.3928/00904481-20090918-08
39. McSwain SD, Marcin JP. Telemedicine for the care of children in the hospital setting. *Pediatr Ann.* (2014) 43:e44–9. doi: 10.3928/00904481-20140127-10
40. Reynolds CA, Maughan ED. Telehealth in the school setting: an integrative review. *J Sch Nurs.* (2015) 31:44–53. doi: 10.1177/1059840514540534
41. Finley JP. Paediatric telecardiology in Canada. *Paediatr Child Health.* (2009) 14:156–8. doi: 10.1093/pch/14.3.156
42. Dullet N, Geraghty E, Kaufman T, Dharmar M, King J, Smith AC, et al. Impact of a university-based outpatient telemedicine program on time savings, travel costs, and environmental pollutants. *Value Health.* (2017) 20:542–6. doi: 10.1016/j.jval.2017.01.014

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 01 December 2020
doi: 10.3389/fdgth.2020.585656

The Upcoming Role for Nursing and Assistive Robotics: Opportunities and Challenges Ahead

Eftychios G. Christoforou[1], *Sotiris Avgousti*[2], *Nacim Ramdani*[3], *Cyril Novales*[3] and *Andreas S. Panayides*[4]*

[1] Department of Mechanical and Manufacturing Engineering, University of Cyprus, Nicosia, Cyprus, [2] Department of Nursing, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus, [3] Université d'Orléans, INSA CVL, PRISME EA 4229, Orléans, France, [4] Department of Computer Science, University of Cyprus, Nicosia, Cyprus

As an integral part of patient care, nursing is required to constantly adapt to changes in the healthcare system, as well as the wider financial and societal environment. Among the key factors driving these changes is the aging of population. Combined with an existing shortage of nursing and caregiving professionals, accommodating for the patients and elderly needs within hospitals, elderly-care facilities and at a home setting, becomes a societal challenge. Amongst the technological solutions that have evolved in response to these developments, nursing and assistive robotics claim a pivotal role. The objective of the present study is to provide an overview of today's landscape in nursing and assistive robotics, highlighting the benefits associated with adopting such solutions in standard clinical practice. At the same time, to identify existing challenges and limitations that essentially outline the area's future directions. Beyond technological innovation, the manuscript also investigates the end-users' angle, being a crucial parameter in the success of robotics solutions operating within a healthcare environment. In this direction, the results of a survey designed to capture the nursing professionals' perspective toward more informed robotics design and development are presented.

Keywords: nursing robots, socially-assistive robots, physically-assistive robots, healthcare robotics, connected health, robotics, eHealth

OPEN ACCESS

Edited by:
Wei Chen,
Fudan University, China

Reviewed by:
Maria Vittoria Bulgheroni,
Ab.Acus, Italy
Siti Anom Ahmad,
Putra Malaysia University, Malaysia

***Correspondence:**
Andreas S. Panayides
panayides@cs.ucy.ac.cy

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *21 July 2020*
Accepted: *06 November 2020*
Published: *01 December 2020*

Citation:
Christoforou EG, Avgousti S, Ramdani N, Novales C and Panayides AS (2020) The Upcoming Role for Nursing and Assistive Robotics: Opportunities and Challenges Ahead. Front. Digit. Health 2:585656. doi: 10.3389/fdgth.2020.585656

INTRODUCTION

Nurses constitute the backbone of the healthcare industry and the nursing profession itself has typically been the largest segment of the healthcare workforce. Steadily rising healthcare costs and a population that is gradually aging are both factors impacting the healthcare systems and the nursing profession. A notable fact is that population aging is becoming a global phenomenon with wider financial and social implications. In the European Union alone, it is projected that older people ($\geq$65) will rise from 101 million in 2018 to 149 million by 2050. From a percentage angle, there will be an increase of 17.6 and 60.5% of people aged between 65–74 and 75–84, respectively in the EU-28, while the highest expansion growth is expected for very old people ($\geq$85) at a rate of 130.3%. On the opposite end, people aged $<$55 years will shrink by 9.6% during this period. At the same time, the old age dependency ratio (OADR) is expected to climb from 30.5% in 2018 to 49.9% in 2050 (i.e., from ~three persons of working age between 15 and 64 for every older adult in 2018 to two persons in 2050), with the global OADR projected to reach 28% (1, 2). At

a personal level, elderly individuals are challenged in various aspects including social (neglect, isolation, fear, loneliness, boredom), financial (low income, fear of becoming a burden, lack of insurance), psychological (depression, poor memory, dementia, insomnia), physiological (decline of mental abilities, less efficient reflexes, muscle weakness, weak body balance, falls, fragile bones) (3). For the above reasons, older people require special care that friends and relatives are often unable to provide and this commonly leads to institutionalization.

In response to the existing shortage of nursing and caregiving professionals, along with the rising healthcare costs, the employment of various technological solutions has been proposed. Technologies which have evolved to support the independent living and aging-in-place concepts include "Ambient Assisted Living" (4). The purpose of these technologies, also referred to as "smart home" technologies, is to support independent living using a combination of sensors appropriately installed in a house setup (stationary or wearable). Such sensors include magnetic switches, temperature sensors, photosensors, water flow sensors, motion sensors, force sensors, smoke detectors, and biosensors for vital signs (5). Ambient monitoring systems may capture activities of everyday living, which can then be exploited in two distinct ways: identify short-term emergencies; identify long-term variations in health status (6). Despite the relevance of these technologies to patient and elderly care, further consideration is beyond the scope of the present work, which focuses on robotics-related technologies.

From a robotics perspective, specially-designed systems have the potential to ease the burden on nursing staff within hospitals and nursing homes but also to undertake general assistive roles at home, without compromising quality of care while improving quality of life. Consistent to the above roles is the distinction between nursing and assistive robots:

a. *Nursing robots* may serve as supplemental healthcare workers in hospitals, elderly-care facilities, and at home. They can perform logistics and laborious physical tasks, combat loneliness and inactivity in the elderly population, or assigned routine tasks such as measuring patients' vital signs. Remote-controlled telerobots can handle interactive caretaker duties and serve as interfaces for doctors and/or nurses to communicate with patients and/or the elderly over distance.
b. *Assistive robots* may enable disabled and/or elderly people to pursue healthy, independent and productive lives. Depending on their primary role, assistive robots are grouped into: "Socially-assistive" and "Physically-assistive." The former, provide assistance to end-users through social interaction while the latter through physical interaction.

Enhanced capabilities for the above robotics technologies exist within the wider scope of telerobotics and telemedicine. Nursing and assistive robots are in fact part of the wider field of healthcare robotics, which also include the medical robotic systems. The latter has been an area of active research and various systems have already been established in clinical practice. Robotic systems are currently involved in surgical specialties including general surgery, orthopedic and neurosurgery, as well as other therapeutic procedures, such as radiation treatments (7). Realistically, employment of nursing and assistive robotics involves numerous challenges: technological, clinical, financial, insurance, psychological, social, ethical and legal. From a technological perspective challenges include indoor navigation, manipulation, safety, telecommunications, and integration of robots with existing in-hospital technologies. Key integration examples involve the connection to the hospitals' enterprise resource planning (ERP) and electronic health records (EHR) software systems. Corresponding tasks for nursing robots include performing logistics operations and vital signs measurements, respectively. On the other end, user perceptions and attitudes toward nursing and assistive robots are expected to have a decisive role on the future and the impact of these technologies. This is relevant both from the patients and the elderly point-of-view, as well as the nursing professionals and caregivers (8).

The purpose of this work is to provide an overview of the emerging fields of nursing and assistive robotics in order to highlight their potential and identify the involved challenges. The latter provides guidelines for robot design and directions for future developments in these areas. Informed design may constitute robotic solutions more usable and effective allowing them to better serve their purpose. The paper is organized as follows. It starts with an overview of nursing robots in section Nursing Robots and the discussion extends to socially-assistive and physically-assistive robots in sections Socially-Assistive Robots and Physically-Assistive Robots, respectively. In these sections, the added-value and potential of these robotic solutions are portrayed. Then, the enhancements in nursing and assistive robots facilitated via the integration with telerobotics technologies is highlighted in section A Role for Telerobotics. The emerging robotics role in disease outbreaks is discussed in section Robots in Times of Disease Outbreaks. Section Robots in Healthcare Environments: Endorse Concept Case Study examines the introduction of robots in the healthcare environment through a case study of the EU-funded ENDORSE project and a survey designed to capture end-user views upon different aspects of nursing robots. Challenges pertinent to future developments in the areas of nursing and assistive robotics are the topic of section The Challenges Ahead. The last section presents the conclusions.

NURSING ROBOTS

A role for nursing robots exists both in hospitals and elderly-care facilities. Robots may effectively relieve burden from nurses allowing them to concentrate on tasks pertinent to their primary duties. Robotic machines have already been considered to support processes including distribution of food trays, medicines, and laboratory specimens throughout a hospital. Robots may also automate logistics tasks relevant to medical equipment and supply storage. Beyond these tasks, an upgraded role for robots includes working alongside or collaborating with nurses to support their work and enhance efficiency. Moreover, robot nurses can help reduce occupational exposure of human nurses to hazardous infections or chemicals. Following special training, nurses may undertake the role of coordinating and overseeing the

duties of a robotic fleet within a hospital; thus, creating a new professional specialization.

Specially-designed robotic systems that help with patient transfers, ambulation, and lifting may significantly reduce physical stress on nurses. It is common for caregivers to suffer from back pain and job-related illnesses. Specially-designed robotic devices may be assigned laborious tasks, such as transferring and moving patients (9). This aspect also directs to the wider research on wearable exoskeleton devices. Exoskeletons may enhance a person's physical capabilities allowing lifting of heavier weights (power extenders), while preventing musculoskeletal disorders. In fact, exoskeletons provide an alternative to fully-automated robotic solutions, effectively preserving the human skills in the job.

Nursing robots may also provide services for telemedicine purposes (10). Robotic nurses accommodating telepresence platforms can effectively serve as interfaces for doctors to communicate with patients over distance. Typical scenarios involve routine virtual visits where the robot navigates to hospital wards employing the onboard screen to establish the required visual contact with the examined patients. Toward this direction, endowing robots with autonomous navigation capabilities is a particularly attractive feature, which relieves the necessity of operators manually navigating robots until a specific patient is located. Additionally, the robot may also capture the patient's vital signs at various intervals as required for a diagnosis and typical clinical protocols. In principle, the latter scenario further extends to the patient's home setup bringing specialized care to citizens and healthcare centers situated in remote and isolated areas.

Overall, electromechanical caregivers have unique advantages over their human counterparts including the capacity to work continuously throughout the day. Being programmable machines, robots have the potential to personalize care and adapt to varying needs. Importantly, robots can be integrated with other hospital technologies, such as cloud-based EHR systems, facilitating access to a patient's complete medical history and thus ensuring continuity of care.

SOCIALLY-ASSISTIVE ROBOTS

A socially-assistive robot is a type of assistive robot, which provides assistance to end-users through social interaction (11). A natural human tendency to attribute human characteristics and intentions to mobile physical entities, constitutes robots more effective than any computer program or a smartphone mHealth application. Potential uses of socially-assistive robots suggested in literature are discussed below and include: (i) companion robots; (ii) supporting adults with dementia; (iii) motivating physical exercise; and (iv) providing post-stroke rehabilitation.

Companion robots have emerged as a special category within assistive robotics. A primary role has been to act as interfaces for the elderly to enrich their social lives, while connecting with their families and friends (12). Among the capabilities of socially-assistive robots is to monitor elderly patients via video and also provide alerts to caregivers on patient activity. Moreover, robots may provide older citizens with news and entertainment information, reminders for medication adherence, as well as facilitate physical exercise. On a different note, robotic pets have received considerable attention in an attempt to reduce stress and depression, while avoiding the effort and risks involved in animal care (11). Another key area of socially-assistive robots concerns supporting people suffering from dementia (13–16).

Regular physical exercise is essential in elderly individuals to maintain and improve health status, support mental and physical well-being, and reduce the likelihood of depression. Robots have been designed to engage elderly users in physical exercise (17) facilitating workout sessions, while evaluating user performance and providing real-time feedback. Two potential implementation challenges for these robots were identified in Görer et al. (18). First, is the automatic analysis of the coach's gestures toward being adequately reproduced, and second, is the different physical embodiment that a robot possesses compared to the coach. Use of robotics also extents to post-stroke exercising and rehabilitation, which typically involves carefully designed repetitive, passive or active exercises (11). In either case, a movement therapy robot may provide a diagnostic (measurement and assessment) or therapeutic (improvement of function) benefit.

PHYSICALLY-ASSISTIVE ROBOTS

Two key elements of independent living that are directly associated to quality of life of both the elderly and patients (19) are: (a) The preservation of mobility; (b) The ability to manipulate objects. In elderly populations, a wide variety of medical conditions ranging from strokes and neurodegenerative diseases, to bone fractures and decline of muscular power, lead to the loss of mobility. To combat this situation, robotic solutions have been proposed to provide assistance required to stand-up, sit and walk (20). Robotic wheelchairs provide users with autonomy, enhanced mobility and safety (21). With an appropriate mechanical structure for the robotic wheelchair, architectural barriers may be overcome, including curb ascending and descending (22). In terms of control, a robotic wheelchair may hierarchically combine: (i) low-level functions (e.g., obstacle/collision avoidance, corridor centering) and (ii) high-level functions (e.g., directing the wheelchair) (23).

Appropriately-designed assistive robotic manipulation systems can support people with motor impairments such as limited hand and arm movements, high-level spinal injuries or tremors. Surveys have identified the relevant needs of disabled people in this group regarding assistive devices to carry out activities (23): eating and drinking (feeding assistive devices); personal care (washing, shaving, applying cosmetics); handling objects (books, devices); mobility and access (opening doors); general reaching and moving tasks. Manipulation systems addressing aforementioned challenges can be either fixed or wheelchair-mounted (24).

A ROLE FOR TELEROBOTICS

Within the wider field of healthcare, teleoperated medical robotic systems have been successfully employed allowing procedures such as surgeries, treatments, and diagnoses to be conducted across distances, while utilizing wired and/or wireless

communication networks. Recent developments in telerobotics and their enabling technologies [robotic manipulation, video streaming; (25), telecommunications] constitute nursing and assistive robots more effective and widen their application fields (26). Robotics hardware enhances telepresence to a more natural and effective level through mobility and performance of manipulation tasks in the remote environment. Telerobotics solutions pertinent to nursing may facilitate the doctors' virtual visits scenario. Using an onboard adjustable camera, the user may remotely drive the robot to locate a patient in the clinic and/or provide a set of destination points to which the robot will autonomously navigate. Bidirectional video conferencing then allows the doctor to appear on the robot's screen and engage in a dialogue with the patient to assess his/her current clinical status (telehealth). Real-time medical charts can further complement and enhance this remote clinical assessment using a robot-mounted device equipped with vital signs acquisition capabilities and EHR connectivity, such as the ENDORSE concept discussed in section Robots in Healthcare Environments: Endorse Concept Case Study.

Telepresence robots supporting elderly persons at home may facilitate social interaction, help the elderly to remain socially engaged, and allow relatives to make virtual visits and experience the feeling of close proximity. They also enable contact with doctors and nurses to remotely monitor their health and provide the required support. Compared to video calls, a telepresence robot enhances interaction to a more natural level through the mobility of the system. Enhanced capabilities for telerobotic systems are possible through their inherent compatibility with IT technologies, including internet-of-things (IoT), as for example the IoT-enabled telerobotics application in home care proposed in Zhou et al. (27). In (28), various telepresence robotic systems are reviewed and three main areas of application of telerobotics in elderly care become apparent, namely: telemedicine, remote interactions with other people, and telehealth monitoring.

ROBOTS IN TIMES OF DISEASE OUTBREAKS

In times of outbreaks of contagious diseases, healthcare workers are in high risk for infection due to direct contact with patients. This exposure can be minimized when robots undertake some nursing duties (29). In that case, nursing robots play analogous roles to emergency response robots deployed in contaminated sites (e.g., following a nuclear plant accident). In either case, robots become frontline actors preventing human exposure to health hazards. Through the novel coronavirus crisis (COVID-19) in 2020 has emerged a renewed interest in robotics solutions as effective resources to combat a pandemic (30). Despite the research and development in the fields of nursing and service robotics, the robotics community was found unprepared to drastically deploy effective solutions following COVID-19 pandemic. However, an upgraded role for nursing robots has emerged regarding their potential to reduce personal physical contact and exposure. Various robot applications have been identified including autonomous robots deployed to disinfect hospital wards using non-contact ultraviolet (UV) surface disinfection methods, deliver medicine, food trays and medical supplies, or handling of contaminated waste within a hospital. An indirect benefit is that robots help reduce the usage (and need for reuse) of personal protective equipment and also avoid contamination during its removal.

When large-scale screening programs are implemented, robots may contribute in the collection of samples, while limiting physical contact and increasing the coverage of a study. Robotic manipulation systems can then be employed in laboratory testing by automating the processing of large sample quantities. For diagnosis and screening purposes robots can also undertake temperature measurements in public areas and ports of entry.

The previously mentioned roles envisioned for assistive robotics (see sections Nursing Robots, Socially-Assistive Robots, Physically-Assistive Robots, A Role for Telerobotics) also become relevant toward addressing quarantine and social distancing implications. Socially assistive robots may provide patients and elderly with companionship and sustain social contact, while physical visits are not possible. Also, robots may physically support elderly/patients at home and facilitate health monitoring when family, friends or caregivers become less available. Rehabilitation therapies may continue without the physical presence of a physiotherapist, and physical exercise sessions at home or elderly care facilities can be carried out without an instructor. On a different note, mobile robots can be used to supervise social distancing rules in public areas, check usage of protective equipment and provide reminders and alerts. Furthermore, ground or aerial robotic vehicles can assist in policing quarantine areas and border control operations.

Moreover, teleoperated robotic manipulation systems can play a role in diagnosis and health monitoring without physical presence of a medical expert. In particular, telesonography robots [e.g., (31, 32)] can be used for pulmonary condition examinations and prevent cross-contamination in suspected patients. Recently, a telerobotic ultrasound system was considered for cardiopulmonary assessment of COVID-19 patients, as presented in Zhou et al. (27).

ROBOTS IN HEALTHCARE ENVIRONMENTS: ENDORSE CONCEPT CASE STUDY

Despite the documented clinical value and commercial potential, there is a limited market penetration of mobile robotic solutions in hospital environments today. The latter is further amplified by the fact that only a handful of solutions exist, which are purpose-oriented and do not adequately scale to accommodate the wide range and high demand of clinical services and logistics tasks that would translate into wider adoption. Moreover, vendors often overlook the importance and necessity of integrating their solutions with existing healthcare systems, while robotic fleets are vulnerable to cybersecurity attacks, and typically involve time consuming and costly infrastructure setups (33–35). These areas are currently attracting considerable research interest worldwide.

FIGURE 1 | ENDORSE project proof-of-concept mobile robotic fleet [http://www.endorse-project.eu/]. Robotnik's RB-1 robot is used with different, easily swappable, and mountable hardware modules. From left to right, a robotic arm (Universitat Politecnica de Valencia, Spain), a carrier component for logistic tasks (Robotnik, Spain), and an e-diagnostic module for vital signs acquisition (StreamVision, France) and communication to a cloud-based electronic health record (EHR) system (University of Cyprus, Cyprus).

The ENDORSE Concept

ENDORSE Concept is a European funded project aiming to address afore-described technological challenges and broaden the functional scope of mobile robotic solutions in indoor healthcare settings (36) (**Figure 1**). More specifically, innovation in ENDORSE is centered around the following four pillars:

(i) infrastructure-less indoor navigation of a mobile robots fleet;
(ii) intelligent Human-Robot Interaction (HRI) toward optimizing the seamless sharing of crowded spaces between humans and robots;
(iii) integration of ENDORSE software modules with corporate software solutions, complying with the latest EU regulations on data security;
(iv) development of modular hardware mechanisms to accommodate a diverse set of tasks and services by simply swapping reconfigurable component modules.

ENDORSE functionality will be demonstrated via the integration of an e-diagnostic support module for vital signs monitoring on a fleet of mobile robots, facilitating connectivity to cloud-based Electronic Health Records (EHR), and validated in an operational hospital environment for realistic assessment.

Survey on Nursing and Assistive Robots

Within the context of the ENDORSE project, a questionnaire was drafted aiming to capture end-users views upon different aspects of the ENDORSE concept in particular and robotics solutions in general. As such, the questionnaire was tailored with a focus primarily on nurses and secondarily on healthcare professionals and stakeholders, involved in the provision of clinical care in healthcare indoor settings. The latter, was the result of three focus groups that took place prior to finalizing the questionnaire. The first two focus groups involved senior nurses and experiences researchers, respectively, while the final one brought together both groups. The questionnaire was primarily circulated amongst the students and alumni of the Department of Nursing, Faculty of Health Sciences, of the Cyprus University of Technology (CUT) during September 2019. Secondarily, the questionnaire study involved experienced researchers in the broader electronic health and robotics areas. Here, it is important to highlight that a second questionnaire is scheduled as a part of ENDORSE research activities, aiming to capture the perceptions of patients. A total of 115 responders participated in the survey, of which more than 80% were nurses (16% of which were university students), 15% researchers with academic experience, 75% were aged between 18 and 34 years old, 75% were University graduates, and approximately two-thirds were females.

The questionnaire consisted of the following sections: (i) Demographics, (ii) Perceived behavioral control, (iii) Subjective norm, (iv) Safety and privacy considerations, (v) Operational perspective, and (vi) Management and financial perspective. An explanatory section highlighting the ENDORSE project's objectives (see section The ENDORSE Concept) preceded the survey questions, aiming to introduce the involved concepts to all participants. A link to the project's website and contact details for additional information were further provided. In what follows, key observations extracted from the analysis of 115 responses that were collected are described.

A Perceived Behavioral Control

The opening section of the questionnaire consisted of 6 questions and its primary goal was to capture the end-users perceived behavioral control. A promising 72% of the participants, as depicted in **Figure 2A**, considered themselves technologically competent, while only 7 out of 115 participants (disagree: ∼6%) considered themselves the opposite. Approximately 1/5 neither agreed nor disagreed with being technologically competent and familiar with technology. A very large percentage (∼65%), received some form of robotic education/interaction during his/her undergraduate and graduate studies. This percentage aligns with the percentage of responders that stated being technologically competent. The latter emphasizes and reiterates the importance of bringing robotic education in university courses, and especially in health sciences and not just in computer science and engineering disciplines. Importantly, an impressive ∼61% were aware of the DaVinci surgical robot, also showing the wide acceptance and penetration in clinical care this robotic solution enjoys over the past decade. Moreover, an interesting 36.5%, ∼34, and 27%, were familiar with exoskeleton robotic solutions, robots used for assisted living applications, and mobile robots used for logistic applications, respectively. On the other hand, approximately a fifth of the participants were not aware of any of the listed robotic solutions. More than 80% expressed their strong confidence in their ability to learn how to interact

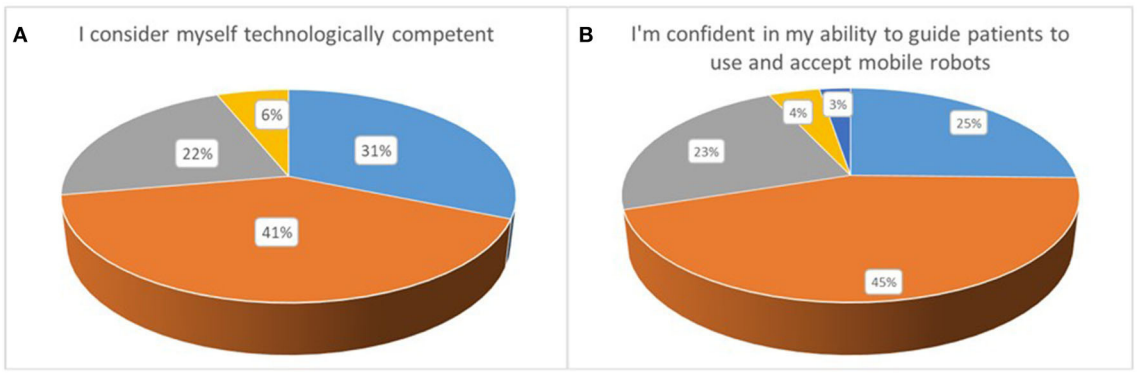

FIGURE 2 | (A,B) Perceived behavioral control.

and operate a mobile robot if one was to become a part of their healthcare unit. A similar percentage of ∼73% strongly agreed or agreed that they are confident in their ability to learn how to guide their colleagues in operating mobile robots working in indoor healthcare spaces. Only 8.6% disagreed while 18.3% neither agreed nor disagreed. An elevated 23% (compared to the previous two questions) neither agreed nor disagreed with respect to the statement that they feel confident in guiding patients on how to use and accept the use of mobile robots in their daily care routine. On the opposite end, ∼70% (strongly agree: ∼25%; agree: ∼45%) felt that learning how to guide patients coexist with indoor mobile robots should not be a challenge, as shown in **Figure 2B**. The last three questions do show a trend, that between 20 and 30% of responders are somehow skeptical with the idea of adopting a mobile robot in standard clinical care, with the underlying cause being how to convince patients or educate new colleagues in operating/interacting with mobile robots.

Subjective Norm

The subjective norm section of the questionnaire (7 questions) complements the perceived perspective, and was designed with the dual objective of first, capturing how healthcare professionals believe their colleagues would react in the adoption of mobile robots operating in a healthcare environment, and second, what they actually expect of that robot in practice.

An impressive 89.5% of the participants was enthusiastic of adopting an ENDORSE-like solution in their workplace, should this was linked to increasing the quality of the provided care, as highlighted in **Figure 3**. Importantly, there was only one response disagreeing on this particular question. However, in the following question, the participants appear to hesitate on how this would be received by their colleagues, with ∼44% responding as neither agreeing nor disagreeing in the statement that their broader workplace sees the adoption of robotic solutions in indoor healthcare environments in a positive angle. Approximately 10% were pessimistic (believe the opposite) while ∼46% were indeed optimistic (agree with the statement). The same trend was further documented in a follow-up question, trying to capture one's view with respect to their immediate colleagues. Again, 53% (against 44% in the previous question) selected a neutral response, depicting that they do believe that adopting mobile robots into daily healthcare routine and tasks is not a trivial task. More than one third however responded positively (strongly agree and agree: ∼38%) while only a minor percentage of ∼8% disagreed. In a similar question, that was phrased a bit differently, stating that colleagues would strongly resist the adoption of such solutions, a ∼27% agreed (strongly agree: ∼9%; agree: ∼18%). This was the highest documented response for a potentially negative statement, showing that healthcare professionals do not take for granted that their colleagues share the same perceptions with respect to adopting a potentially transforming robotic solution. Still, the highest percentage of ∼41% was neutral while almost 3 out of 10 (∼29%) disagreed; being convinced that no technology-oriented opposition would appear. A great sign of solidarity was document in the next question, where about two thirds or ∼68% responded that healthcare professionals would help each other, with any matter that should arise directly or indirectly, with the adoption of robotic solutions (see **Figure 3**). The next two questions revealed two key characteristics expected of robotic solutions that should be taken into careful consideration during design and development to facilitate user-acceptance. More specifically, ∼76 and 85%, expect from a mobile robot to respond promptly to its tasks and be available 24/7, respectively (see **Figure 4A**). Only ∼7 and 3.5% of responders did not have these expectations (the remaining being neutral), respectively.

Safety and Privacy Considerations

Six questions, addressing broader security (physical and cyber-security) and privacy considerations composed the present section of the questionnaire. The introductory question aimed at assessing whether the end-users felt that the introduction of mobile robots tasked with various operations that have so far been undertaken by clinical personnel (e.g., nurses), can potentially pose a threat to their occupation in the future. The responses were balanced, with 35% acknowledging that the latter scenario could become a reality, while 41% did not perceive the abovementioned statement as a threat, as demonstrated in **Figure 4B**. Approximately 24% gave a neutral response.

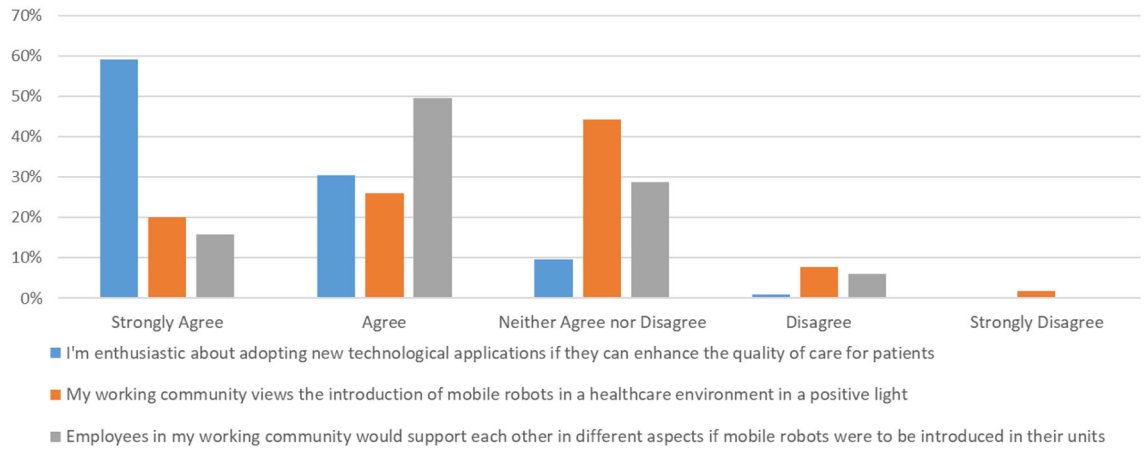

FIGURE 3 | Subjective norm.

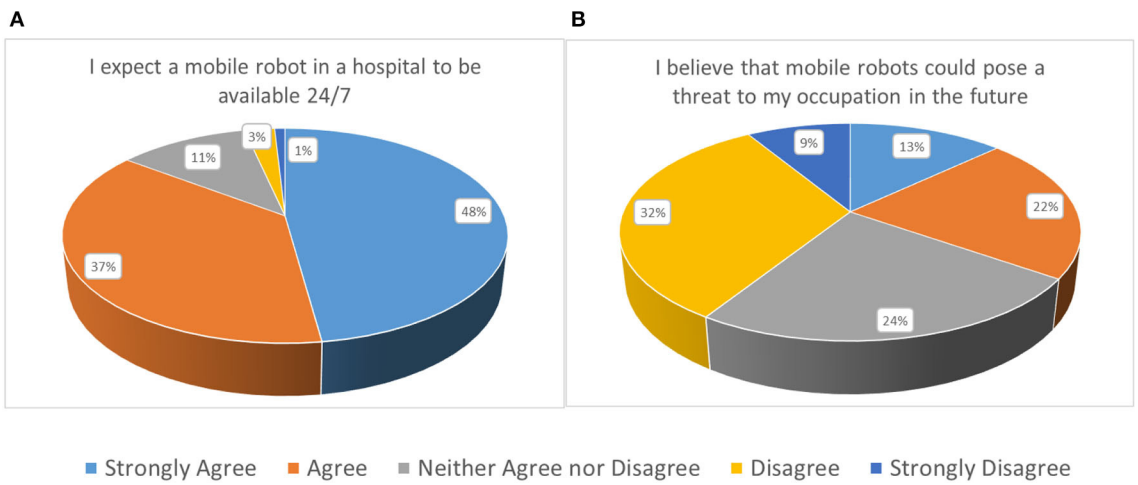

FIGURE 4 | (A) Subjective norm and **(B)** Safety and privacy considerations.

Here, it is important to highlight an observed trend that has been evident throughout the questionnaire. Individuals that strongly feel technologically competent have a more positive predisposition and are typically less concerned with any negative developments that might arise from robotic solutions in a healthcare indoor setting. The opposite holds for individuals that do not feel adequately secured from a technological competencies angle.

The responses to the next question, whether such a development could incur any security issues, deliberately phrased in a high-level manner, were again balanced. The highest percentage, or 43%, gave neutral responses, while ∼27 and 23% responded positively and negatively, respectively (see **Figure 5A**). Safer conclusions can be drawn from the next question, where specific examples concerning physical, infrastructure, and robot security were listed. More than half of the responders or ∼56% considered the specific examples possible, which emphasizes that a secure-by-design robotic fleet development is of primary essence and a catalytic factor in user acceptance in a full-scale deployment scenario.

With respect to privacy considerations, responses to the general question if such a deployment could pose privacy issues, positive responses were slightly elevated compared to the corresponding question on security, as depicted in **Figure 5A**. In particular, ∼38% believe that a privacy compromise is likely to occur, with 36% giving a neutral response, and ∼27% (the same as above) giving a negative response (i.e., do not believe that a privacy breach is a true risk). Indeed, responses in the specific examples in the following question were slightly elevated. In fact, a ∼37% was concerned of a privacy compromise of his/her personal data. Again, more than 4 out of 10 participants or ∼43% considered that all listed privacy concerns are possible.

Operational Perspective

The operational perspective section consisted of 6 questions and provided significant insights with respect to the projected

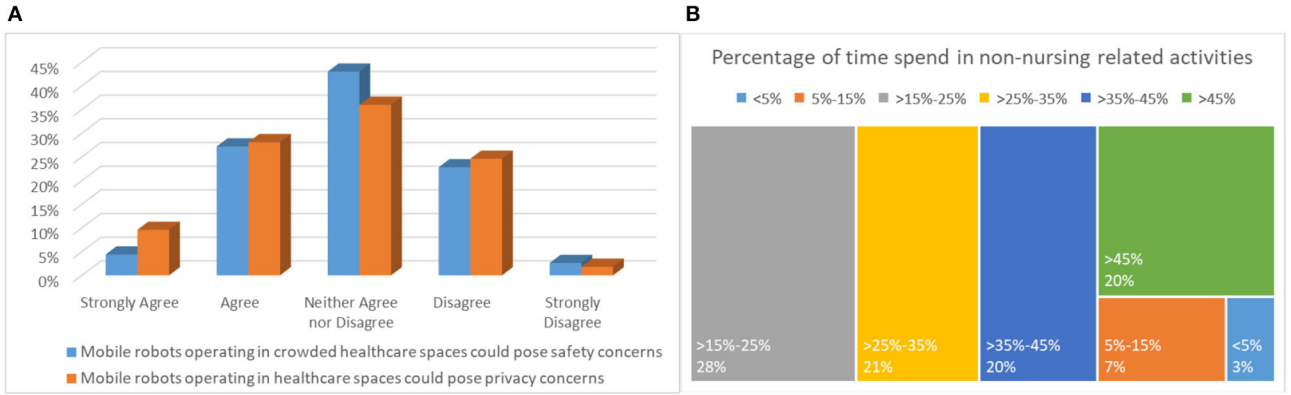

FIGURE 5 | (A) Safety and privacy considerations and **(B)** Operational perspective.

operational benefits of adopting an ENDORSE-like solution in routine daily care. An impressing 83.5% expects that physical burden will be significantly reduced, while 2/3 project that they will save time from tasks that do not relate to their primary mission of providing care, such as transfer of linens, food, waste, and other. In fact, only one responded that mobile robots are not suitable for such operations, demonstrating the physical and time burden experienced by healthcare professionals attributed to non-clinical tasks. Perhaps one of the most alerting responses of this questionnaire is associated to the next question, highlighted in **Figure 5B**, documenting the amount of time spent on non-clinical tasks. An extraordinary 89% responded that non-clinical tasks consume more than 15% of their time, of which ∼65% more than 25%, ∼40% more than 35%, and 20% more than 45%. The latter, is a key driving and motivating factor for designing efficient and effective robots that would assist healthcare professionals in their clinical, but more importantly, in their non-clinical tasks, allowing more time to be allocated for providing the appropriate levels of clinical care. Importantly, ∼73% believe that patients would welcome mobile robots undertaking the above-mentioned non-clinical tasks.

Further extending their acceptance to clinical tasks, more than half of the end-users participating in this survey, responded positively to the 5 listed clinical operations suggested in the next question. In fact, in 3 out of the 5 examples, the acceptance rate climbed to two out of three participants. In line with the aforementioned, 60% responded that they would trust a mobile robot undertaking certain clinical tasks such as the ones mentioned in this questionnaire (i.e., vital signs and medical data capture, electronic health records connectivity and medical data display, broader telemedicine and telehealth services, etc.), with only 10.5% declaring the opposite. Moreover, a noteworthy ∼69% believe that patients would welcome and accept a robotic solution undertaking certain clinical tasks, should the involved healthcare professionals allocated the time to explain and convince them that such a development is for their own benefit (see **Figure 6**). On the other hand, ∼31% replied that their feeling is that patients would be skeptical about such a scenario.

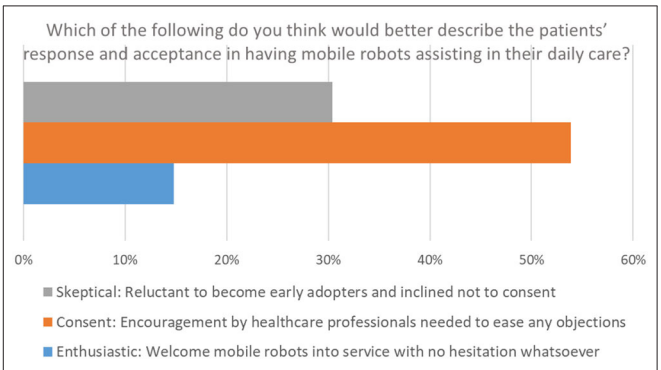

FIGURE 6 | Operational perspective.

Management and Financial Perspective

The key objective of these two sections is to document an initial reaction of potential end-users concerning ENDORSE solution incurred costs as well as associated healthcare expenditures savings. While the questionnaire does not provide adequate data for a fair expenditure estimation, the documented opinions are nevertheless important for future reference. Moreover, a second objective was to document potential barriers to adoption. The first question indeed reveals the two most important barriers to wide mobile robots deployment in indoor healthcare spaces according to end-users. The first is the associated financial burden, with more than 80%, showing that based on the provided information, end-users expect ENDORSE solution to be costly. The second, with ∼60%, is the patient acceptance, another key factor that has to be taken into consideration, provided that it is not technology-related. Other barriers that were ranked relatively high included end-user acceptance (∼53%) and privacy issues (∼52%).

Importantly, ∼89% of the participants, were not aware of robotic solution like the one ENDORSE is proposing, highlighting the innovation potential of the proposed solution. However, for the financial part of this questionnaire, we opted

to include only those who responded positively to the above question, in an attempt to receive more realistic cost estimations. Unfortunately, we only received 13 responses, of which ∼46% anticipate the ENDORSE solution to cost between 100 and 200 k and another 23% more than 200 k. The latter suggests that robots are still considered an expensive technology and not a commodity. In the last question, results were balanced, with half of the responders believing that ENDORSE solution would save <100 k annually, for a 3-story hospital with 100 beds, and the other half more than a 100 k.

THE CHALLENGES AHEAD

Among the key challenges to successful implementations of nursing and assistive robotics is the user acceptance. Of relevance are the perceptions of patients/elderly as well as nurses/caregivers. The role of geographical and cultural differences may not be underestimated, as for example the case of China that is discussed in (37). On behalf of the patients/elderly there exists a natural concern that robot deployments at home may replace personal contact and assistance, leading to a loss of companionship and increased isolation. Robots also have the ability to learn and process personal information, which is in effect a privacy violation. The presence of robotic systems at home creates the feeling of being under continuous surveillance. Interestingly, from a different perspective, robotic technologies may in fact increase the level of privacy by avoiding the need for human assistance for tasks which are perceived as private. From a psychological perspective, a prominent non-physical risk is the attachment to the robot and deception about its abilities.

Many nursing professionals are accustomed to emerging technologies impacting their work and daily duties. Relevant robotics content is slowly becoming part of nursing education (38). However, an innate concern that the introduction of robotics is likely to threaten their job security may become an obstacle to the adoption of nursing robotic solutions. Along user acceptance, technological challenges pertinent to the robotics technology itself remain to be addressed. For mobile robots operating within crowded, dynamic spaces (e.g., a hospital or a house setup) (36), standard sensing, localization and navigation techniques as applied in structured environments (e.g., a production facility) are not readily applicable. Likewise, safety concepts applied to manipulation systems, which have been well-established in industrial setups, also need to be reconsidered. The ENDORSE concept has been contributing toward that direction. Specifications for nursing robots should include sterilizability so that the robot is prevented from becoming a contamination agent itself; this is particularly important in robot deployments during contagious disease outbreaks. Pertinent to communications and integration of cloud technologies, a potential safety risk related to robotic nursing is the possibility of unauthorized access to healthcare databases and sensitive private information; hence, data security technologies become relevant.

The success of nursing and elderly-care robots requires the safe and effective interaction between human and machine, which is a topic for human factors engineering. For elderly-care purposes, interfaces for human-robot interaction should be usable and appealing to older generations, also considering their relatively limited exposure to modern ICT applications. In that respect, personalization can be used, which is an advantage inherent in programmable robots. For an effective human factors design, a requirement is the sound understanding of the user characteristics, possibly associated to diseases, accidents, aging, and birth defects. Specific to elderly populations there exist three principal categories of disabilities (39): (1) Physical impairments, including motor limitations, limit an individual's ability to reach and manipulate controls. (2) Perceptual impairments (sensory limitations) impair an individual's ability to receive information and feedback. (3) Cognitive limitations impair an individual's ability to process information.

The appearance and aesthetics of physically and socially-assistive robots is in general considered important to users (23). Appearances may take different forms including machine-like, humanoid, and software agents with human faces. In the bibliography it has been widely recognized that a robot's physical appearance leads to social expectations; a human appearance may lead to unrealistic expectations beyond the actual capabilities of the robot (40). Regarding motion systems, the majority of mobile robots are wheeled, given the advantage of less mechanical and control complexity. Despite their complexity, a main advantage of legged/anthropomorphic robots is their readiness to operate in environments and use tools originally designed for humans.

To fulfill their duties both nursing and assistive robots require some degree of autonomy, but is important that high-level control remains in the hands of the user. Widening the use of autonomous robotic technologies will require a legal as well as ethical framework to provide a foundation for further

TABLE 1 | Design requirements summary using the "Design for X" framework.

Life-cycle phase	X design parameter		
Development	• Simplicity • Safety • Reliability • Quality	• Modularity • Reprogrammability • Interchangeability • Expandability	• Upgradability • Integrability • Standards/ Regulations • Price
Production/ Manufacturing	• Manufacturability • Assembly	• Testing • Integration	• Cost • Materials
Use	• Usability • Human Factors • Ergonomics • Error-Resistance • Aesthetics • User-Friendliness • Customizability • Personalization • Clinical Relevance	• Multi-Use • Autonomy • Energy Autonomy • Mobility and Speed • Maneuverability • Manipulability • Stability • Energy-Efficiency	• Cost Effectiveness • Load Capacity • Sterilizability • Maintainability • Serviceability • Physical Safety • Logistics • Cyber-Security • User Privacy • Ethics
Disposal	• Recyclability	• Reusability	• Sustainability

The relevant X design parameters are grouped according to the life-cycle phase relevance.

developments. Among the unresolved issues is the attribution of civil and criminal liability should an autonomous robot produce damages (41, 42). The technological nature of nursing and care robots makes this issue rather complex.

Throughout the present study several key challenges pertinent to the introduction and use of nursing and assistive robotics have been identified and discussed, which eventually translate into corresponding design requirements. Informed design will constitute new robotic solutions more usable and effective, while facilitating acceptance by end-users. Toward that direction, the "design for X" (design for excellence) concept becomes relevant and it allows here to effectively summarize key requirements; the X variable is associated to different attributes of the system (e.g., safety) (43, 44). The identified design parameters are collected in **Table 1**. This design framework spans the whole life-cycle of nursing and assistive robots (the study outcomes provide input pertinent to the development and use phases). Noticeably, the requirements relevant to the use phase outnumber the requirements associated to the other life-cycle phases, which are mostly engineering and technological in nature. It is also pointed out that the compiled list of use phase requirements relates to the perspective of all stakeholders (nurses, patients, management).

CONCLUSIONS

Robots are currently impacting many aspects of our lives and their applications extend beyond their traditional applications in production. Nursing and assistive robotics are categorized within the broader scope of service robotics—the non-industrial uses of robots. In that context, autonomous and/or tele-operated robots, when employed in healthcare, can improve efficiency without compromising quality of care while reducing expenditures. Their mission further extends to elderly-care supporting the aging-in-place concept. Recently, an upgraded role for nursing robotics has emerged as effective means to combat outbreaks of infectious diseases. From a technical standpoint the involved technologies are mature. Yielding productive solutions necessitates integration of robotic components (mobile robots, manipulation systems, end-effectors, etc.) together with other enabling technologies (vision and image processing, video streaming, security, etc.). Toward this direction, robotic systems are inherently compatible and can be integrated with other contemporary technologies (i.e., internet-of-things, electronic health, etc.) to effectively increase their capabilities and clinical practice adoption.

Despite the potential of nursing and assistive robots there exist challenges that remain to be addressed prior to effective robot deployments of scale. Among key technological challenges, one can identify robot autonomy, indoor navigation and safe operation in healthcare settings. These areas are subject to further fundamental and applied scientific research. Inevitably, beyond the technological challenges, the perceptions and concerns of end-users toward these technologies will play a decisive role in future developments. The latter was the topic of the questionnaire study presented herein consolidating the nursing professionals' perspective on such pressing aspects as summarized next.

Interestingly, the majority of participants consider themselves technologically competent, confident with the idea of operating robots and interacting with them, as well as learning how to provide the required guidance to their colleagues and patients. Hence, it is no surprise that the same pool of responders appeared enthusiastic about adopting robotic solutions in their workplace. The latter, is rooted in the established expectations that robots will possess the ability to operate continuously throughout the day while promptly responding to the assigned tasks. In particular, nursing professionals anticipate that the adoption of robots in healthcare spaces will eventually alleviate the physical burden they currently experience that is attributed to non-clinical tasks, allowing them to concentrate on their primary clinical duties. Results highlighted that a considerable amount of time is actually consumed on often tedious, non-clinical tasks, such as logistics, transfers of linens, food and waste. In that context, the prevailing feeling among nursing professionals was that patients themselves would also be supportive of such robot implementations. Favorable responses were further recorded with respect to the acceptance of robots in reliably performing clinical tasks (e.g., capture of vital signs). The majority or responders believe that patients would react positively to the idea of robots undertaking clinical tasks, as long as healthcare professionals appropriately introduce the process.

In the opposite end, skepticism was indeed expressed by a certain percentage regarding the adoption of robots in clinical care due to the difficulties in convincing patients and educating new colleagues. The latter can be associated with the fact that responders did not have a clear view of whether their colleagues would indeed support such a transformative change. A significant concern that cannot be overlooked, although not being the prevalent impression, involves the scenario where robotic solutions pose a job security threat in the future. Toward this direction, primary user-acceptance concerns extend over security and privacy. Particular concerns have emerged with respect to both physical safety and cyber security. It is vital that these issues are thoroughly addressed via a secure-by-design approach (45–49). Likewise, concerns surfaced regarding the potential privacy compromises emanating from the presence of the robots and the corresponding sensitive data processing. To overcome these justified apprehensions, a clear and unambiguous regulatory framework overseeing nursing robot operations should be jointly developed by all involved stakeholders in collaboration with national authorities.

A key aspect affecting wider adoption in clinical practice involves the management and financial perspective. Demand exists for safe, reliable and cost-effective solutions, facilitating fast deployment and integration to existing IT infrastructure. However, the present robotic landscape market does not yet meet these expectations. As a result, there is a need for technological breakthroughs such as the ones pursued by the ENDORSE project that will remove existing financial barriers toward the large-scale adoption of nursing robots in healthcare environments. The realization of such advancements will in turn trigger the documented clinical and non-clinical benefits and effectively materialize reduction of healthcare expenditures in the near future.

The current review shows the potential that exists for nursing and assistive robotics, which was documented through the survey results. Clearly, challenges to be addressed extend beyond the technological and clinical issues to user acceptance. Despite an overall positive attitude that was recorded toward the introduction of robotics technologies some skepticism was also evident. Adequately addressing these concerns will be important for their future and informed robot design becomes critical toward that direction.

To promote the use of these technologies there exist three axes for targeted action, directed toward the end-users. Firstly, nursing and assistive robotics can become part of the nursing professionals' education, familiarizing and allowing them to effectively utilize these tools but also appropriately present them to patients and older adults. It is important that the capabilities of the robots are clarified and their role is transparent to the users. The second direction is to ensure direct involvement of all stakeholders in the product development stage, with patients' associations engaged in a leading role. Beyond the desirable effect on user acceptance it will eventually result to more efficient clinically-oriented solutions. Finally, wider adoption of nursing and assistive robotics will depend on successful implementations and demonstrations in clinical practice, while keeping in mind that evaluation will be on the basis of healthcare quality and cost effectiveness.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 823887.

REFERENCES

1. Albertone G, Allen S, Redpath A. *Ageing Europe*. 2019 ed, Luxembourg: Publications Office of the European Union (2019).
2. United Nations. *World Population Ageing*. United Nations, Department of Economic and Social Affairs (2019).
3. Kumar ES, Sachin P, Vignesh BP, Ahmed MR. Architecture for IOT based geriatric care fall detection and prevention. In: *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems* (Madurai: ICICCS). (2017). p. 1099–104.
4. Christoforou EG, Panayides AS, Avgousti S, Masouras P, Pattichis CS. An overview of assistive robotics and technologies for elderly care. *IFMBE Proc.* (2020) 76:971–6. doi: 10.1007/978-3-030-31635-8_118
5. Uddin Z, Khaksar W, Torresen J. Ambient sensors for elderly care and independent living: a survey. *Sensors.* (2018) 18:2027. doi: 10.3390/s18072027
6. Tsukiyama T. In-home health monitoring system for solitary elderly. *Proc Comput Sci.* (2015) 63:229–35. doi: 10.1016/j.procs.2015.08.338
7. Avgousti S, Christoforou EG, Panayides AS, Masouras P, Vieyres P, Pattichis CS. Robotic systems in current clinical practice. In: *20th IEEE Mediterranean Eletrotechnical Conference IEEE MELECON* (Palermo: IEEE) (2020).
8. Turja T, Taipale S, Kaakinen M, Oksanen A. Care workers' readiness for robotization: identifying psychological and socio-demographic determinants. *Int J Soc Robot.* (2020) 12:79–90. doi: 10.1007/s12369-019-00544-9
9. Tashiro T, Aoki K, Lee Y, Sakaki T. Research and development of wearable auxiliary tool for behavior assistance of elderly who requires nursing care. In: *International Conference on Control, Automation and Systems.* Jeju: IEEE Computer Society (2017). p. 1501–4.
10. Kyriacou E, Pattichis MS, Pattichis CS, Panayides A, Pitsillides A. M-health e-emergency systems: current status and future directions. *IEEE Antennas and Propagation Magazine* (2007, June 11).
11. Tapus A, Mataric MJ, Scassellati B. Socially assistive robotics [grand challenges of robotics]. *IEEE Robot Automat Mag.* (2007) 14:35–42. doi: 10.1109/MRA.2007.339605
12. Portugal D, Alvito P, Christodoulou E, Samaras G, Dias J. A study on the deployment of a service robot in an elderly care center. *Int J Soc Robot.* (2019) 11:317–41. doi: 10.1007/s12369-018-0492-5
13. Marti P, Bacigalupo M, Giusti L, Mennecozzi C, Shibata T. Socially assistive robotics in the treatment of behavioural and psychological symptoms of dementia. In: *Proceedings of the First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006.* Pisa: IEEE (2006). p. 483–8.
14. Kramer SC, Friedmann E, Bernstein PL. Comparison of the effect of human interaction, animal-assisted therapy, and AIBO-assisted therapy on long-term care residents with dementia. *Anthrozoös.* (2009) 22:43–57. doi: 10.2752/175303708X390464
15. Shibata T. Therapeutic seal robot as biofeedback medical device: qualitative and quantitative evaluations of robot therapy in dementia care. *Proc IEEE.* (2012) 100:2527–38. doi: 10.1109/JPROC.2012.2200559
16. Begum M, Wang R, Huq R, Mihailidis A. Performance of daily activities by older adults with dementia: the role of an assistive robot. *IEEE Int Conf Rehabil Robot.* (2013) 2013:6650405. doi: 10.1109/ICORR.2013.6650405
17. Fasola J, Matarić MJ. Using socially assistive human-robot interaction to motivate physical exercise for older adults. *Proc IEEE.* (2012) 100:2512–26. doi: 10.1109/JPROC.2012.2200539
18. Görer B, Salah AA, Akin HL. A robotic fitness coach for the elderly. *Lect Notes Comput Sci.* (2013) 8309:124–39. doi: 10.1007/978-3-319-03647-2_9
19. Balaguer C, Gimenez A, Jardon A, Cabas R, Correal R. Live experimentation of the service robot applications for elderly people care in home environments. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems.* Edmonton, AB: IEEE, IROS (2005). p. 2733–8.
20. Takahara S, Jeong S. Prototype design of robotic mobility aid to assist elderly's standing-sitting, walking, and wheelchair driving in daily life. In: *International Conference on Control, Automation and Systems.* Seoul: IEEE Computer Society (2014). p. 470–3.
21. Carlson T, Demiris Y. Collaborative control for a robotic wheelchair: evaluation of performance, attention, and workload. *IEEE Trans Syst Man Cybern B Cybern.* (2012) 42:876–88. doi: 10.1109/TSMCB.2011.2181833
22. Candiotti JL, Daveler BJ, Kamaraj DC, Chung CS, Cooper R, Grindle GG, et al. A heuristic approach to overcome architectural barriers using a robotic wheelchair. *IEEE Trans Neural Syst Rehabil Eng.* (2019) 27:1846–54. doi: 10.1109/TNSRE.2019.2934387
23. Hersh M. Overcoming barriers and increasing independence–service robots for elderly and disabled people. *Int J Adv Robot Syst.* (2015) 12:114. doi: 10.5772/59230
24. Ktistakis IP, Bourbakis NG. A survey on robotic wheelchairs mounted with robotic arms. In: *2015 National Aerospace and Electronics Conference (NAECON).* Dayton, OH: IEEE (2015). p. 258–62.

25. Panayides AS, Pattichis MS, Pantziaris M, Constantinides AG, Pattichis CS. The battle of the video codecs in the healthcare domain - a comparative performance evaluation study leveraging VVC and AV1. *IEEE Access.* (2020) 8:11469–81. doi: 10.1109/ACCESS.2020.2965325
26. Avgousti S, Christoforou EG, Panayides AS, Voskarides S, Novales C, Nouaille L, et al. Medical telerobotic systems: current status and future trends. *BioMed Eng.* (2016) 15:96. doi: 10.1186/s12938-016-0217-7
27. Zhou H, Yang G, Lv H, Huang X, Yang H, Pang Z. IoT-enabled dual-arm motion capture and mapping for telerobotics in home care. *IEEE J Biomed Health Inform.* (2020) 24:1541–9. doi: 10.1109/JBHI.2019.2953885
28. Reis A, Xavier R, Barroso I, Monteiro MJ, Paredes H, Barroso J. The usage of telepresence robots to support the elderly. In: *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW).* Thessaloniki: IEEE (2018). p. 1–6.
29. Li Z, Moran P, Dong Q, Shaw RJ, Hauser K. Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas. In: *Proceedings - IEEE International Conference on Robotics and Automation.* Singapore: Institute of Electrical and Electronics Engineers Inc. (2017). p. 3581–6.
30. Yang GZ, Nelson BJ, Murphy RR, Choset H, Christensen H, Collins SH, et al. Combating COVID-19-the role of robotics in managing public health and infectious diseases. *Sci Robot.* (2020) 5:eabb5589. doi: 10.1126/scirobotics.abb5589
31. Vieyres P, Poisson G, Courreges F, Merigeaux O, Arbeille P. The TERESA project: from space research to ground tele-echography. *Ind Robot.* (2003) 30:77–82. doi: 10.1108/01439910310457742
32. Vieyres P, Poisson G, Courrèges F, Smith-Guerin N, Novales C, Arbeille P. A tele-operated robotic system for mobile tele-echography: the Otelo project. *M-Health.* (Boston, MA: Springer) (2006) 461–73. doi: 10.1007/0-387-26559-7_35
33. Bonaci T, Herron J, Yusuf T, Yan J, Kohno T, Chizeck HJ. To make a robot secure: an experimental analysis of cyber security threats against teleoperated surgical robots. *arXiv[Preprint].arXiv:1504.04339.* (2015).
34. ENISA. *Cyber Security and Resilience for Smart Hospitals* (2016).
35. Reportsanddata. *Healthcare Cybersecurity Market to Reach USD 27.1 Billion by 2025.* (2019). Available online at: https://www.globenewswire.com/news-release/2019/08/26/1906602/0/en/Healthcare-Cybersecurity-Market-To-Reach-USD-27-10-Billion-By-2026-Reports-And-Data.html
36. Ramdani N, Panayides A, Karamousadakis M, Mellado M, Lopez R, Christophorou C, et al. A safe, efficient and integrated indoor robotic fleet for logistic applications in healthcare and commercial spaces: the endorse concept. In: *Proceedings - IEEE International Conference on Mobile Data Management.* Hong Kong: IEEE (2019). p. 425–30.
37. Sifeng Z, Min T, Zehao Z, Zhao Y. Capturing the opportunity in developing intelligent elderly care robots in china challenges, opportunities and development strategy. In: *Proceedings of IEEE Workshop on Advanced Robotics and Its Social Impacts, ARSO.* Shanghai: IEEE Computer Society (2016). p. 61–6.
38. Mudd SS, McIltrot KS, Brown KM. Utilizing telepresence robots for multiple patient scenarios in an online nurse practitioner program. *Nurs Educ Perspect.* (2019) 41:260–62. doi: 10.1097/01.NEP.0000000000000590
39. Denno S, Isle BA, Ju G, Koch CG, Metz SV, Penner R, et al. *Human Factors Design Guidelines for the Elderly and People with Disabilities*. Technical Report SSDC-SYS/AI-C92-009. Minneapolis, MN: Honeywell (1992).
40. Feil-Seifer D, Mataric M. Socially assistive robotics. *IEEE Robot Autom Mag.* (2011) 18:24–31. doi: 10.1109/MRA.2010.940150
41. Brozek B, Jakubiec M. On the legal responsibility of autonomous machines. *Artif Intell Law.* (2017) 25:293–304. doi: 10.1007/s10506-017-9207-8
42. Abbott R, Borges G, Dacoronia E, Devillier N, Jankowska-Augustyn M, Karner E, et al. *Liability for Artificial Intelligence and Other Emerging Digital Technologies*. Report from the Expert Group on Liability and New Technologies – New Technologies Formation European Union. (2019). Available online at: https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608
43. Bralla JG. *Design for Excellence*. New York, NY: McGraw-Hill Professional Publishing (1996).
44. Pahl G, Beitz W. *Engineering Design - A Systematic Approach*. 2nd ed. London; Berlin; New York, NY: Springer Science & Business Media (1996).
45. Dieber B, Breiling B, Taurer S, Kacianka S, Rass S, Schartner P. Security for the robot operating system. *Robot Auton Syst.* (2017) 98:192–203. doi: 10.1016/j.robot.2017.09.017
46. Dieber B, White R, Taurer S, Breiling B, Caiazza G, Christensen H, et al. Penetration testing ROS. *Stud Computat Intell.* (2020) 831:183–225. doi: 10.1007/978-3-030-20190-6_8
47. Guerrero-Higueras ÁM, DeCastro-García N, Matellán V. Detection of cyber-attacks to indoor real time localization systems for autonomous robots. *Robot Auton Syst.* (2018) 99:75–83. doi: 10.1016/j.robot.2017.10.006
48. Vilches VM, Gil-Uriarte E, Ugarte IZ, Mendia GO, Pisón RI, Kirschgens LA, et al. Towards an open standard for assessing the severity of robot security vulnerabilities, the Robot Vulnerability Scoring System (RVSS). *arXiv[Prerpint].arXiv:1807.10357.* (2018).
49. Vilches VM, Kirschgens LA, Calvo AB, Cordero AH, Pisón RI, Vilches DM, et al. Introducing the Robot Security Framework (RSF). p. a standardized methodology to perform security assessments in robotics. *arXiv[Preprint].arXiv:1806.04042.* (2018).

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 07 December 2020
doi: 10.3389/fdgth.2020.572671

Image Descriptors for Weakly Annotated Histopathological Breast Cancer Data

Panagiotis Stanitsas[1], *Anoop Cherian*[2], *Vassilios Morellas*[1], *Resha Tejpaul*[1], *Nikolaos Papanikolopoulos*[1]* *and Alexander Truskinovsky*[3]

[1] Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, United States, [2] Australian Center for Robotic Vision, Australian National University, Canberra, ACT, Australia, [3] Department of Pathology & Laboratory Medicine, Roswell Park Cancer Institute, Buffalo, NY, United States

Introduction: Cancerous Tissue Recognition (CTR) methodologies are continuously integrating advancements at the forefront of machine learning and computer vision, providing a variety of inference schemes for histopathological data. Histopathological data, in most cases, come in the form of high-resolution images, and thus methodologies operating at the patch level are more computationally attractive. Such methodologies capitalize on pixel level annotations (tissue delineations) from expert pathologists, which are then used to derive labels at the patch level. In this work, we envision a digital connected health system that augments the capabilities of the clinicians by providing powerful feature descriptors that may describe malignant regions.

Material and Methods: We start with a patch level descriptor, termed Covariance-Kernel Descriptor (CKD), capable of compactly describing tissue architectures associated with carcinomas. To leverage the recognition capability of the CKDs to larger slide regions, we resort to a multiple instance learning framework. In that direction, we derive the Weakly Annotated Image Descriptor (WAID) as the parameters of classifier decision boundaries in a Multiple Instance Learning framework. The WAID is computed on bags of patches corresponding to larger image regions for which binary labels (malignant vs. benign) are provided, thus obviating the necessity for tissue delineations.

Results: The CKD was seen to outperform all the considered descriptors, reaching classification accuracy (ACC) of 92.83%. and area under the curve (AUC) of 0.98. The CKD captures higher order correlations between features and was shown to achieve superior performance against a large collection of computer vision features on a private breast cancer dataset. The WAID outperform all other descriptors on the Breast Cancer Histopathological database (BreakHis) where correctly classified malignant (CCM) instances reached 91.27 and 92.00% at the patient and image level, respectively, without resorting to a deep learning scheme achieves state-of-the-art performance.

Discussion: Our proposed derivation of the CKD and WAID can help medical experts accomplish their work accurately and faster than the current state-of-the-art.

Keywords: connected health for breast cancer, image descriptors, annotated data, histopathological data, connected health and computer vision

OPEN ACCESS

Edited by:
Andreas S. Panayides, University of Cyprus, Cyprus

Reviewed by:
Parisis Gallos, National and Kapodistrian University of Athens, Greece
Constantinos S. Pattichis, University of Cyprus, Cyprus

***Correspondence:**
Nikolaos Papanikolopoulos
npapas@cs.umn.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *15 June 2020*
Accepted: *26 October 2020*
Published: *07 December 2020*

Citation:
Stanitsas P, Cherian A, Morellas V, Tejpaul R, Papanikolopoulos N and Truskinovsky A (2020) Image Descriptors for Weakly Annotated Histopathological Breast Cancer Data. Front. Digit. Health 2:572671. doi: 10.3389/fdgth.2020.572671

INTRODUCTION

About one in eight U.S. women (about 12%) will develop invasive breast cancer over the course of her lifetime[1]. Even though there is a widespread adoption of mammography, interpretation of these images remains challenging. Some of the fundamental morphological characteristics of malignant tumors includes (i) an increased number of cell nuclei per unit area, (ii) increased size of the nuclei, (iii) the nuclei staining darker than those of benign cells (nuclear hyperchromasia), (iv) greater than normal variability in the size and shape of nuclei, and (v) irregular nuclear contours. Therefore, the number, irregularity, and contrast of edges are all expected to increase in malignant tumors compared with benign tissues as noted by Basavanhally et al. (1) and Irshad et al. (2). The diagnostic questions pathologists face depend on the clinical situation and the required characteristics for determining whether a lesion is cancerous. The use of Computer Aided Diagnosis (CAD) schemes can better assist medical experts with their everyday tasks in determining whether a lesion is cancerous or not, the geometric characteristics of the location of the tumor, size, and its relation to the surgical margins with anatomic and histological landmarks.

Cancerous tissue recognition (CTR) from histopathological data is a particularly challenging task since it requires a close examination of tissue slides from suspected regions under a microscope which can be time-consuming hence constraining the number of cases pathologists can handle daily.

An automated identification of the regions that are highly likely to be cancerous can assist experts in finding them among the surrounding tissues efficiently, resulting in faster diagnosis. This is a part of a larger vision in digital connected health that will enable clinicians not matter where they are located to provide more informed assessments and decision-making than the current state-of-the-art.

In order to be trained effectively, most available cancerous tissue recognition (CTR) schemes require pixel level annotations, collected in the form of tissue delineations from expert pathologists [e.g., Sirinukunwattana et al. (3), Spanhol et al. (4), Xu et al. (5), and Xu et al. (6)], which are then used to produce labels at the patch level. Nevertheless, collecting such delineations is error prone and depends on individual experts' judgment toward identifying accurate transition boundaries between healthy and tumorous tissues. In contrast, relaxing the requirement for such tight tissue delineations and instead asking for annotations only at the bounding box (or whole slide) level can significantly reduce the effort from the experts. Similar considerations have appeared in the medical image analysis literature [e.g., Bejnordi et al. (7), Dundar et al. (8), Xu et al. (5), and Xu et al. (6)]. However, to the best of our knowledge, such studies have not looked at weakly-supervised inference from the perspective of representation learning, which is the primary contribution of this work. In this work, we propose a framework for training cancerous tissue recognition (CTR) schemes in the presence of weakly annotated data to expedite the analysis of Hematoxylin & Eosin (H & E)-stained tissue samples.

[1]Breast Cancer: Statistics. Available online at: https://www.breastcancer.org/

We propose a two-step framework for recognition in breast cancer data. Our key insight comes from the process by which the tissue slides are stained, specifically, the Hematoxylin & Eosin (H&E) staining scheme. This process gives unique color and texture to the tissue samples, and our approach is to derive a feature descriptor that leverages on these image properties. First, we derive the Covariance-Kernel Descriptor (CKD), a patch level descriptor that compactly describes tissue architectures associated with malignant areas and achieves superior performance on the problem of Cancerous Tissue Recognition (CTR) against a diverse collection of image descriptors including deep learning derived features. The origins of the Covariance-Kernel Descriptor (CKD) in his area can be traced in a previous work from our group by Stanitsas et al. (9). Second, we devise the Weakly Annotated Image Descriptor (WAID), an image descriptor geared toward larger slide regions that capitalizes on the covariance-kernel descriptor (CKD). The weakly annotated image descriptor (WAID) provides inference on larger image regions, while uplifting the requirement for pixel level annotations.

MATERIALS AND METHODS

Data Description

Fully Annotated Breast Cancer Database (FABCD)

For FABCD, tissue samples collected are Hematoxylin & Eosin (H&E) stained (10), followed by high-resolution (10K $\times$ 9K pixels) scans of tissue sections taken at x50 magnification on a digital slide scanner. Medical experts (surgical pathologists) were responsible for providing annotations corresponding to the malignant and benign image regions. The annotated regions are then divided into smaller disjoint patches of 150 $\times$ 150 pixels. Twenty-one annotated images of carcinomas and 19 images of benign tissue taken from 21 patients were combined toward constructing the FABCD. Binary class labels are assigned to each of the image patches in **Figure 1**. That is, those patches for which more than 80% of the pixels correspond to carcinomas are treated as the positive class, while patches in the negative class are devoid of any cancerous regions.

Breast Cancer Histopathological Database (BreakHis)

BreakHis (11) contains data from 82 patients at four different digital magnifications (40X, 100X, 200X, and 400X). For every magnification level approximately 2,000 H&E-stained tissue slides are collected of size 700 x 460 pixels, while binary labels (benign vs. malignant) and ordinal (four types of malignant and four types of benign) are provided. The magnification of 40x is aligned with the objectives of this study. Medical expert is requested to provide images in the form of bounding boxes surrounding suspicious regions of the whole slide as shown in **Figure 2**.

Covariance-Kernel Descriptors (CKD)

In this work, we compute the region covariance descriptors (RCDs) as proposed by Porikli et al. (12) over a set of features extracted from every pixel in the image patch. In their basic form, RCDs (denoted C_z) by Tuzel et al. (13) are generated as

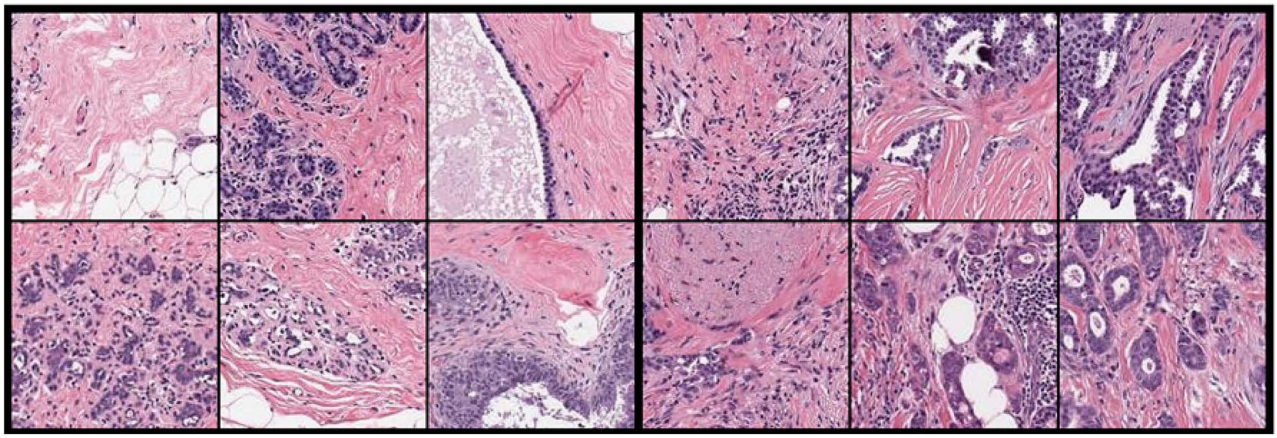

FIGURE 1 | Breast tissue H&E stained patches from 12 samples of the breast cancer dataset with the first three columns illustrating benign cases, while columns 4–6 depict image patches associated with carcinomas.

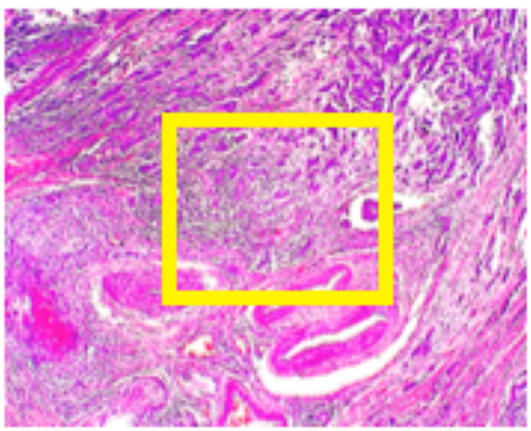

FIGURE 2 | BreakHis sample at 40x magnification for a malignant segment tissue with a yellow bounding box indicating the approximate location of the malignant tissue (for visualization purposes) surrounded by healthy tissue.

described in Equation (1), where $f_i \in R^d$, are d-dimensional features extracted from each pixel $i \in \{1,2,\cdots,N\}$ of an image patch $\mathbf{z}$, and μ is the mean feature given by $\mu = \frac{1}{N}\sum_{i=1}^{N} f_i$.

$$C_z = \frac{1}{(N-1)} \sum_{i=1}^{N} (f_i - \mu)(f_i - \mu)^T. \quad (1)$$

We consider a 5-dimensional RCD consisting of the normalized intensities of the three channels **R**, **G**, and **B** of a color patch combined with first-order gradient information along the x and y axis, as denoted by $\mathrm{Gr}_{\mathrm{i}}^{\mathrm{x}}$ and $\mathrm{Gr}_{\mathrm{i}}^{\mathrm{y}}$ respectively. That is, our f_i has the following form (for pixel i in the image patch):

$$f_i = \left[\mathrm{R_i}\ \mathrm{G_i}\ \mathrm{B_i}\ \mathrm{Gr}_{\mathrm{i}}^{\mathrm{x}}\ \mathrm{Gr}_{\mathrm{i}}^{\mathrm{y}}\right]^T. \quad (2)$$

Covariance-kernel descriptors (CKDs) are computed as the fusion of the Region Covariance Descriptors (RCDs) (12) and Normalized Color Histograms (NCHs) (in conjunction with the work in (14)) that are used to reveal information uncovered by the Hematoxylin & Eosin (H&E) staining. Toward deriving the NCH, for a given patch, we computed color histograms consisting of 256 bins each for the R, G, and B color channels; this histogram is normalized to sum to one and concatenated to form a 768-dimensional feature descriptor for the respective patch. RCDs compute the feature correlations at the pixel level (local) in a patch and in that way capture texture and shape in the patch implicitly. In contrast, NCH represents global color information at the patch's vicinity. The combination of both global and local information captures complementary cues for recognition which are essential. However, rather than concatenating the three histograms, as in the case of NCH, we combine them to formulate a matrix $\boldsymbol{H} \in R^{3\times b}$, where each row corresponds to the b-bin histogram on a channel and enables us to capture global color correlations *via* the modality $\boldsymbol{HH^T}$. In that way, for an image patch $\boldsymbol{z}$, the CKD is computed in the form of a compact block diagonal symmetric positive definite (SPD) matrix descriptor that contains in its first block the RCD denoted by C_z, while the second block captures the correlations between the histograms computed on the three color channels of the image patch, as formally defined in Definition 1.

Definition 1. (Covariance-Kernel descriptor). The Covariance-Kernel descriptor, for an image patch $\mathbf{z}$ is defined as:

$$D_z = \begin{bmatrix} C_z + \epsilon\, I_{d1} & 0_{d1} \\ 0_{d2} & H_z H_z^T + \epsilon\, I_{d2} \end{bmatrix} \quad (3)$$

where $\epsilon > 0$ is a very small constant, d_1 and d_2 are equal to the dimensionality of C_z and $H_z H_z^T$ respectively, 0_{d1} and 0_{d2} are square zero matrices of dimension d_1 and d_2 respectively,

while I_{d1} and I_{d2} are the identity matrices of dimension d_1 and d_2 respectively.

Given that the 3 x 3 histogram correlation matrix $\boldsymbol{H_z H_z^T}+$ is positive definite, and thus a valid Mercer kernel, we further improve its representational power by computing the correlations *via* a kernel function. That is, suppose $h_c \in R^b$ denotes a histogram vector (where $\in \{R, G, B\}$), then we replace the Gram matrix $\boldsymbol{H_z H_z^T}$ in (3) by a kernel matrix $\boldsymbol{K_z}$ defined by $K\left(h_c1, h_c2\right) = \varphi\left(h_c1\right)^T \varphi(h_c2)$ for $c1, c2 \in \{R, G, B\}$ and a feature map φ. For our task, the linear kernel performed the best among the χ^2, Radial Basis Function (RBF) and polynomial kernels.

Theorem 1 (positive definiteness of the CKD). For an image patch $\boldsymbol{z}$, its corresponding CKD, $\boldsymbol{D_z}$ is an SPD matrix. That is:

$$v^T D_z v > 0, \forall v \in \mathbb{R}_d - \{0_d\} \tag{4}$$

Proof: Let $v = \left[v_C^T \quad v_H^T\right]^T$, where $v_C \in \mathbb{R}^{d1}$, $v_H \in \mathbb{R}^{d2}$ with d_1 and d_2 *cor*responding to the size of $\boldsymbol{C_z}$ *and* $\boldsymbol{H_z H_z^T}$ respectively. That way

$$\begin{aligned} v^T D_z v &= \left[v_C^T \quad v_H^T\right] \begin{bmatrix} C_z + \epsilon I_{d1} & 0_{d1} \\ 0_{d2} & H_z H_z^T + \epsilon I_{d2} \end{bmatrix} \left[v_C^T \quad v_H^T\right]^T \\ &= \boldsymbol{v_C^T}\left(\boldsymbol{C_z} + \boldsymbol{\epsilon I_{d1}}\right) \boldsymbol{v_C} + \boldsymbol{v_H^T}\left(\boldsymbol{H_z H_z^T} + \boldsymbol{\epsilon I_{d2}}\right) \boldsymbol{v_H} \end{aligned} \tag{5}$$

Since $C_z \succeq 0$ and $\boldsymbol{H_z H^T} \succeq 0$ they both become SPD *via* a small additive perturbation on their diagonal. Thus, both terms of the summation become positive, validating that $v^T D_z v > 0$.

Geometry of CKD

While the CKD already uses rich non-linearities to capture useful higher-order cues in the data, the positive definiteness structure, as shown in Theorem 1, further allows the use of non-linear geometries to significantly improve the recognition performance. That is, instead of using a Euclidean distance to measure the similarity between two SPD matrices, a non-linear measure is used which governs the geometry of the space of these matrices.

In our experiments, we adopt two such measures for efficiently computing similarities between SPD matrices, namely (i) the Log-Euclidean Riemannian metric, and the recently introduced (ii) Jensen-Bregman Logdet Divergence. Of these two, (i) also defines a Riemannian geometry to the space of SPD matrices and is a geodesic distance, while (ii) defines an information geometry-based similarity measure.

First, the Log-Euclidean Riemannian Metric (LERM) Arsigny et al. (15) is described in Equation (6) for a pair of CKDs D_i and D_j. In Riemannian geometry, the set of symmetric matrices forms a tangent space for the Riemannian manifold of SPD matrices, and the space of symmetric matrices is isomorphic to the Euclidean space. Thus, taking the matrix logarithm embeds the SPD matrices into a flat tangent space of symmetric matrices on which the usual Euclidean distance can be used for similarity computations. The Euclidean distance is:

$$LERM\left(\mathrm{D_i}, \mathrm{D_j}\right) := \left\| \mathrm{Log}\left(\mathrm{D_i}\right) - \mathrm{Log}\left(\mathrm{D_j}\right) \right\|_F \tag{6}$$

where $Log(.)$ is the matrix logarithm and $\|.\|_F$ is the Frobenius norm.

Second, the Jensen-Bregman LogDet Divergence (JBLD), first proposed by Cherian et al. (16), is also considered for similarity computations. In contrast to LERM, JBLD retains the rich non-linear geometry of the space of SPD matrices, and at the same time is computationally cheaper as the matrix logarithms are replaced by matrix determinants which can be computed efficiently *via* Cholesky factorization.

$$JBLD\left(\mathrm{D_i}, \mathrm{D_j}\right) := \left[\log\left|\frac{\mathrm{D_i} + \mathrm{D_j}}{2}\right| - \frac{1}{2}\log\left|\mathrm{D_i D_j}\right|\right]^{1/2} \tag{7}$$

where $|A|$ is the determinant of SPD matrix A.

Weakly Annotated Image Descriptor (WAID)

In an effort to broaden the recognition abilities of the CKD to larger tissue regions (and potentially whole slides) we resort to Multiple Instance Learning (MIL) (17). In the MIL setting, we only need to know if there is at least one patch that is benign or malignant in a whole slide, usually called a bag, and the MIL formulation needs to incorporate the task of inferring which instance in the bag belongs to the concerned class. Similar considerations were presented in Wang and Cherian (18) for activity recognition in a deep learning framework. Our scheme differs from the work in Wang and Cherian (19) in that WAID is computed on symmetric positive definite (SPD) matrices whose geometry is different from descriptors used in action recognition in Wang and Cherian (19). The proposed Weakly Annotated Image Descriptor (WAID) uses the MIL setup to provide annotations at the bag level thus relaxing the requisite for tissue delineations and is devised as the parameters of decision boundaries between positive bags and negative bags.

To formalize the derivation of the WAID, we let a weakly annotated image i (malignant or benign disease) be denoted by Z_i^+. Performing a random sub-sampling of m patches of size $n \times n$ for each image allows for expressing Z_i^+ as the set $\{Z_i^+[1], Z_i^+[2], \ldots, Z_i^+[m]\}$. For a bag to be characterized as positive the requirement is that at least one of the contained instances is positive which in this work translates to containing tumor tissue (benign disease or malignant). In contrast, for a bag to be negative all instances need to be negative, which is equivalent to containing neither benign diseased nor malignant patches. To achieve this, we contrast our positive bags against negative bags of background classes. In particular, we devise three strategies for populating negative bags with instances namely, (i) random noise images, (ii) images from a surrogate texture recognition dataset [Mallikarjuna et al. KTH (20)] and, (iii) patches depicting healthy regions from H&E breast tissue. In that way, we let $\tilde{Z}_j^-$ denote a negative bag, containing $\{Z_j^-[1], Z_j^-[2], \ldots, Z_j^-[M]\}$ instances derived from a background class. Prior to adopting the MIL machinery to our problem, it is required that we provide a compact description of the patches organized in bags; for this task, we employ the CKD. The CKD is a mapping from the space of image patches to that of SPD matrices as $f: \mathbb{R}^{n \times n} \longrightarrow S_{++}^d$. In that way, we express $\tilde{Z}_i^+$ and $\tilde{Z}_i^-$ as the sets $\left\{D_1^+, D_2^+, \ldots, D_m^+\right\}$ and $\{D_1^-, D_2^-, \ldots, D_m^-\}$ respectively.

The WAID is devised based on variants of the SparseMIL (21) framework, originally designed for applications which exhibit sparse positive bags (containing few positive instances); such an application is image region classification. In particular, we compute the WAID by solving an SVM objective. In that way, for every image i we identify the optimal decision boundary parametrized by $\mathbf{w_i}$ and $\mathbf{b_i}$ such that the percentage of classifiable positive instances is $\geq \eta$.

Given a positive bag $\mathcal{Z}_i^+$ and at least one negative bag $\mathcal{Z}_j^-$ we aggregate their instances in $\{D_1, D_2, \ldots, D_N\}$ along with their associated instance level labels $\{y_1, y_2, \ldots, y_N\}$ such that $y_i = +1$ if $D_i \in \mathcal{Z}_i^+$ and -1 otherwise; N here is the total number of instances in the considered bags. For all D_i's we compute their matrix logarithm [*via* the operator Log $(\cdot)$] which is equivalent to projecting the CKDs to the tangent to the cone plane which was shown to have a positive effect on similarity computations for SPD matrices (15) as described for LERM [refer to (6)].

Toward allowing for non-linear classification boundaries in the SVM model, we compute explicit feature maps $\boldsymbol{\Psi}(\cdot)$ which linearly approximate the Jensen-Shannon's homogenous kernel based on the work by Vedaldi and Zisserman (22). This allows for the computation of a linear SVM on the feature maps while encapsulating important non-linearities for separating instances belonging to the positive bag from instances in the negative bag(s). As a result, the parameters of the classification boundary are easily captured in $\mathbf{w_i}$, which for the non-linearized case becomes less trivial. Then χ^2 and the intersection kernel were also considered with the Jensen-Shannon's kernel achieving the highest performance among them. For simplifying the notation we let $\mathbf{d}_i$ denote the vector resulting from concatenating the columns of $Log(D_i)$. The classifier is, in that way, computed in a kernel Hilbert space H for which the inner product is defined as $\langle \boldsymbol{\Psi}(d_i), \boldsymbol{\Psi}(\mathbf{d_j}) \rangle_H = \boldsymbol{\Psi}(\mathbf{d_i})^T \boldsymbol{\Psi}(\mathrm{d_j})$.

With the above notation, we propose our multiple-instance max-margin WAID learning as:

$$\begin{aligned} &\min_{w_i, b_i, \xi} \|w_i\|_2^2 + C\sum_{k=1}^{N} \xi_k \quad \text{subject to} \\ y_k \quad &\left(w_i^T \Psi(d_k) + b_i\right) \geq 1 - \xi_k, \quad \forall k \in \{1, \ldots, N\} \\ &\xi_k > 0, \quad \forall k \in \{1, \ldots, N\} \\ &\frac{|y_i^+|}{|\mathcal{Z}_i^+|} \geq \eta \end{aligned} \tag{8}$$

where y_i^+ denotes the set of instances that receive a positive label by the trained ∞-SVM (23), that has three important components, namely (i) the WAID descriptor defined by the pair (w, b), (ii) the class labels y that is -1 for all instances in the negative bag, however is either $+1$ or -1 depending on whether the optimization decides the instance in the positive bag is positive or negative, and (iii) a proportionality constraint that says that we know a proportion defined by η of the positive bag has positive instances. The hyper-parameter η needs to be decided *via* cross validation or from experience.

To accommodate constraints that are difficult to cater to, we incorporate slack variables denoted by ξ_k to handle the non-separability of the samples as defined below.

$$sign\ (w\frac{T}{i}\psi\ (d_l) + b_i) = +1,\ \forall\ d_l\ \in\ y_i^+. \tag{9}$$

Even though the conventional SVM part of the formulation is convex, and thus can be solved efficiently *via* standard optimization machinery, the η-constraint makes it combinatorial. An important observation for solving this problem is the effect of the regularization parameter C on the objective; larger values of C penalize more steeply misclassified instances. Toward satisfying the η-constraint, computed on the ratio $\frac{|y_i^+|}{|\mathcal{Z}_i^+|}$, the SVM objective is iteratively solved for increasing values of C. In particular, starting with a small value for the parameter C we retrieve a solution and check if the η-constraint is satisfied based on that. In the case that the condition is not satisfied, the parameter C is rescaled to a larger value making the formulation less tolerant to misclassifications and thus steering it toward making more positive predictions. In the case that the condition is met, the SVM objective is solved for that value of C and the parameters of the classifier ($\mathrm{W_I}$ and b_I) are extracted and used to form the WAID. More formally, the WAID (**Figure 3**) for an image I is presented in Definition 2.

Definition 2 (WAID). The Weakly Annotated Image Descriptor for an image I is defined as:

$$W_I = \left| w_I^T \ \ b_I \right|^T \tag{10}$$

Once the WAID is computed for every sample in a given set of images, standard machine learning techniques are implemented toward learning based on the patterns uncovered by the descriptor.

The overall pipeline for processing the aforementioned benchmark is illustrated in **Figure 4**. First, images are sub-sampled and for the generated patches CKD descriptors are computed. Second, for every group of patches the WAID is computed then an SVM model is computed on the resulting WAID representations.

RESULTS

In this section, we present our experiments on the two databases described in the methods. First, we present a thorough evaluation of the CKD on the FABCD against a very large collection of image descriptors computed at the patch level. Following that, we evaluate the WAID on the BreakHis dataset against Multiple Instance Learning (MIL) alternatives as well as schemes that have been previously proposed for providing inference on the dataset.

FABCD

We present comparisons using SVMs, while for all the learned models we evaluate the classification performance using two different metrics, namely (i) classification accuracy (ACC), and (ii) the Area Under the Curve (AUC) computed from Receiver

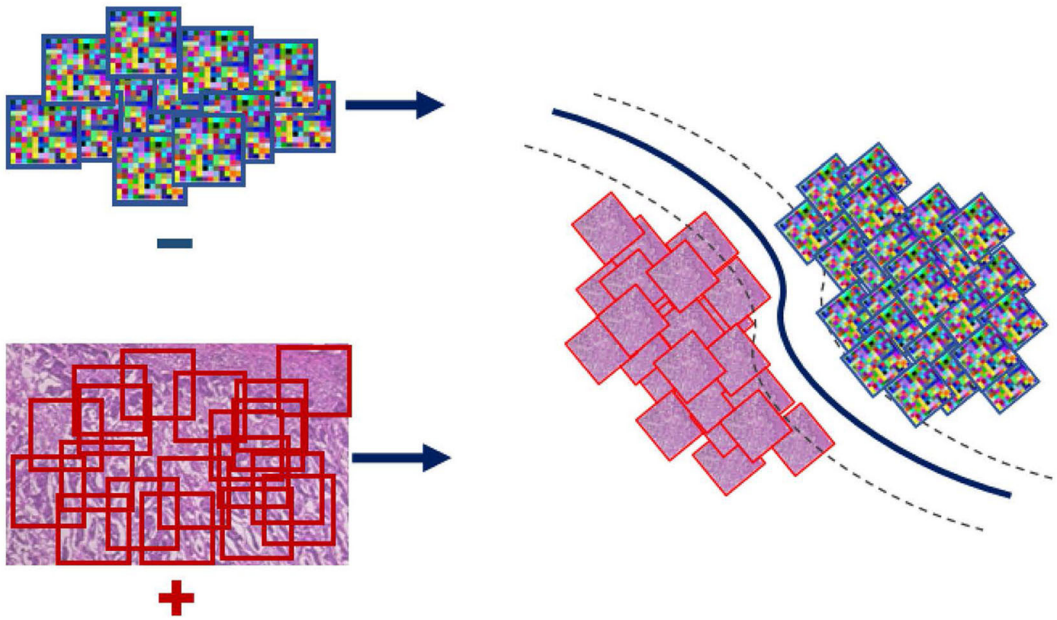

FIGURE 3 | WAID computation. Images are sub-sampled, and descriptors are computed for every derived patch. The WAID is computed as the vector containing the parameters of an SVM model computed in a multiple instance learning framework.

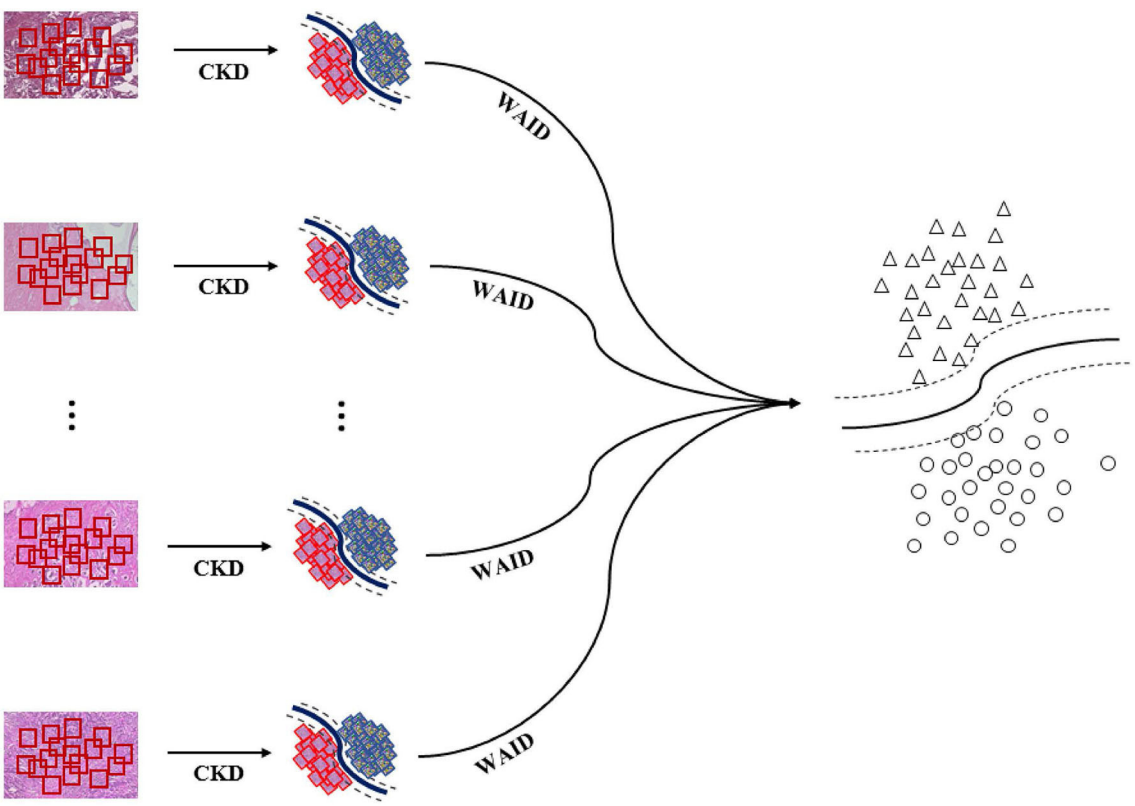

FIGURE 4 | Weakly annotated data processing. First, images are sub-sampled and CDKs are computed on every patch. Second, the WAID is computed for every group of patches. Finally, an SVM model is computed for classifying malignant images.

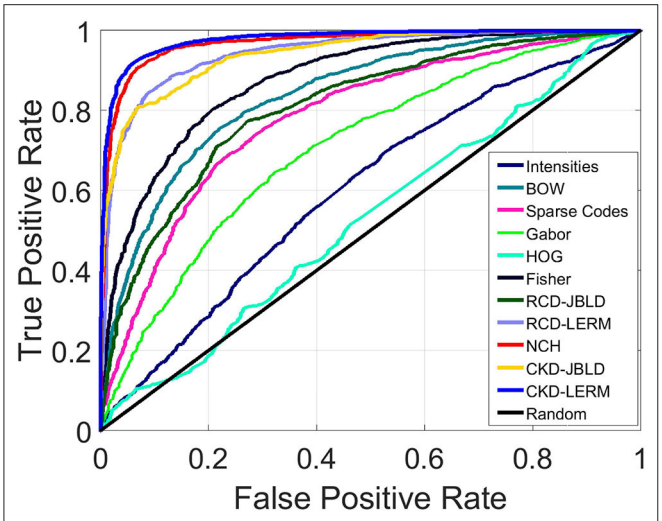

FIGURE 5 | ROC curve for intermediate level descriptors on FABCD (Best if viewed in color).

TABLE 1 | Experimental results on FABCD.

Features	ACC	AUC
Intensities	57.91%	0.60
HOG	51.86%	0.53
Gabor	65.60%	0.71
Fisher	79.66%	0.88
Sparse codes	72.31%	0.78
BOW	76.46%	0.84
RCD-JBLD	74.26%	0.81
RCD-LE	87.66%	0.94
NCH	91.63%	0.97
CKD-JBLD	85.51%	0.94
CKD-LE	**92.83%**	**0.98**

The bold values highlight the best performance.

TABLE 2 | Experimental results on FABCD against CNNs.

Features	ACC	AUC
CNN(AlexNet)	89.23%	0.96
CNN(VGG-16)	**93.91%**	**0.99**
CKD-LE	92.83%	0.98

The bold values highlight the best performance.

Operating Characteristic (ROC) curves in a 10-fold cross-validation. For RCDs and CKDs, we use Radial Basis Function (RBF) Mercer kernels based on the LERM and the JBLD measures stated in the **Supplementary Material**. For the rest of the tested descriptors, a collection of different kernels and parameter configurations were tested. In particular, the tested kernels were linear, polynomial, RBF, and Sigmoid. In **Figure 5** we can see that for almost all features represented, linear kernels achieved the highest performance and were used to report our results. The only exception is the kernel utilized for the Gabor features which is a polynomial kernel of third degree.

Figure 5 above presents the resulting ROC curves for the conducted experiments. Among edge-based descriptors, Fisher Vectors (FVs) appear to achieve the highest accuracy as well as AUC, reaching accuracy of 79.66%. The NCH IV-A outperformed all the edge-based descriptors achieving a high accuracy value of 91.63%, accompanied by very high AUC. RCDs reported accuracy that was on par with the performance of the NCHs. Finally, the CKD was seen to outperform all the considered descriptors, reaching ACC of 92.83% and AUC of 0.98. **Table 1** below aggregates the results obtained on FABCD for all the described intermediate level descriptors in terms of ACC and AUC, as computed for the extracted ROC curves.

Comparisons Against CNNs

Even though CNN based representations would require patch level annotations for their crafting, we believe that presenting comparisons against popular CNN topologies is very important. It should be noted though, that in the general weakly supervised setup patch level annotations are not necessarily available. Since we have data limited to a few thousand samples, we fine-tuned two popular CNN topologies with weights learned on the 1M image database of the ILSVRC challenge. For this study, we established a comparison against the Alexnet (24) and VGG16 (25) topologies. We compare against well-known CNN models that are often found to be generically useful for a variety of tasks. However, our experiments show that in small-data regimes, training such large topologies leads to overfitting and thus reduced performance in comparison to feature representations that are tailored to the task, as is the case with our proposed CKD descriptor.

The results of this section are delivered in the form of ACC and AUC in a 10-fold validation setup in **Table 2**. The CKD when combined with LE similarities is seen to outperform the Alexnet topology which achieved ACC of 89.23%. Finally, the VGG-16 was able to outperform the CKD achieving ACC of 93.91% and AUC of 0.99.

BreakHis

Parameter Tuning

Our experimentation indicated that working with tissue slides collected at 40x magnification level and patches of size 50×50 yielded the highest training accuracy as also illustrated in **Figure 6A**. Similarly, **Figure 6B** presents a parameter exploration with respect to the number of patches sub-sampled from the initial slide. We found that sampling 25 patches yielded the optimal recognition accuracy since it balances between training accuracy and over-fitting. Furthermore, as illustrated in **Figure 6C**, we can see that working with more than 15 negative bags did not improve the performance of the WAID. Finally, **Figure 6D** depicts the performance of the devised scheme against different values of the parameter η. We select $\eta = 0.9$.

Among the three different types of background bags (KTH, healthy tissue, and random noise), we found that working with random noise images yielded the highest accuracy. Low dimensional embeddings by der Maaten and Hinton (26) of

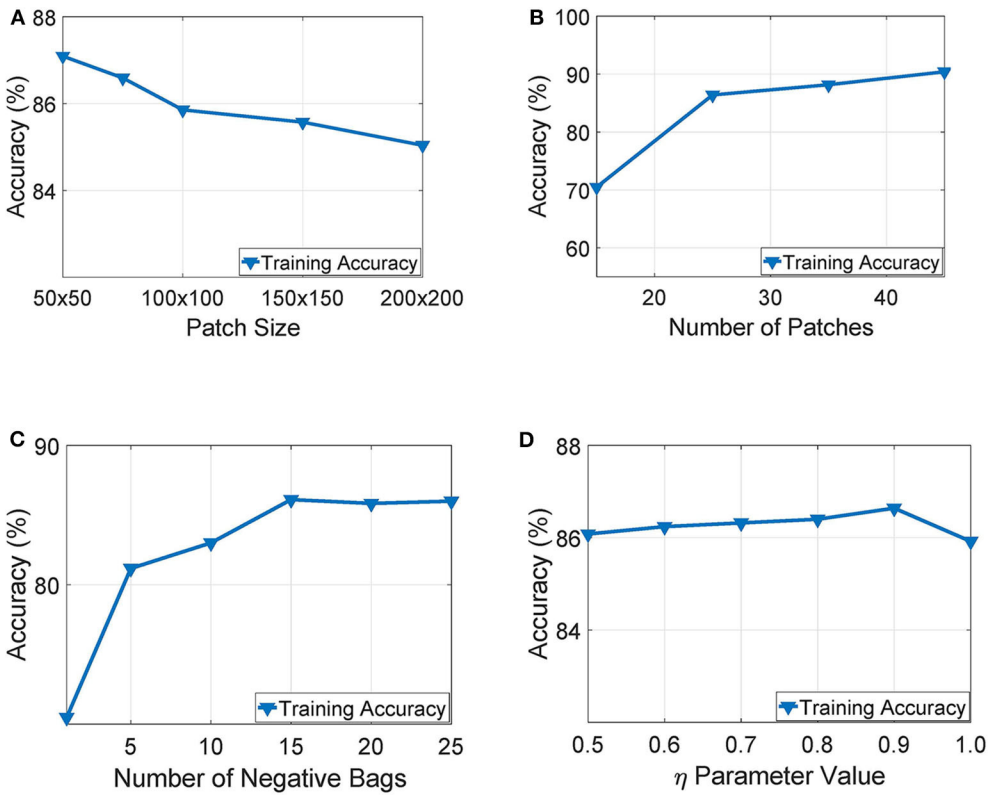

FIGURE 6 | Parameter exploration for, **(A)** patch size, **(B)** number of patches, **(C)** number of negative bags and, **(D)** η parameter.

CKDs computed on instances of the aforementioned bags (blue dots) are plotted against CKDs computed on patches of BreakHis (red dots) in **Figure 7**. For the case of healthy tissue patches, the overall performance was hindered by the risk of steering the decision boundaries around healthy samples since the positive bags also contain instances corresponding to healthy tissue deteriorating the overall performance. This can result in the inaccurate enclosure of the benign or malignant tumor instances as also suggested by **Figure 7A**. In addition, the KTH database offers a large variability in the types of contained textures resulting in a less firm cluster formation when plotted against CKDs on the histopathological data as also illustrated in **Figure 7B**. Finally, when working with random images for the background class, as presented in **Figure 7C**, it resulted in a better separation from the tissue samples which was also imprinted in our results.

Comparisons Against MIL Schemes

The comparisons of the WAID against MIL based alternatives and a baseline corresponding to computing a CKD descriptor on the whole image termed Single-CKD (S-CKD) was reported. The results are shown in terms of accuracy and the area under the curve averaged across the 5-folds provided with the benchmark. First, we considered the MIL-SVM Andrews et al. (27) scheme, the Sparse-MIL (21) was included in this set of experiments, since it takes into account the sparse distribution of positive instances in the positive bags which we set to 0.3 (an estimate of the percentage of cancerous tissue against healthy in the image). In both, we used a linear kernel. Third, in a boosting setup we present comparisons against the MIL-Boost (28) and the MCIL-Boost (6) schemes, for which we use 50 weak classifiers. The number of weak classifiers was identified *via* a trial and error process in an effort to control the amount of over-fitting of the model on the training sets. Additionally, for the MCIL-Boost (6) scheme we present results for two and three clusters in the positive bags. The number of clusters in the data was aligned with the characteristics of the dataset according to which samples contain malignant tissue surrounded by healthy tissue and potentially transition areas between the two. Finally, to further motivate the adaptation of a MIL based scheme we present results based on image descriptors computed at the whole view (S-CKD). For S-CKD and the WAID, we use an SVM model with an RBF kernel with $\gamma = 0.00025$. For the experiments involving the following schemes, namely, (i) MIL-SVM, (ii) SIL and, (iii) Sparse-MIL, we used the MISVM python module distributed in support of Doran and Ray (29). Furthermore, for the MIL-Boost and MCIL-Boost schemes we used the distribution, accompanying the work by Xu et al. (5, 6).

Summarizing the contents of **Table 3**, we see that MIL-SVM and Sparse MIL achieved the lowest performance achieving an average 0.76 and 0.70 AUC across the five computed folds. Following that, the S-CKD achieved an AUC of 0.83 underlining

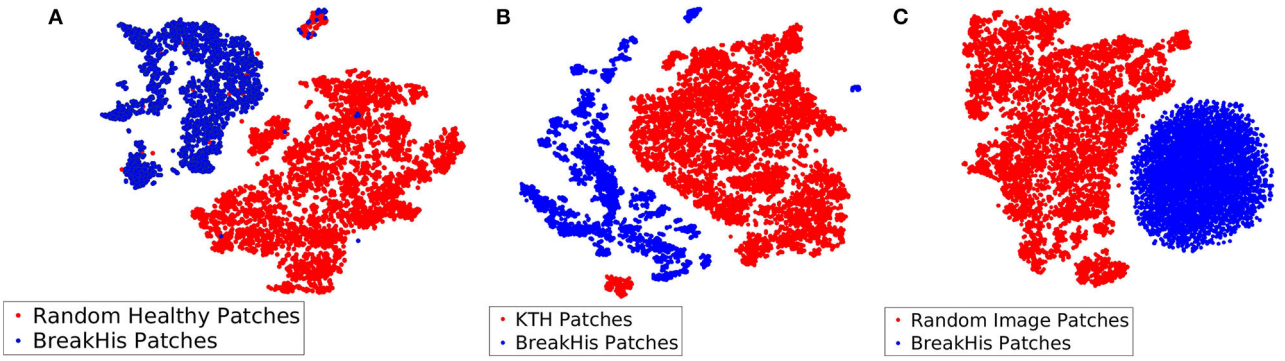

FIGURE 7 | Low dimensional embeddings of CKDs on sub-sampled patches of BreakHis images against CKDs computed on **(A)** patches of healthy tissue, **(B)** KTH patches, and **(C)** random noise images.

TABLE 3 | Comparisons against different frameworks for weakly supervised data on BreakHis.

Method	ACC		AUC
	Patient	Image	Image
MIL-SVM (27)	73.24	71.42	0.76
Sparse-MIL (21)	71.53	71.27	0.70
MIL-Boost (28)	79.54	79.68	0.87
MCIL-Boost (6) (c = 2)	80.44	79.73	0.89
MCIL-Boost (6) (c = 3)	80.14	80.13	0.89
S-CKD	77.99	77.40	0.83
WAID (KTH)	84.05	82.02	0.87
WAID (Healthy)	80.63	79.98	0.86
WAID (Random)	**85.50**	**83.57**	**0.90**

The bold values highlight the best performance.

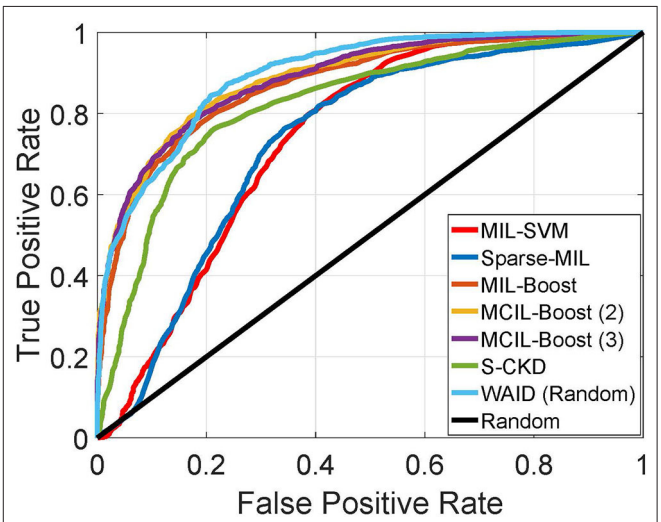

FIGURE 8 | ROC curve for experiments against MIL schemes on BreakHis (Best if viewed in color).

the necessity of the MIL paradigm. The fusion of boosting and MIL was shown to be sufficient to exceed the three aforementioned baselines, and its performance was exceeded by allowing for multiple clusters in the data through MCIL-Boost. The latter achieved an AUC value of 0.89 accompanied by ACC of 80.13 and 80.14% at the image and patient level, respectively. MCIL-Boost was only outperformed by the proposed WAID which reached an AUC of 0.90 accompanied by ACC of 83.57 and 85.50% at the patient and image level, respectively.

Comparisons Against State-of-the-Art Schemes

In this section, we establish comparisons against visual learning schemes that have been previously deployed for providing inference on the selected magnification (x40) of the BreakHis dataset. We compare against the work by Spanhol et al. (11) for which Parameter Free Threshold Adjacency Statistics (PFTAS) features (30) were computed and coupled with different classifiers namely, (i) 1-Nearest-Neighbor, (ii) Quadratic Discriminant Analysis, (iii) Random Forests, and (iv) Support vector machines. Comparisons are also established with the work by Spanhol et al. (4) which proposed a CNN based on the (24) topology. Furthermore, we present results against the work by Song et al. (31) that utilized a Fisher Vector based scheme. Finally, a CNN based scheme capitalizing on the GoogleNet topology (32) was presented by Das et al. (33). It should be noted that we are not concerned with fusion rules on the predictions of multiple images as in Spanhol et al. (4) and Das et al. (33), and we focus our evaluation on the predictions at the image and patient levels as reported in the respective studies. In the aforementioned studies, results were presented in the form of Correctly Classified Malignant (CCM) instances, at the slide level as well as the patient level. It should be noted that for the PFTAS based schemes the authors did not provide slide level performance statistics, while for Das et al. (33) the CCM at the patient level is based on majority voting in contrast to averaging as deployed in all other schemes. **Figure 8** shows the ROC curve for the experiments.

Table 4 summarizes the results obtained on the BreakHis by the WAID against recently published schemes on this dataset. The proposed framework outperforms existing approaches achieving state-of-the-art performance on BreakHis with its CCM reaching 91.27 and 92.00% at the patient and image level, respectively, without resorting to a deep learning scheme,

TABLE 4 | Comparisons against state-of-the-art on BreakHis.

Method	CCM	
	Patient	**Image**
PFTAS-1NN (11)	80.90	–
PFTAS-QDA (11)	83.60	–
PFTAS-RF (11)	81.80	–
PFTAS-SVM (11)	81.60	–
CNN-Alexnet (4)	89.60	88.60
Adaptive-Fisher (31)	87.00	90.00
CNN-Googlenet (33)	–	91.26
WAID (Random)	**91.27**	**92.00**

The bold values highlight the best performance.

thus making WAID a computationally attractive and easier to implement alternative.

DISCUSSION

In this work, we presented a framework for the analysis of histopathological breast cancer data in the presence of weak supervision. The proposed Covariance-Kernel descriptor (CKD) manages to capture higher order correlations between edges and color information (as the result of the staining process) that are very important for the recognition of malignant areas while enclosing them in a compact representation. Although the CKD successfully characterizes tissue architectures at the patch level, its performance deteriorates as the targeted slide regions increase in size. This can be attributed to the fusion of different tissue types in larger slide regions (healthy, benign disease, and malignant regions). To address this shortcoming, while leveraging the recognition capability of the CKD to larger regions of the slide (and potentially the whole slide), we derive an image descriptor in a Multiple Instance Learning (MIL) (17) framework that builds upon the CKDs. The MIL paradigm was selected due to its ability to provide inference for data organized in the form of bags (larger slide regions or whole slides) containing not individually labeled instances (patches). In pursuance of obviating the necessity for pixel level annotations, we propose the weakly annotated image descriptor (WAID) which solely requires weakly annotated samples in the form of binary labels (malignant vs. benign) and is capable of characterizing larger slide regions. Based on the results gathered from the experiments, we concluded that WAID is able to achieve state-of-the-art performance on a database that contains weakly annotated images.

As personalized medicine becomes prevalent, medical experts are faced with high demands to create automation of their most recurrent tasks and for a more complex set of analyses to be done (34). The average patient waits approximately 10 days for a pathology result, which can be critical for some patients when it comes to treatment plans as their safety and health are at risk[2]. Samples containing a large set of data require substantial effort and time from medical experts who have to manually segment the data. With these challenges, it is essential to address real-world medical challenges, solve clinical or public health problems, and recognize patients' needs (35). An automated model will allow medical diagnosis to be made at a timelier and prompter rate, thereby allowing patients to receive their results earlier which minimizes both anxiety and delayed treatments. Our model does not require an extensive amount of effort from medical experts, hence eliminating human errors. In addition, this allows medical experts to focus their time on treatment plans and patient consultations which will further improve the quality of care patients are to receive. This will not only help to improve the patient's health outcomes but also enhance the quality of health management.

Some limitations of this study are: (1) the dataset is not large enough, (2) the descriptors may not work for other cancers since they may need different weights, and (3) future studies are needed to validate the model.

CONCLUSION

In this work, we presented a framework for the analysis of histopathological breast cancer data in the presence of weak supervision. This work was concerned with the derivation of a scheme demanding less annotation effort by medical experts. We initiated our analysis with the derivation of an intermediate image representation (patch level), termed CKD, which outperformed a very large collection of popular computer vision descriptors on a private, fully supervised H&E Breast cancer dataset (FABCD). Following that, we proposed an image descriptor, termed WAID, which was derived in a MIL setup for characterizing larger image regions. WAID achieved state-of-the-art performance on the considered magnification level of the BreakHis both against MIL-based schemes as well as prior methods on the database.

Delays in diagnosing cancer is either from providers simply not consider cancer in their differential diagnosis[3] or by the waiting time of 10 or more days depending on the workload and skills of the expert to collect complicated analysis of H&E slides[2]. By implementing the derivation of a scheme that demands less annotation effort from medical experts, H&E slides can be read at a faster pace without compromising on accuracy enabling providers to determine the diagnose and treatment plan that will lessen the stress, anxiety, and unwanted burden on their patients. In regard to patients' well-being, our proposed derivation of the CKD and WAID can help medical experts accomplish their work accurately and faster.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

[2]Tests-and-procedures: biopsy. Available online at: www.cancer.net.

[3]Available online at: https://www.frontiersin.org/research-topics/13170/use-of-primary-care-datasets-for-high-risk-and-early-cancer-detection

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UMN IRB. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

PS has contributed 30% of the paper's content while the other authors have equal contributions. All authors contributed to the article and approved the submitted version.

FUNDING

Research reported in this publication was supported by the National Cancer Institute of the NIH under Award Number R01CA225435. This material was also based on work supported by the NSF through grants #CNS-1439728 and #CNS-1939033. AC was funded by the Australian Research Council Center of Excellence for Robotic Vision (project number CE140100016).

ACKNOWLEDGMENTS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2020.572671/full#supplementary-material

REFERENCES

1. Basavanhally A, Ganesan S, Feldman M, Shih N, Mies C, Tomaszewski J, et al. Multi-field-of-view framework for distinguishing tumor grade in er+ breast cancer from entire histopathology slides. *IEEE Trans Biomed Eng.* (2013) 60:2089–99. doi: 10.1109/TBME.2013.2245129
2. Irshad L, Veillard A, Racoceanu D. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. *IEEE Rev Biomed Eng.* (2014) 7:97–114. doi: 10.1109/RBME.2013.2295804
3. Sirinukunwattana K, Raza S, Tsang Y, Snead D, Cree I, Rajpoot N. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging.* (2016) 35:1196–206. doi: 10.1109/TMI.2016.2525803
4. Spanhol F, Oliveira L, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. In: *International Joint Conference on Neural Networks*. Vancouver, BC: IEEE (2016).
5. Xu Y, Zhu J, Chang E, Lai M, Tu Z. Weakly supervised histopathology cancer image segmentation and classification. *Med Image Anal.* (2014) 18:591–604. doi: 10.1016/j.media.2014.01.010
6. Xu Y, Zhu J, Chang E, Tu Z. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: *Conference on Computer Vision and Pattern Recognition*. Hubei: IEEE (2012).
7. Bejnordi B, Balkenhol M, Litjens G, Holland R, Bult P, Karssemeijer N, et al. Automated detection of dcis in whole-slide h&e stained breast histopathology images. *IEEE Trans Med Imaging.* (2016)35:2141–50. doi: 10.1109/TMI.2016.2550620
8. Dundar M, Badve S, Raykar V, Jain R, Sertel O, Gurcan M. A multiple instance learning approach toward optimal classification of pathology slides. In: *International Conference on Pattern Recognition*. Cape Town: IEEE (2010).
9. Stanitsas P, Cherian A, Li X, Truskinovsky A, Morellas V, Papanikolopoulos N. Evaluation of feature descriptors for cancerous tissue recognition. In: *2016 23rd International Conference on Pattern Recognition (ICPR)* Daejeon: IEEE (2016).
10. Fischer A, Jacobson K, Rose J, Zeller R. Hematoxylin and eosin staining of tissue and cell sections. *CSH Protocols.* (2008) 2008:pdb.prot4986. doi: 10.1101/pdb.prot4986
11. Spanhol F, Oliveira L, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. *IEEE Trans Biomed Eng.* (2016) 63:1455–62. doi: 10.1109/TBME.2015.2496264
12. Porikli F, Tuzel O, Meer P. Covariance tracking using model update based on lie algebra. In: *International Conference on Computer Vision and Pattern Recognition*. Istanbul: IEEE (2006).
13. Tuzel O, Porikli F, Meer P. Region covariance: a fast descriptor for detection and classification. In: *European Conference in Computer Vision*. Graz (2006).
14. Maji S, Berg AC, Malik J. Classification using intersection kernel support vector machines is efficient. In: *Computer Vision and Pattern Recognition*. Beijing: IEEE (2008).
15. Arsigny V, Fillard P, Pennec X, Ayache N. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magn Reson Med.* (2006) 56:411–21. doi: 10.1002/mrm.20965
16. Cherian A, Sra S, Banerjee A, Papanikolopoulos N. Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE Trans Pattern Anal Mach Intell.* (2013) 35:2161–74. doi: 10.1109/TPAMI.2012.259
17. Dietterich T, Lathrop R, Lozano-Perez T. Solving the multiple instance problem with axis-parallel rectangles. *Artif Intell.* (1997) 89:31–71. doi: 10.1016/S0004-3702(96)00034-3
18. Wang J, Cherian A, Porikli F, Gould S. Action representation using 13 classifier decision boundaries. In: *CVPR*. Salt Lake City, UT (2018).
19. Wang J, Cherian A. Learning discriminative video representations using adversarial perturbations. In: *ECCV*. Munich (2018).
20. Mallikarjuna P, Targhi AT, Fritz M, Hayman E, Caputo B, Eklundh J-O. *THE KTH-TIPS 2 Database* (2006).
21. Bunescu R, Mooney R. Multiple instance learning for sparse positive bags. In: *International Conference on Machine Learning*. Corvalis, OR: ACM (2007).
22. Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps. *IEEE Trans Pattern Anal Mac Intell.* (2012) 34:480–92. doi: 10.1109/TPAMI.2011.153
23. Yu FX, Liu D, Kumar S, Jebara T, Chang SF. ∞-SVM for learning with label proportions. In: *ICML'13: Proceedings of the 30th International Conference on International Conference on Machine Learning*. Vol. 28 (2013). p. 504–12.
24. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. Lake Tahoe, NV (2012).
25. Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. CoRR (2015).
26. der Maaten LV, Hinton G. Visualizing data using t-sne. *J Mach Learning Res.* (2008) 9:85. Available online at: http://jmlr.org/papers/v9/vandermaaten08a.html
27. Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning. *Adv Neural Inf Process Sys.* (2003) 15:577–84.
28. Viola P, Platt J, Zhang C. Multiple instance boosting for object detection. In: *Conference on Neural Information Processing Systems*. Vancouver, BC (2005).

29. Doran G, Ray S. A theoretical and empirical analysis of support vector machine methods for multiple-instance classification. *Mach Learn.* (2014) 97:79–102. doi: 10.1007/s10994-013-5429-5
30. Hamilton N, Pantelic R, Hanson K, Teasdale R. Fast automated cell phenotype image classification. *BMC Bioinformatics.* (2007) 8:110. doi: 10.1186/1471-2105-8-110
31. Song Y, Zou J, Chang H, Cai W. Adapting fisher vectors for histopathology image classification. In: *International Symposium on Biomedical Imaging*, VIC: IEEE (2017).
32. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Conference on Computer Vision and Pattern Recognition.* Santa Clara, CA: IEEE (2015).
33. Das K, Karri S, Roy A, Chatterjee J, Sheet, D. Classifying histopathology whole-slides using fusion of decisions from deep convolutional network on a collection of random multi-views at multi-magnification. In: *International Symposium on Biomedical Imaging, IEEE.* (2017).
34. Van Eycke Y-R, Foucart A, Decaestecker C. Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images. *Front Med.* (2019) 6:222. doi: 10.3389/fmed.2019.00222
35. Kostkova P. Grand challenges in digital health. *Front Public Health.* (2015) 3:134. doi: 10.3389/fpubh.2015.00134

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 20 January 2021
doi: 10.3389/fdgth.2020.566595

Wearable Based Calibration of Contactless In-home Motion Sensors for Physical Activity Monitoring in Community-Dwelling Older Adults

*Narayan Schütz[1], Hugo Saner[1,2], Angela Botros[1], Philipp Buluschek[3], Prabitha Urwyler[1,4], René M. Müri[1,4] and Tobias Nef[1]**

[1] ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland, [2] Sechenov First Moscow State Medical University, Moscow, Russia, [3] DomoSafety S.A., Lausanne, Switzerland, [4] Department of Neurology, University Neurorehabilitation Unit, University Hospital Bern (Inselspital), University of Bern, Bern, Switzerland

Passive infrared motion sensors are commonly used in telemonitoring applications to monitor older community-dwelling adults at risk. One possible use case is quantification of in-home physical activity, a key factor and potential digital biomarker for healthy and independent aging. A major disadvantage of passive infrared sensors is their lack of performance and comparability in physical activity quantification. In this work, we calibrate passive infrared motion sensors for in-home physical activity quantification with simultaneously acquired data from wearable accelerometers and use the data to find a suitable correlation between in-home and out-of-home physical activity. We use data from 20 community-dwelling older adults that were simultaneously provided with wireless passive infrared motion sensors in their homes, and a wearable accelerometer for at least 60 days. We applied multiple calibration algorithms and evaluated results based on several statistical and clinical metrics. We found that using even relatively small amounts of wearable based ground-truth data over 7–14 days, passive infrared based wireless sensor systems can be calibrated to give largely better estimates of older adults' daily physical activity. This increase in performance translates directly to stronger correlations of measured physical activity levels with a variety of age relevant health indicators and outcomes known to be associated with physical activity.

Keywords: sensor calibration, pervasive computing, passive infrared, physical activity, older adults, outing imputation, ambient assisted living, telemonitoring

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Iraklis Paraskakis, South East European Research Center, Greece
Parisis Gallos, National and Kapodistrian University of Athens, Greece

***Correspondence:**
Tobias Nef
tobias.nef@artorg.unibe.ch

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *28 May 2020*
Accepted: *03 September 2020*
Published: *20 January 2021*

Citation:
Schütz N, Saner H, Botros A, Buluschek P, Urwyler P, Müri RM and Nef T (2021) Wearable Based Calibration of Contactless In-home Motion Sensors for Physical Activity Monitoring in Community-Dwelling Older Adults. Front. Digit. Health 2:566595. doi: 10.3389/fdgth.2020.566595

INTRODUCTION

Population aging poses unprecedented global challenges to modern health care systems, economies and last but not least, society as a whole (1, 2). Modern information and communication technology has the potential to contribute in overcoming some of these challenges (3–5). This includes the use of pervasive computing technology, such as microprocessor enhanced objects of everyday life. Small sensing devices like smartwatches or smart home appliances may be used to provide continuous remote monitoring of relevant health indicators and outcomes (4), increasingly referred to as digital biomarkers (6–8). These may allow for early detection of health deteriorations, enabling for instance better preventive measures or earlier interventions (9, 10). Additionally, monitoring

of relevant digital biomarkers by means of pervasive computing technologies could allow for continuous assessments of chronic conditions and help in evaluating intervention efficacy (9, 11).

Physical activity (PA) is associated with a wide range of health benefits, including lower rates of all-cause mortality, non-communicable diseases, cardiorespiratory and muscular fitness across all age groups. Regular PA also helps to protect against frailty, sarcopenia, and cognitive decline (12–14). Wearable technologies, known as wearables, that can track individual's PA behavior are popular consumer items with a worldwide distribution, particularly in younger and middle-aged populations. Also, wearable accelerometers are a well-accepted method to objectively measure PA in everyday life (15–17).

While wearable devices like smartwatches, smartphones or fitness trackers would be ideal to track a variety of health relevant markers like physical activity, post-implementation based experience, including our own, point toward a clear preference for unobtrusive contactless sensing devices (9). Reasons for that may include a certain social stigma associated with visibly wearing devices amongst peers (18, 19), difficulty in handling them, added discomfort of having to think about charging and wearing a device (20), as well as skin irritations related to long-term biosensor wear (intensified by sweat in summer). While some of the mentioned issues are related to the perception of the current generation of older adults toward technology, handling wearable devices that need regular maintenance, can also be problematic for older adults with motor, cognitive, and especially memory related, issues. However, the alternative, wireless ambient sensors, are oftentimes either less accurate (for instance infrared sensors or bed motion sensors) or overly intrusive (for instance video or audio-based recording devices).

The use of wearable devices for initial calibration of less accurate but unobtrusive ambient sensors for PA quantification is a novel approach that could minimize the burden of wearing a device, while improving the reliability and thus usefulness of unobtrusive ambient sensors for physical activity tracking significantly. A similar strategy was employed with passive infrared (PIR) sensor based gait-speed estimation, where calibration was performed using a sensor array as ground-truth, but as the authors state, another source, such as a wearable device, could have been used (21).

PIR motion sensors are rather inexpensive, contactless, and unobtrusive. Therefore they are commonly used in long-term in-home monitoring settings with older adults (9, 11, 22–26). We have previously shown that in-home physical activity, quantified by PIR motion sensors can be used to approximate physical activity in old and oldest-old community-dwelling adults (26). However, the PIR motion sensor-based approach has two main disadvantages: (1) baseline activity comparisons of absolute values between participants are difficult if apartments and sensor placements differ and (2) it is unclear how to address outings correctly. We aim to address both problems by using the much more accurate and well-validated accelerometer based physical activity, to initially calibrate the ambient sensor systems.

METHODS

Participants

The data used for this work stems from a study where modern pervasive computing systems were evaluated for telemonitoring in older adults (26). Participants were part of the StrongAge cohort in Olten (Switzerland) (27) and should represent a naturalistic population sample of community-dwelling, alone-living, old and oldest-old adults in Switzerland. We included all participants that had at least 60 days of wearable activity data recorded (first 30 days reserved for calibration and $\geq$30 days for evaluation) in this analysis, totalling 20 participants (age = 88 $\pm$ 8 years). The 60 days were chosen to include as many participants in the dataset as possible while guaranteeing a minimal number of data points.

The original study was conducted based on principles defined in the Declaration of Helsinki and approved by the Ethics Committee of the canton of Bern, Switzerland (KEK-ID: 2016-00406). All subjects signed and handed in an informed consent before study participation.

Pervasive Computing Systems

In this work we made use of the DomoCare® home monitoring system for older adults (DomoSafety S.A., Lausanne, Switzerland), the same as in (26). The system consists of PIR motion sensors (sampling at 0.5 Hz) placed in the participant's apartment. Kitchen, toilet, living-room, entrance, and bedroom were always equipped with at least one sensor, if a separate bathroom was present it was equipped with a sensor as well. In addition, a magnetic door sensor was placed on the entrance and fridge door, respectively. All sensing units communicate via the ZigBee protocol with a base unit, that then sends data to a secure cloud in real-time. The PIR system allows motion detection in individual rooms based on changes in infrared radiation caused by human activity (28). The door sensors allow outings to be calculated based on entrance door opening and closing events, as explained in (29).

For the calibration of the PIR sensor system we used the medical grade Everion® biosensor worn on the upper arm (Biovotion AG, Zürich, Switzerland). Amongst other sensors, the device contains a 3-axis accelerometer that samples at 50 Hz and outputs/stores aggregated and standardized activity (vector magnitude) at 1 Hz. The participants wore the device throughout the daytime and put it on an inductive charger overnight. While charging the device, data was transmitted to a smartphone via Bluetooth Low Energy which was then encrypted and automatically transferred to a secure cloud. Data from DomoCare® systems was first stored on cloud instances from DomoSafety S.A. located in Switzerland and data from the Everion® was initially stored on instances at the University of Bern. Post collection, all data was subsequently transferred to local servers and ingested into an OmniSci (OmniSci, San Francisco, CA, United States) analytics database instance after quality control. A schematic including the data structure is available in the **Appendix**. To initially ensure accelerometer validity, we compared values from the Everion® with the widely used and validated (30) Axivity AX3 (Axivity

Ltd., Newcastle, UK), 3-axis accelerometer [calibrated to local gravity and temperature, as described in (31)] and found good overall agreement.

Problem Definition

There are three major limitations related to the use of PIR sensors for PA quantification: (1) Motion measured by the commonly used simple PIR motion sensors is converted to a binary response, zero if there was no change in infrared radiation above the sensor's sensitivity threshold and one otherwise. It is thus apparent, that simple PIR motion sensors cannot differentiate between the intensity of the motion, unlike a body attached accelerometer; (2) the angle and distance to a sensor can influence if and how long motion is being detected; (3) the size of equipped rooms and the apartment layout in general can lead to different results for the same amount of physical activity exerted by a person. As a result, even if the same person performed the exact same finite set of activities $A = \{a_1, \ldots, a_n\}$ in different PIR motion sensor equipped apartments, measures of these activities between the PIR motion sensor measurement functions $f_{PIR} : A \rightarrow M_{PIR}$; $M_{PIR} \in \mathbb{R}_+$ and the accelerometer $f_{acc} : A \rightarrow M_{acc}$, $M_{acc} \in \mathbb{R}_+$ would likely differ widely. Now in reality, this simplification is not exactly true, because certain activities a_i will be measured by the accelerometer but not by the PIR sensors—for instance when a person is outside the apartment, outside the field of view of the PIR sensors or in a non-equipped room. This gives rise to a subset of all measured activities $\tilde{A} \subseteq A = \left\{ a_i \mid a_i \in A,\ a_i \in dom(f_{PIR}) \right\}$. We will henceforth refer to f_{acc} that is only defined over this subset as $\tilde{f}_{acc} : \tilde{A} \rightarrow \tilde{M}_{acc}$.

The idea of initial calibration is then to find a mapping $\hat{f}_{PIR} : \tilde{A} \rightarrow \hat{M}_{PIR}$, such that for a given activity a_i, the Euclidean distance between the calibrated PIR motion measurement function $\hat{f}_{PIR}$ and the domain restricted accelerometer measurement function $\tilde{f}_{acc}$ is minimized, which can be thought of as a classic regression objective:

$$\min \sqrt{\left(\hat{f}_{PIR}\left(a_i\right) - \tilde{f}_{acc}\left(a_i\right)\right)^2} \ \forall\, i = 1 \ldots n$$

Well-calibrated $\hat{f}_{PIR}$ functions from different PIR sensor equipped apartments should then allow that somewhat similar results are obtained for a given activity, since f_{acc}, given a certain activity a_i should be similar across apartments. This assumption is only true if the difference between $\tilde{f}_{acc}$ and f_{acc} is not too large and the accelerometer intensity measurements between participants is mostly comparable. The latter assumption is likely true as for instance described in (32), while the former is largely apartment and person specific but may be improved upon by including an estimate for activity while outside.

Learning Calibration Function

To find a suitable function $\hat{f}_{PIR}$, or in this case a distribution over $\hat{f}_{PIR}$ we propose to use Gaussian process regression, such that $\hat{f}_{PIR} \sim GP\left(\mu, k\right)$, where $\mu\left(A\right) = 0$ is the standardized activity mean and $\mathrm{k}\left(\mathrm{A}, \mathrm{A}'\right)$ the activity covariance function. Gaussian process regression (GPR) provides various characteristics that are likely useful in our calibration scenario. First, it allows non-linear relationships to be modeled and is non-parametric (33). In addition, GPR is known to work well with relatively little data and allows a predictive distribution to be obtained, which can help in detecting model uncertainty (33, 34). The included epistemic model uncertainty could be helpful post calibration as it could allow for quantification when patterns not seen during calibration occur, and give respective warnings if total uncertainty increases.

In the shown experiments we ended up using $k\left(a_i, a_j\right) = \sigma_0^2 + a_i \cdot a_j + \sigma_n^2 \delta_{ij}$, as kernel defining the covariance function, where δ_{ij} is a Kronecker delta, σ_n^2 is a learnable bias term and σ_0^2 is a learnable noise constant representing additional homogenous aleatoric uncertainty in the activity measurements (33). To give a comparison how other, more traditional algorithms might perform, we additionally evaluated calibration performance with a regular linear regression (LR) algorithm and the popular XGBoost (XGB) implementation (35) of a tree boosting algorithm. The GPR kernel and hyperparameters for the other algorithms were selected by means of 3-fold cross-validation (splitting at the participant level) and random search (36). It is rather difficult to assess the usefulness of the predictive distribution, obtained by the marginal normal of the GP, in a realistic manner. We try to quantify its utility by calculating the linear correlation between the daily average MAE_p^d (see below) and the daily average uncertainty estimate (the σ of the marginal Gaussian distribution).

Data Pre-processing and Representation

We represent individual activities a_i as activity bouts/islands and describe their characteristics in vector space. The activity islands are extracted by first applying a simple moving average low pass filter, with 1-min length, to the total PIR motion activity signal (the sum of the duration where all PIR motion sensors in an apartment were active) and then extracting the activity islands (stretches where low-pass filtered activity is constantly > 0). Based on these islands we calculate the following features that can be used to summarize the islands in vector space: the total duration of the island, the hour of the day, the duration PIR sensors detected activity for each equipped room and the relative activity of each room with respect to the total island duration. Corresponding activity from the wearable accelerometer was also extracted and summed over the island duration, giving us the target activity f_{acc}. The feature matrix was standardized to have zero mean and unit variance across column.

Statistical Evaluation Metrics

First, it should be noted that evaluation was always performed on all available data beyond the initial 30 days that were reserved for calibration. Throughout this work we refer to multiple evaluation metrics that are explained here. First, the mean absolute error MAE_p between the activity estimate for each activity island and its corresponding accelerometer activity was calculated for each participant p. MAE_p^d refers to the same but averaged over a day d. Second, to measure the proportionality of calibration, the Pearson correlation coefficients ρ_p between the sum of estimated post calibration activity $\hat{F}_{PIR}^d = \sum_{a_i \in d} \hat{f}_{PIR}\left(a_i\right)$ per day d and the

sum of total accelerometer activity $F^d_{acc} = \sum_{a_i \in d} f_{acc}(a_i)$ per day d were calculated. Similarly to ρ_p, we calculate $\tilde{\rho}_p$ the Pearson correlation coefficient between the sum of daily calibration activity $\hat{F}^d_{PIR}$ and domain restricted sum of accelerometer activity $\tilde{F}^d_{acc} = \sum_{a_i \in d, a_i \in \tilde{A}} \tilde{f}_{acc}(a_i)$. For all metrics, the sample average over all participants can be calculated, resulting in the global *MAE*, ρ and, $\tilde{\rho}$.

Determining the Amount of Wearable Ground-Truth Data

To assess the relationship between wear-time and calibration performance, we performed calibration with 1 day, 7 days, 14 days 21 days, and 30 days of accelerometer data and calculated $\tilde{\rho}$ and *MAE* for each wear-time point and each learning algorithm (as results may be algorithm dependent).

Evaluation of Post-calibration Performance Evolution

One of the main concerns about this kind of calibration procedure is the potential degradation that calibration quality could be subjected to over time as a result of a shift in the data generating distribution (e.g., as a result of changing behavior or seasonal patterns). To assess the potential for degradation, we calculated the weekly average *MAE* for all participants over thirty consecutive weeks (if data was available). To ensure similar scales, we first standardize weekly averages by removing the median and scaled data with respect to the interquartile range. Finally, for each week, the global average was taken, and a regression line estimated. The *p*-value of the slope coefficient was then used to determine whether the parameter differed significantly (based on $\alpha = 0.05$) from 0, allowing one to decide whether any relevant trend might be present.

Effect of Calibration on Correlations With Clinical Assessments

To evaluate how calibration influences overall relationships with health indicators and outcomes, we calculate the non-parametric Spearman's rank correlation coefficients r between median daily total activity and the mean of the respective clinical assessments (if multiple were taken per participant over the same duration). Clinical assessments include: the fall-risk focused Timed Up and Go (TUG) (37), the balance and gait focused Tinetti Performance Oriented Mobility Assessments (POMA-b and POMA-g) (38), the late life depression focused Geriatric Depression Scale (GDS) (39), the cognition focused Montreal Cognitive Assessment (40), the frailty focused Edmonton Frail Scale (EFS) (41) as well as muscle force focused handgrip, hip flexor and knee extensor strength. To assess whether there are statistical differences between pre- and post-calibration, we apply the non-parametric Wilcoxon signed rank test to the absolute correlation values under the alternative hypothesis that post-calibration values are on average greater compared to pre-calibration values.

Time Spent Outside

To assess overall PA, time spent outside the home needs to be considered. In our case, using PIR in-home sensors, PA outside the home can be seen as blocks of missing data.

A common strategy to deal with missing data is called imputation, which refers to replacing missing data with substitutes (for instance a variable's mean over all observed values) (42). Imputation can often work reasonably well, if the data is "missing completely at random" or "missing at random" (42). Given that outing likely involves more physical activity than being inside, it may be impossible to correctly impute physical activity of outing periods. Fortunately, access to calibration data from a wearable (given the wearable is also worn outside, which is true in our case), allows us to estimate a factor τ_p (for each participant p) by which the expected inside activity should be multiplied with. To calculate this factor, we first calculate outings according to (29) and then for each outing we divide the physical activity measured by the accelerometer with the average activity of the accelerometer during the same time of day, when the person was at home. Eventually, the median of these ratios gives us τ_p. A global factor τ can then be calculated by averaging over all individual participant's τ_p. As we are dealing with missing time blocks, we use temporal means—similar to what has been used for imputing non-wear time intervals with accelerometers (17, 43). That is, the expected activity sum for the given time-interval (when the outing occurred) over all observed days. To evaluate the effect of this imputation procedure on overall calibration, the evaluation metric $\tilde{\rho}$ is calculated using (1) temporal mean imputation, (2) temporal mean imputation with factor τ_p and (3) temporal mean imputation with factor τ.

All data processing, analyses and plotting have been performed with the Python (Python Software Foundation) scripting language (version 3.7). For the LR and GPR algorithm implementations from Scikit-learn library (44) were used. In case of the XGB algorithm, the official Python implementation was used.

RESULTS

Calibration Results With Differing Amounts of Data and Learning Algorithms

In **Figure 1**, we visualized evaluation metric $\tilde{\rho}$ and *MAE* for 1, 7, 14, 21, and 30 days of calibration data in combination with the proposed GPR based calibration as well as LR and XGB based calibration. It should be noted, that for both evaluation metrics, the largest increase in performance can be seen between one and seven days of calibration data (from wearable accelerometer). Beyond 14 days, more data leads to increasingly smaller improvements. In case of the correlation coefficient ρ performance saturation seems to be reached by 21 days, while in case of the *MAE* saturation is not completely evident, even after 30 days. In terms of the learning algorithms used to approximate the activity calibration function $\hat{f}_{PIR}$, it is visible how GPR shows the best performance with little data up to 14 days. After that, GPR is mostly on par with LR and starts losing in comparison to XGB. Correlation values ρ show an average of 0.84 after 14 days. Note that all results displayed downstream were based on the 14 days calibration data case.

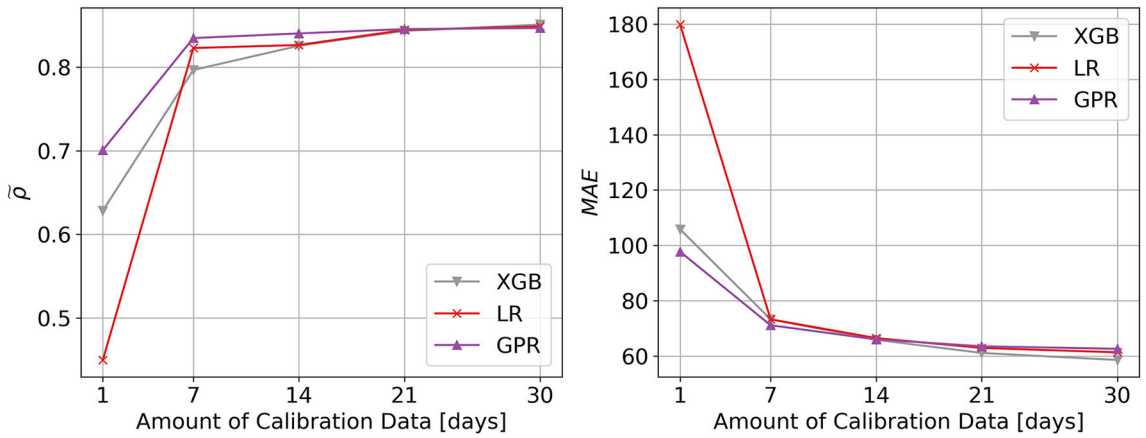

FIGURE 1 | Visualization of data and algorithm dependent calibration performance. Performance of algorithms used for calibration of passive infrared sensor systems, with respect to physical activity measured by wearable accelerometers. The learning curves show the performance across all 20 participants of the calibration method against the number of days of accelerometer reference data. The different line colors show different learning algorithms used for calibration (LR, Linear Regression; XGB, XGBoost; GPR, Gaussian Process Regression). The left plot shows the Pearson correlation coefficient $\tilde{\rho}$ as evaluation criterion (higher is better), while the right plot shows the mean absolute error (*MAE*) evaluation criterion (lower is better). With GPR, only 7–14 days of reference accelerometer data is necessary to obtain a mapping quality, which can only be marginally improved upon with additional data.

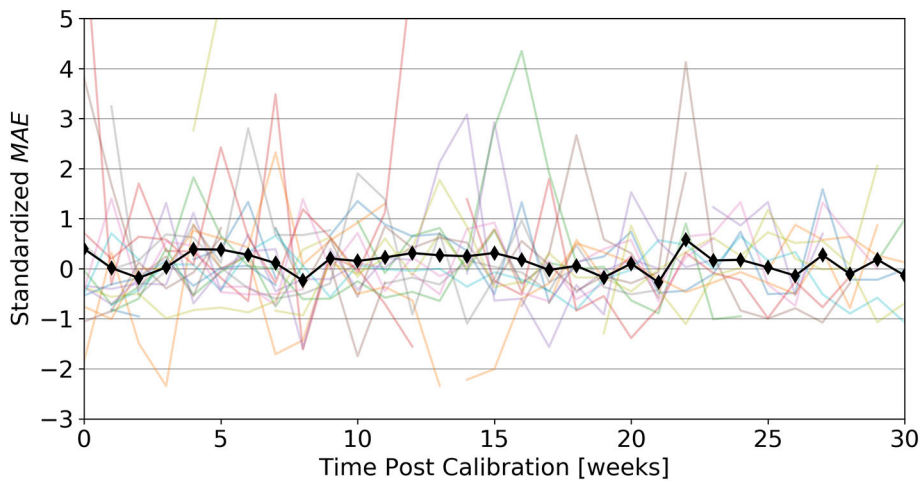

FIGURE 2 | Post calibration performance evolution. shows evolution of average calibration performance, up to 30 weeks post calibration. Individual colored lines represent standardized $\boldsymbol{MAE}_p^d$ for a given week of each participant, while the black line represents the standardized ***MAE*** over all participants for a given week We observe that even up to 30 weeks (∼7 months) post-calibration an initial calibration using 14 days of accelerometer data remains valid.

Post-calibration Performance Evolution

Visually, it is difficult to discern any sort of overall deterioration throughout 30 weeks post calibration, beyond some short-term variation (see **Figure 2**). Regression analysis of *MAE* against time, further reveals that the slope is not statistically significant ($p = 0.262$).

Impact of Calibration on Age Relevant Health Indicators and Outcomes

Results showing correlations of clinical assessments using calibrated and uncalibrated activity from the ambient sensor system as well as the accelerometer, demonstrate how calibration leads to increases in correlation for all assessments except hip extensor strength. Oftentimes post-calibration correlations reach strengths close to the accelerometer gold standard (see **Table 1**). Results based on the Wilcoxon signed-rank test, additionally suggest that the differences in correlations between pre- and post-calibration are statistically significant ($n = 8$, $p = 0.004$).

Handling Outings

We found that for most participants time spent outside the house leads to more activity compared to the average of the same time they spent at home. On average the ratio of activity outside vs. inside was found to be 1.38. However, depending on the person

TABLE 1 | Participant characteristics and demographics.

	Pre-calibration	Post-calibration	Accelerometer
TUG	−0.42	−0.57	−0.54
POMA–b	0.51	0.66	0.66
POMA-g	0.49	0.56	0.65
GDS	−0.43	−0.61	−0.64
MoCA	0.68	0.85	0.80
EFS	−0.56	−0.64	−0.65
Handgrip right	0.39	0.49	0.48
Hip right	0.37	0.36	0.34

Shows Spearman's rank correlation coefficient r_s between clinical assessment results and median daily activity sum of the ambient system, pre- and post-calibration, as well as from the worn accelerometer sensor.

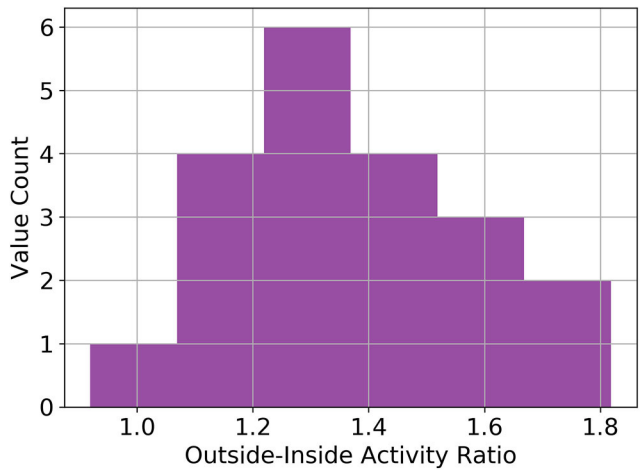

FIGURE 3 | Distribution of inside-outside activity ratios. Histogram of the ratio between time spent outside and inside the home. The average value is 1.38 across all included participants. These values are based on data from a wearable accelerometer sensor.

this ratio can be quite a bit different, ranging from 0.92 up to 1.82. The distribution is visualized as a histogram shown in **Figure 3**.

We further found that by temporal mean imputing, ρ (the correlation to overall daily accelerometer activity) increases in most cases. Regarding the type of temporal mean imputation, using no coefficient seems to lead to significantly lower correlation values compared to using a person specific coefficient ($p = 0.0007$) or a static coefficient value ($p = 0.0007$), between which no significant difference ($p = 0.8$) was found (see **Figure 4**).

Predictive Distribution

To evaluate the potential usefulness of a predictive distribution we assessed how well it correlates with the daily MAE_p^d for each participant. The median correlation coefficient across all participants was 0.49 ± 0.15 (min $= 0.1$, max $= 0.67$). An example of a decent correlation is given in **Figure 5**.

DISCUSSION

We found that using even relatively small amounts of wearable based ground-truth data, PIR based wireless sensor systems can be calibrated to considerably improve estimates of older adults' daily physical activity. We could additionally verify, that this increase in performance directly translates to stronger correlations of the measured physical activity levels with a variety of age relevant health indicators and outcomes, known to be associated with physical activity. This indicates that the performance gained by calibration is not only present on paper but also manifests itself in physical activity readings that capture relations to health significantly better than would be the case without calibration.

Deciding on the necessary amount of wearable data, sufficient for calibration, is a rather subjective and task specific matter, as it is a trade-off between calibration performance and wear-time. In our case, calibrating a PIR motion sensor system, 7–14 days seem to give reasonable results, with diminishing additional benefit employing longer calibration periods. We also observed that the optimal type of algorithm, approximating the calibration function $\hat{f}_{PIR}$, seems dependent on the amount of available calibration data. For small amounts of calibration data, GPR may be considered the best choice—which is a known property of GP based approaches (45). As a side note, in our case a linear kernel proved to be the best parametrization, which would be equivalent to using a Bayesian linear regression algorithm, but the GP view might still be more effective given little data (46). This also explains why the GPR results largely converge to the LR results given more data. On the other hand, the XGB algorithm leads to slightly better performance, given more than 14 days of calibration data, which would be the expected behavior for an algorithm with much more learning capacity. Now, since we want to restrict the necessary wear-time to a minimum, GPR is, as we initially assumed, a suitable algorithm for the task. An additional benefit of GPR's Bayesian nature, is the included predictive uncertainty, which we think can be quite useful as it often indicates a simultaneous increase in model error and may thus be used to diagnose when a calibration model's performance degrades. For our data (see **Figure 5**), however, we found no significant degradation in calibration performance up to 30 weeks post calibration, indicating that calibration is overall relatively stable and resilient toward smaller potential perturbations.

It comes as no surprise that it is important to somehow factor in the time spent outside, else, physical activity of people spending a lot of time outside would be vastly underestimated. The question, as how to best deal with outings in this scenario does however remain open and we did not find any work assessing this in community-dwelling old and oldest old adults. Our findings suggest that just replacing time spent outside with the average activity throughout a given time-interval is a valid strategy, leading to significant calibration improvements but does in most cases underestimate physical activity as old and oldest-old adults tend to be more physically active when outside. We found our participant population to be, on average, 1.38 times as physically active when outside, compared to if they were inside at the same time of the day (see **Figure 3**). Using this

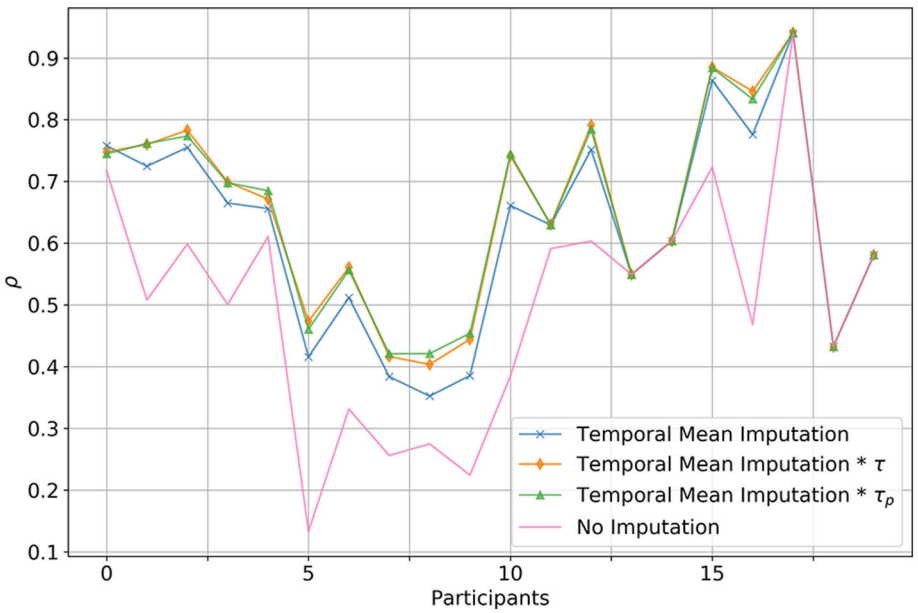

FIGURE 4 | Comparison between multiple imputation strategies to handle outing. Displayed is the correlation between the total daily calibrated activity and the total accelerometer activity. In the case of the blue line, simple temporal imputation has been used to substitute missing physical activity due to outings. The orange line denotes the case where in addition to temporal mean imputation a global correction factor was added, whereas with the green line a person specific correction factor was used. Red denotes the baseline, where outings where not imputed at all.

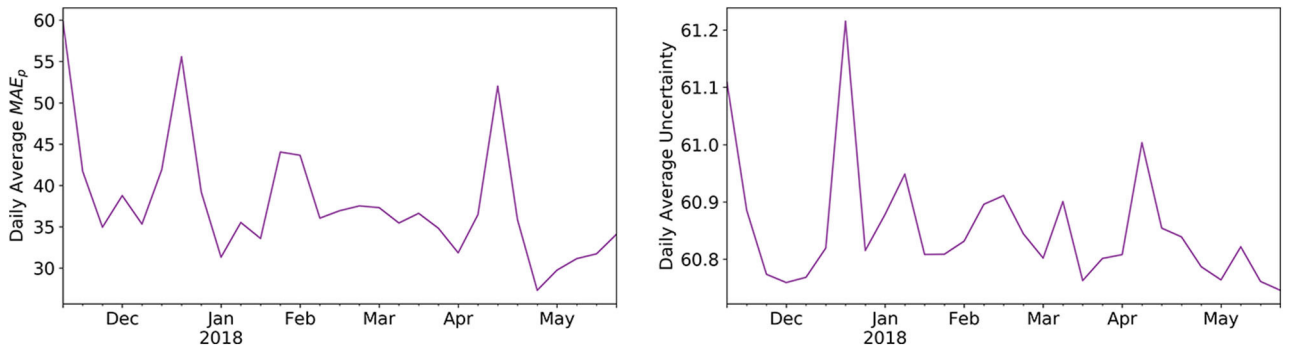

FIGURE 5 | Correlation between daily average mean average errors and predictive uncertainty. Shown is the example of a participant, where we plotted the daily average $\boldsymbol{MAE_p^d}$ and the daily average predictive distribution (both mean aggregated on a week level) as given by the marginal normal of the GPR.

knowledge, it is possible to further improve outing imputation, correcting somewhat for the bias caused by outing. Interestingly, no improvements were seen between using a static global factor and employing a person specific factor, suggesting, that even if no accelerometer ground-truth was available, outings may be corrected by a factor of around 1.4. We are not exactly sure why this is, but it may be due to the fact that we are using a very rough estimate anyways and the exact factor would only have an effect if our estimates were more accurate. However, this finding merits further investigation in different populations and under varying circumstances.

Using short-term data from a more accurate wearable device seems to work well for calibrating wireless PIR ambient sensor systems. Given that previous research on the calibration of PIR sensor systems to measure gait-speed also led to very promising results (21), such relatively simple initial calibration procedures should be considered in future long-term telemonitoring applications and research employing wireless PIR sensors.

After all, our calibration procedure has its obvious limitations and problems. In general, it should be noted that due to the relatively small sample size, generalization of our results involving statistical inference may be limited. Regarding the calibration procedure, most PIR sensors have relatively low sampling rates due to the having a refractory period and a restricted field of view. This makes it virtually impossible to get a completely accurate estimate of the real physical activity,

as we would get by using a high-frequency accelerometer. This means that there will likely always be a certain underestimation of physical activity even after calibration, as certain activities are just missed by the PIR system. Further, we should add that the approach can only function if someone is living alone. Although some work suggests PIR installations may be usable in a multi-person setting, this is likely not the case with physical activity quantification. Another problem is variance in results between participants (as can be easily seen in **Figure 4**). For certain people it did not seem possible to get a good calibration (although still slightly better than baseline), and even after in-depth manual investigation, in two instances we did not find any reasonable explanation for this behavior. Possible explanations could be that there were not enough sensors in a room, that the sensors were not placed ideally, or that the person's behavior makes it inherently difficult to capture physical activity using PIR sensors—for instance someone that is regularly taking care of the neighbor's pet. This is another important argument in favor of using reliable data for calibration of wireless systems. By employing cross-validation it is straight forward to identify installations for which there is a large disagreement before and after calibration, this also allows to manually check for potential biases using Bland-Altman plots. Considering medical applications, the validity of data coming from non-invasive ambient motion sensors is of particular importance for building up trust with this new technology, and may in that way allow for broader application. We would thus advice work related to contactless health monitoring to use more accurate and validated wearable devices for initial calibration and sanity checking of wireless sensors. Future work might evaluate similar calibration procedures applied to other modalities like contactless heart rate or breathing rate sensing. In addition, it would be very interesting to further investigate the found activity outside to activity inside ratio in larger populations of community-dwelling older adults.

CONCLUSION

We found that using calibration data from a wearable accelerometer, collected over 7–14 days, significantly improves physical activity estimates of wireless passive infrared sensor systems. This leads also to significantly stronger correlations with health indicators and outcomes, known to be associated with physical activity. Bayesian methods like Gaussian process regression, that work well with small datasets and provide an inherent predictive distribution, which can help in diagnosing when a calibration function deteriorates over time—for instance due to changes in a person's behavior. Time-spent outside should be imputed with the average activity throughout the same time period at home, multiplied by an individual outing factor. If an individual outing factor is not available, a factor of ∼1.4 may be used.

We conclude that using even relatively small amounts of wearable based ground-truth data over 7–14 days, PIR based wireless sensor systems can be calibrated to give largely better estimates of older adults' daily physical activity. This increase in performance translate directly to stronger correlations with a variety of age relevant health indicators and outcomes known to be associated with physical activity.

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 human participants were reviewed and approved by the Kantonale Ethikkommission des Kantons Bern, Murtenstrasse 31, 3010 Bern (KEK-ID: 2016-00406). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NS, HS, PB, PU, RM, and TN designed and planned the study. NS and HS installed and maintained the system and measured the participants. NS and AB analyzed the data. NS, AB, and HS wrote the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

The work related to this manuscript has been partially funded by InnoSuisse and partially by institutional funding. The authors declare that Innosuisse was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

We would like to thank all subjects for their participation. In addition, we thank everyone involved in gathering the presented data.

REFERENCES

1. Knickman JR, Snell EK. The 2030 problem: caring for aging baby boomers. *Health Serv Res.* (2002) 37:849–84. doi: 10.1034/j.1600-0560.2002.56.x
2. Bloom DE, Canning D, Lubet A. Global population aging: facts, challenges, solutions & perspectives. *Daedalus.* (2015) 144:80–92. doi: 10.1162/DAED_a_00332
3. Koch S. Healthy ageing supported by technology - A cross-disciplinary research challenge. *Informatics Heal Soc Care.* (2010) 35:81–91. doi: 10.3109/17538157.2010.528646
4. Beard JR, Bloom DE. Towards a comprehensive public health response to population ageing. *Lancet.* (2015) 385:658–61. doi: 10.1016/S0140-6736(14)61461-6
5. Conti M, Orcioni S, Madrid NM, Gaiduk M, Seepold R. A review of health monitoring systems using sensors on bed or cushion. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence*

and Lecture Notes in Bioinformatics) (Berlin: Springer Verlag). (2018). p. 347−58. doi: 10.1007/978-3-319-78759-6_32
6. Meister S, Deiters W, Becker S. Digital health and digital biomarkers – enabling value chains on health data. *Curr Dir Biomed Eng.* (2016) 2:577–81. doi: 10.1515/cdbme-2016-0128
7. Coravos A, Goldsack JC, Karlin DR, Nebeker C, Perakslis E, Zimmerman N, et al. Digital Medicine: A Primer on Measurement. *Digit Biomarkers.* (2019) 3:31–71. doi: 10.1159/000500413
8. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digit Med.* (2019) 2:14. doi: 10.1038/s41746-019-0090-4
9. Rantz MJ, Skubic M, Popescu M, Galambos C, Koopman RJ, Alexander GL, et al. A new paradigm of technology-enabled "vital signs" for early detection of health change for older adults. *Gerontology.* (2015) 61:281–90. doi: 10.1159/000366518
10. Skubic M, Guevara RD, Rantz M. Automated health alerts using in-home sensor data for embedded health assessment. *IEEE J Transl Eng Heal Med.* (2015) 3:2700111. doi: 10.1109/JTEHM.2015.2421499
11. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess Alzheimerâ€™s disease progression and intervention efficacy. *Front Aging Neurosci.* (2015) 7:102. doi: 10.3389/fnagi.2015.00102
12. Powell KE, Paluch AE, Blair SN. Physical activity for health: What kind? How much? How intense? On top of what? Annu *Rev Public Health.* (2011) 32:349–65. doi: 10.1146/annurev-publhealth-031210-101151
13. Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. *Total Daily Physical Activity and the Risk of AD and Cognitive Decline in Older Adults.* (2012). Available online at: http://n.neurology.org/content/neurology/78/17/1323.full.pdf (accessed September 16, 2018).
14. Taylor D. Physical activity is medicine for older adults. *Postgrad Med J.* (2014) 90:26–32. doi: 10.1136/postgradmedj-2012-131366
15. Garatachea N, Luque GT, Gallego JG. Physical activity and energy expenditure measurements using accelerometers in older adults. *Nutr Hosp.* (2010) 25:224–30. doi: 10.3305/nh.2010.25.2.4439
16. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: Issues and challenges. *Br J Sports Med.* (2014) 48:197–201. doi: 10.1136/bjsports-2013-093154
17. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank Study. *PLoS ONE.* (2017) 12:e0169649. doi: 10.1371/journal.pone.0169649
18. Porter EJ. Wearing and using personal emergency respone system buttons. *J Gerontol Nurs.* (2005) 31:26–33. doi: 10.3928/0098-9134-20051001-07
19. Wu X, Choi YM, Ghovanloo M. Design and fabricate neckwear to improve the elderly patients' medical compliance. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin: Springer Verlag). (2015). p. 222–34. doi: 10.1007/978-3-319-20913-5_21
20. Peek STM, Wouters EJM, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. *Int J Med Inform.* (2014) 83:235–48. doi: 10.1016/j.ijmedinf.2014.01.004
21. Rana R, Austin D, Jacobs PG, Karunanithi M, Kaye J. Gait velocity estimation using time-interleaved between consecutive passive IR Sensor Activations. *IEEE Sens J.* (2016) 16:6351–8. doi: 10.1109/JSEN.2016.2577708
22. Kaye J. Home-based technologies: a new paradigm for conducting dementia prevention trials. *Alzheimers Dement.* (2008) 4:S60–S66. doi: 10.1016/j.jalz.2007.10.003
23. Rantz MJ, Skubic M, Miller SJ, Galambos C, Alexander G, Keller J, et al. Sensor technology to support Aging in Place. *J Am Med Dir Assoc.* (2013) 14:386–91. doi: 10.1016/j.jamda.2013.02.018
24. Urwyler P, Stucki R, Rampa L, Müri R, Mosimann UP, Nef T. Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living. *Sci Rep.* (2017) 7:42084. doi: 10.1038/srep42084
25. Pais Bruno, Philipp B, Tobias N, Narayan S, Hugo S, Daniel G, et al. De nouvelles technologies au service du maintien à domicile des personnes âgées. *Rev Med Suisse.* (2019) 15:1407–11.
26. Schütz N, Saner H, Rudin B, Botros A, Pais B, Santschi V, et al. Validity of pervasive computing based continuous physical activity assessment in community-dwelling old and oldest-old. *Sci Rep.* (2019) 9:1–9. doi: 10.1038/s41598-019-45733-8
27. Saner H, Schütz N, Botros A, Urwyler P, Buluschek P, du Pasquier G, et al. Potential of ambient sensor systems for early detection of health problems in older adults. *Front Cardiovasc Med.* (2020) 7:110. doi: 10.3389/fcvm.2020.00110
28. Song B, Choi H, Lee HS. Surveillance tracking system using passive infrared motion sensors in wireless sensor network. In: *2008 International Conference on Information Networking* (Busan: IEEE) (2008). p. 1–5. doi: 10.1109/ICOIN.2008.4472790
29. Aran O, Sanchez-Cortes D, Do MT, Gatica-Perez D. Anomaly detection in elderly daily behavior in ambient sensing environments. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin: Springer Verlag). (2016). p.51–67. doi: 10.1007/978-3-319-46843-3_4
30. Rowlands AV, Mirkes EM, Yates T, Clemes S, Davies M, Khunti K, et al. Accelerometer-assessed physical activity in epidemiology: Are monitors equivalent? *Med Sci Sports Exerc.* (2018) 50:257–65. doi: 10.1249/MSS.0000000000001435
31. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva ICM, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol.* (2014) 117:738–44. doi: 10.1152/japplphysiol.00421.2014
32. White T, Westgate K, Hollidge S, Venables M, Olivier P, Wareham N, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. *Int J Obes.* (2019) 43:2333–42. doi: 10.1038/s41366-019-0352-x
33. Williams CKI, Rasmussen CE. *Gaussian Processes for Machine Learning.* Cambridge, MA: MIT Press (2006).
34. Wang JM, Fleet DJ, Hertzmann A. Gaussian process dynamical models for human motion. *IEEE Trans Pattern Anal Mach Intell.* (2008) 30:283–98. doi: 10.1109/TPAMI.2007.1167
35. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: Association for Computing Machinery (2016). doi: 10.1145/2939672.2939785
36. Bergstra J, Ca JB, Ca YB. *Random Search for Hyper-Parameter Optimization Yoshua Bengio.* (2012). Available online at: http://scikit-learn.sourceforge.net (accessed April 22, 2020).
37. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* (1991) 39:142–8. doi: 10.1111/j.1532-5415.1991.tb01616.x
38. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc.* (1986) 34:119–26. doi: 10.1111/j.1532-5415.1986.tb05480.x
39. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clin Gerontol.* (1986) 5:165–73. doi: 10.1300/J018v05n01_09
40. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x
41. Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the edmonton frail scale. *Age Ageing.* (2006) 35:526–9. doi: 10.1093/ageing/afl041

42. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. *J Clin Epidemiol.* (2006) 59:1087–91. doi: 10.1016/j.jclinepi.2006.01.014
43. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS ONE.* (2011) 6:e22922. doi: 10.1371/journal.pone.0022922
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: machine learning in Python. *J Mach Learn Res.* (2011) 12:2825–30. doi: 10.5555/1953048.2078195
45. Faul S, Gregorči,č G, Boylan G, Marnane W, Lightbody G, Connolly S. Gaussian process modeling of EEG for the detection of neonatal seizures. *IEEE Trans Biomed Eng.* (2007) 54:2151–62. doi: 10.1109/TBME.2007.895745
46. Griffiths TL, Lucas CG, Williams JJ, Kalish ML. *Modeling Human Function Learning With Gaussian Processes.* Vancouver, BC (2009).

Conflict of Interest: PB was employed by Domo-Safety SA, which is the manufacturer of the displayed sensor system.

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

APPENDIX

While most fields are self-explanatory, we describe some details regarding fields used in the calibration procedure. The "duration" attribute of the PirMotions table refers to how many seconds a given sensor was reporting motion. The location attribute of the same table describes the room the sensor was in and the time_ the exact time of the firing (in UTC). The activity field of the Biovotion1 table represents the normalized activity values stemming from the device's accelerometer and the time_ describes the exact measurement time (in UTC). The DoorSensors table's location field refers to the location the sensor was placed—in this work only entrance sensors were relevant. The status attribute describes whether the door was opened or closed and the time_ attribute denotes the exact time of this event (in UTC).

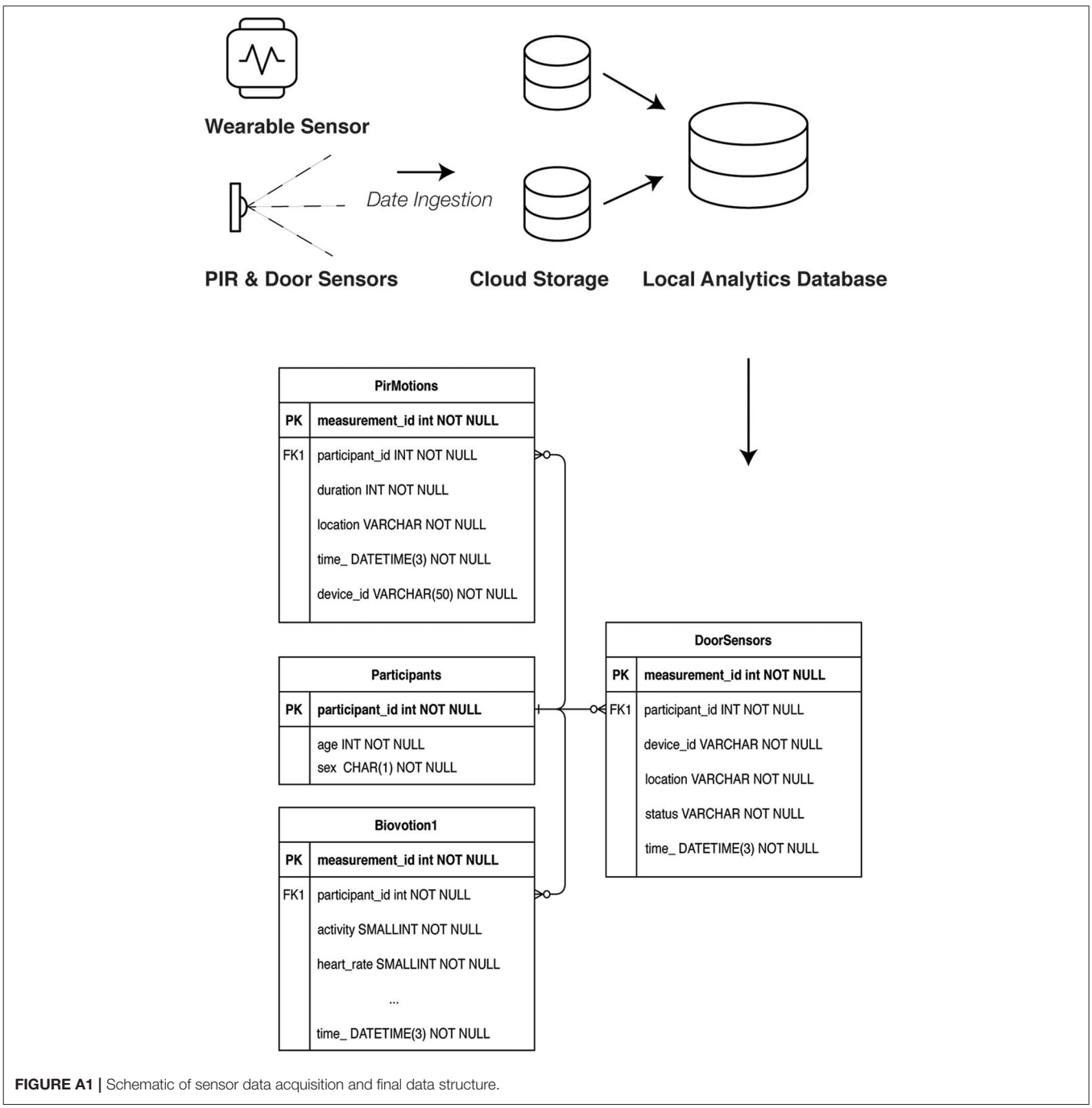

FIGURE A1 | Schematic of sensor data acquisition and final data structure.

frontiers in Digital Health

METHODS
published: 04 February 2021
doi: 10.3389/fdgth.2020.576076

The Use of Immersive Environments for the Early Detection and Treatment of Neuropsychiatric Disorders

Robert F. K. Martin[1], *Patrick Leppink-Shands*[1], *Matthew Tlachac*[1], *Megan DuBois*[2], *Christine Conelea*[2], *Suma Jacob*[2], *Vassilios Morellas*[1], *Theodore Morris*[1] *and Nikolaos Papanikolopoulos*[1]*

[1] Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, United States, [2] Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, United States

Neuropsychiatric disorders are highly prevalent conditions with significant individual, societal, and economic impacts. A major challenge in the diagnosis and treatment of these conditions is the lack of sensitive, reliable, objective, quantitative tools to inform diagnosis, and measure symptom severity. Currently available assays rely on self-reports and clinician observations, leading to subjective analysis. As a step toward creating quantitative assays of neuropsychiatric symptoms, we propose an immersive environment to track behaviors relevant to neuropsychiatric symptomatology and to systematically study the effect of environmental contexts on certain behaviors. Moreover, the overarching theme leads to connected tele-psychiatry which can provide effective assessment.

Keywords: tourette, autism, behavioral disorder, neuropsychiatric, immersive environment

OPEN ACCESS

Edited by:
Andreas S. Panayides, University of Cyprus, Cyprus

Reviewed by:
Dario Salvi, Malmö University, Sweden
Rebeca Isabel García-Betances, Polytechnic University of Madrid, Spain

***Correspondence:**
Nikolaos Papanikolopoulos
papan001@umn.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *25 June 2020*
Accepted: *06 November 2020*
Published: *04 February 2021*

Citation:
Martin RFK, Leppink-Shands P, Tlachac M, DuBois M, Conelea C, Jacob S, Morellas V, Morris T and Papanikolopoulos N (2021) The Use of Immersive Environments for the Early Detection and Treatment of Neuropsychiatric Disorders. Front. Digit. Health 2:576076. doi: 10.3389/fdgth.2020.576076

1. INTRODUCTION

Neuropsychiatric disorders, especially those which begin in childhood, are highly prevalent conditions that often remain chronic without effective treatment, which can lead to persistent disability and impairment across the lifespan. Recent research suggests that about 15% of youth are diagnosed with a mental disorder before 18 years of age (1). The ability to detect the onset of these conditions at the earliest possible time is critical for improved heath outcomes later in life. Critically, effective care hinges on accurate analysis of symptom presence and severity: clinicians must measure symptoms precisely in order to reach the proper diagnosis, identify an appropriately targeted treatment plan, and track the benefit of interventions. However, current methods for neuropsychiatric symptom assessment are severely limited.

Existing tools are limited to paper and pencil self-report measures of symptom severity and impairment, questionnaires, clinician-administered interviews, and observations (2, 3) that are prone to numerous rater biases and entail significant training burden for a clinician to achieve reliability. Reliance on observations can lead to differing results based on the context of those observations [e.g., parent vs. clinician or school vs. home; (4)]. These limitations can have detrimental consequences, including misdiagnosis of children, leading to unnecessary medical or psychiatric treatments (5), or under-diagnosis, such that children do not receive necessary mental health interventions. It is therefore critical to develop sensitive, quantitative tools to accurately identify early subtle neuropsychiatric abnormalities, improve diagnostic precision, and inform objective monitoring of treatment effects.

Many neuropsychiatric disorders display observable behavioral abnormalities that could be used to inform a quantitative evaluation. For example, motor behaviors are core symptoms in several neurodevelopmental disorders, such as motor and vocal tics in Tourette Syndrome (TS), restricted-repetitive behaviors in autism, compulsions in obsessive-compulsive disorder (OCD), and hyperactivity and motor overflow behaviors in attention-deficit hyperactivity disorder (ADHD).

Discussions in medical practice today focus on the value of early detection and intervention since they are often connected to better patient outcomes. Thus, the field of mental health is putting major emphasis on early detection (6). However, neuropsychiatric disorders typically begin with very subtle symptoms or prodromal indicators that can go missed by existing assays. The result is that detection and intervention may come too late in the disease process. Thus, tools that enable screening for the earliest manifestations of neuropsychiatric illness may help improve our ability to alter the illness trajectory by delivering prevention or intervention as close to disorder onset as possible. The major challenge is to discover quantifiable signs of neuropsychiatric illness (e.g., behaviors, neuromotor differences, physiological responses) that enable intervention at the first opportunity. Finally, clinicians want to accomplish these objectives in a connected health setting which enables effective assessments no matter where the patient or clinician are located.

2. BACKGROUND

2.1. Existing Tools to Assess Neuropsychiatric Disorders

Significant research in recent years has sought to develop comprehensive and reliable assessments to measure the presence and severity of symptoms of neuropsychiatric disorders. Current best-practices generally include administering batteries of multi-informant assessment measures comprised of psychometrically validated parent, teacher, and self-reported questionnaires and structured, clinician-administered rating scales. Parent-report measures typically include questions about observed child behavior and are intended to provide information about the extent to which symptoms exist and interfere with daily life. Clinician-rated scales typically involve a clinician making ratings of symptoms and their functional impact based on observation and information obtained via structured or semi-structured interview. Reports for these measures assume accurate reporting from parents, caregivers, and clinicians, but their subjective nature means that results can be influenced by rater biases [i.e., variability among raters due to differential interpretations of the question/rating scale or unique/divergent perceptions of the topic or individual being rated; (7)], response biases (e.g., social response biases, demand characteristics, recall/recency effects), or instrument biases (e.g., question ordering; cultural, semantic, or conceptual biases in the design of a questionnaire). Children, especially with disorders that affect communication, may not be able to articulate symptoms fully and will subsequently under-report symptoms. Similarly, parents may be consistently susceptible to under- or over-reporting depending on their understanding of the questionnaire and their desire for treatment (8). In addition, there are known challenges for interpreting these questionnaires. For example, research shows a lack of concordance between and parent and child reports of symptoms and symptom severity in ASD and ADHD (9). High rates of comorbidity among neuropsychiatric disorders further complicates interpretation of assessment measures because symptoms often overlap beyond a single categorical diagnosis or multiple independent diagnoses (10).

Parent-report questionnaires and clinician-administered interviews are intended to inform a categorical diagnosis about the presence or absence of a neuropsychiatric disorder and to assess symptom severity. This is an important first step for intervention access and planning, as well as for understanding how pervasive a disorder might be, but may not always provide specialized information about how to personalize treatment planning. There is a pressing need for measurement techniques that capture quantifiable symptoms objectively.

2.2. Early Detection

Although the onset and course of symptoms can differ between specific neuropsychiatric disorders, symptoms of some conditions can emerge as early as infancy and progress throughout development. Symptoms can also emerge during various developmental windows where risk for certain disorders is heightened, such as adolescence and the transition to adulthood. These developmental windows are characterized by periods of high neuroplasticity in which certain individuals are especially susceptible to developing symptoms of psychiatric disorders (11). Screening during these particular times in development can lead to earlier detection, which can ultimately help attenuate symptoms and slow the progression of a disorder.

In disorders that emerge during critical developmental periods, as well as in disorders that emerge in infancy, the delay between onset and diagnosis is significant, and treatment often does not begin until years later (12). Detecting symptoms at an early age is crucial for an earlier and more accurate diagnosis, and is crucial for improving outcomes long-term. For example, research has shown that detecting schizophrenia early, before debilitating symptoms emerge, can slow the course of the disorder and significantly improve outcomes (13, 14). Furthermore, because there are high rates of comorbidity in neuropsychiatric disorders, early intervention may curb effects of other related disorders that emerge later in the course of a disorder and worsen both physical and mental health.

2.3. Technological Approaches

Technology has many times been created for and applied to performing labor-intensive and tedious tasks. In the past century, advancements in computer technology have solved problems that typically involved intense computation or inordinate/vast amounts of data. However, with the low cost of cameras and increasing computational power of computers problems involving observation or surveillance are now being addressed. Neuropsychiatric disorders, such as TS, OCD, and autism usually require a close visual assessment by a trained clinician as a

component of the diagnosis. As the time of clinicians is a scarce resource, this is an area ripe for exploration.

As behavior disorders involve motion, there have been a number of different approaches to analysis. Bernabei et al. (15) used wearable accelerometers to track the body movement of a number of individuals with TS. Participants were visually recorded standing still and walking for a short period of time. The videos were annotated by trained clinicians and the acceleration data were analyzed for motion spikes. The researchers found good agreement between automatic and clinician tic detection but found that the method was limited in the range of tics that could be captured. One approach to Parkinson's disease (PD), a nervous system disorder that affects movement, was to insert pressure sensors into the soles of shoes (16). Using stride data gathered from the sensors, a hidden Markov model (HMM) was trained and used to discriminate between healthy subjects and those suffering from PD. Acoustic signals are another sensing modality that have been applied (17), along with a variety of machine learning techniques, to diagnose patients with PD with good results. Using recordings of speech patterns researchers were able to classify patients suffering from PD with an accuracy rate of larger than 95%.

Various studies have also used images and videos of a participant's face to detect psychiatric disorders. The human face not only expresses emotion but its motion also indicates an individual's mood. Scherer et al. (18) found four distinct descriptors using the eyes and lips to determine a correlation with symptoms of depression, anxiety, and post-traumatic stress disorder (PTSD). A wider range of emotions and mental states were recognized in Grafsgaard et al. (19) as predictor of educational outcomes. Using smaller features like the eyebrows, eyelids, or chin, researchers were able to determine when a participant was engaged in learning and when they became frustrated.

Autism spectrum disorder (ASD) is another neuropsychiatric disorder where technology can be applied to great effect. ASD is typically diagnosed through close observation during a clinic visit, which requires focused attention on the part of the clinician and may not be the most comfortable setting for the patient. A recent survey (20) focused on the applications of mobile computing for ASD early detection and remote monitoring and concluded that these efforts "will have significant clinical impact." Using multiple cameras, a child and parent were recorded doing structured play in a natural setting in Rehg et al. (21). Various behaviors were detected, such as gaze direction, smile detection, and object engagement, and these behaviors can be used as metrics to assist in the diagnosis of ASD. A recent study (22) has used an overhead camera and a small humanoid robot to detect and quantify interactions between a child and the robot. These can then be used to help diagnose and treat those with ASD. Other research groups (23, 24) used visual features like the Histogram of Gradients (HOG), the Histogram of Optical Flow (HOF), or SURF features to facilitate the analysis.

Video has also been used to analyze OCD. In Zor et al. (25), participants were recorded performing specific actions and the video was manually annotated. Using this annotation, researchers were able to discriminate between two types of compulsions: cleaning and checking. While the video was not automatically analyzed, a main conclusion from this research was that video analysis "offers an objective and practicable method by which to facilitate discernment" of OCD. Kim et al. (26) used a mixed group of participants with and without OCD to navigate through a virtual reality (VR) world to perform a variety of tasks. Using their interaction data with this virtual world researchers were able to differentiate between the two groups and further tests showed that checking behaviors could be elicited in virtual environments.

2.4. Immersive Environments

Immersive environments allow a person to interact with a computer-generated world. There are varying degrees of immersion based on which senses (e.g., sight or hearing) are simulated and to what degree a participant can control the world. Simple setups include head-mounted displays that allow 360° viewing while more complex uses include display, headphones, and haptic feedback to navigate the world. There are obvious applications in the domains of entertainment, education, and architecture, but in the last decade virtual reality has also been applied to medical diagnostics and treatments.

The ability to control an environment to induce stimuli has shown promise for general assessment (27) as well as possible treatments for specific disorder (28). Recent work (29) has shown that immersive environments, tailored to an individual's fear stimulus, can be used along with cognitive behavioral therapy (CBT) for children with ASD. In general, immersive environments have shown great potential benefit in this area (30). Recent research has looked at the applicability of VR toward the study of behavioral disorders. Oagaz et al. (31) created a virtual environment where the participants were asked to perform a task that tested cognitive ability, memory, balance, and motor skills. This study showed the validity of such a system toward the study of neuropsychiatric disorders.

3. INSTRUMENTATION AND METHODS

3.1. Components

Robustly sensing and recording movement for detecting behaviors such as tics, compulsions, and restrictive-repetitive behaviors will require a wide variety of different types of data. To this end, several sensors have been added or planned for inclusion. The challenge of synchronizing multi-modal sensor readouts was a major consideration when selecting the specific sensors and sensor types. A typical setup with a subject can be seen in **Figure 1**.

3.1.1. CAVE

The basis of the system is the Cave Automatic Virtual Environment (CAVE) (32). It is a virtual reality system that projects the user's surrounding onto a set of walls around them. This project uses a VisCube C4, which projects onto three walls and the floor. Unlike more widely available head-mounted virtual reality systems the CAVE allows users to see their own body rather than a rendered replacement. This feature provides an increased level of immersion, though this is lessened by the fact

FIGURE 1 | A co-author interacting with the CAVE system.

that a user will see outside of the virtual environment if they turn around or look up. Like traditional VR there is a hard limit on the area that a user can traverse and an external controller is necessary if the environment calls for a user to traverse a significant distance.

3.1.2. VICON System

The VICON system provides Vantage infrared cameras and a software tracker to locate objects affixed with retroreflective markers. Four infrared cameras are mounted in the top four corners of the CAVE. It allows for simultaneous tracking of objects which will be very useful during object manipulation tasks. It also includes functionality to record position and orientation information which can be played back in the tracker or exported. Recording can be triggered from the application GUI or via a UDP message. The software package utilizes a Virtual Reality Peripheral Network (VRPN) which relays real-time tracking information from the cameras.

3.1.3. Stereo Glasses

While in the CAVE, users wear Volfoni Edge glasses. These active glasses synchronize with the projectors and allow for the perception of 3D objects, both past the projector walls and inside the area where the user can move. The glasses also include a passive retroreflective tracker array which is tracked by the infrared cameras. The tracking determines the position and orientation of the user's head at all times. This is critical to the functioning of the CAVE as it determines what should be displayed on each projector as the user walks and looks around.

3.1.4. Video Camera Array

Four high-FOV Mobius Maxi cameras are mounted overlooking the CAVE. While the VICON cameras are useful for tracking objects, they are unable to record normal video. The Mobius cameras have a setting which allows them to auto-record which is particularly useful for synchronizing data. Each camera can be triggered to record simultaneously via a serial message to a typical micro-controller board such as an Arduino, which in turn triggers a relay module to the cameras.

3.1.5. High Fidelity Microphone

An Earthworks QTC40 omnidirectional condenser microphone is mounted hanging above the CAVE. It was selected because it is designed to detect very quiet sounds and capture a true representation of high and low frequencies. It is connected to an Apollo Twin USB audio interface which interfaces with the computer that runs the CAVE. In combination with the other sensors it can be used to help try to identify correlations between the subjects environment and the expression of behavioral anomalies, such as vocal tics.

3.1.6. Xsens Human Body Tracking System

The VICON and Mobius cameras both are susceptible to subject and object occlusion. To complement the cameras, an Xsens MVN Awinda will help track body position even when views are occluded. The system includes a vest and straps to attach 17 wireless inertial measurement units (IMUs) to major joints on a user's body. The accompanying software, MVN Analyse, fuses the IMU data, and provides a robust pose. UDP messaging is used to start and stop recording pose information making it easy to synchronize with the other data.

3.2. Integration

Because the projection onto the walls of the CAVE is informed by the head tracking of the user from the VICON cameras, a communication protocol is needed to transfer information from the tracker. This protocol should also work between separate machines, as it can be desirable to run the tracker and the CAVE on different computers for the sake of computational efficiency or as a result of the physical location of hardware. The VRPN is designed for this task. It allows for simple communication of tracking information, as well as any other desired peripherals such as game controllers, across a system.

The software MiddleVR is used to describe the physical projected display layout and to then update the perspective Left/Right eye view transformations used to render the world on each of the 4 display surfaces (33), using the acquired 3D position and orientation of the subjects head through the VRPN. MiddleVR also interfaces directly with other peripheral devices that can used by applications for testing and development (game pads, 3D wands, Mouse, Keyboards, and so on). The software also coordinates the 3D rendering pipeline to the projected 2D displays represented within the Windows graphics device driver. The representation appears as a second monitor display comprised of the four 2D display surfaces joined vertically together into a single tall display. The system architecture is shown in more detail in **Figure 2**.

MiddleVR also provides a plugin for the Unity3D game engine to aid in world creation for the CAVE. This allows for any information from VRPN and any other devices monitored by MiddleVR described previously to be read directly in Unity3D. The VICON tracking data is particularly useful as it allows a simulation to react to the user's position and the motion of any custom controllers with retroreflective arrays attached to them.

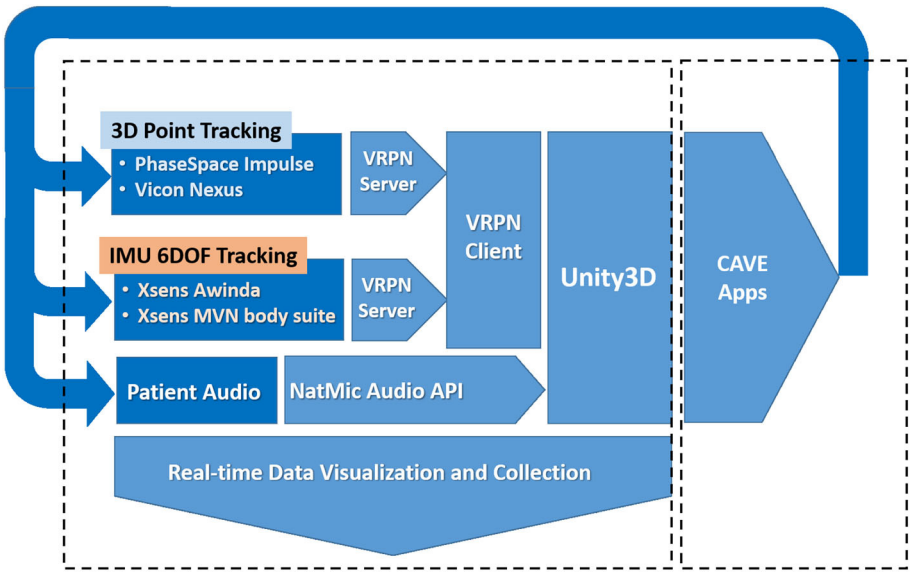

FIGURE 2 | Diagram of the system components.

The plugin also allows the world to be compiled in such a way that it can be properly displayed on the CAVE.

3.2.1. Synchronization

With many sources of data being recorded, synchronizing the data collection is vital. While it would theoretically be possible to read live streams of data from every sensor and then collect and save all the available data at a given time, this doesn't leverage the built-in recording functions of the various programs and devices and would likely result in reduced frequency of data sampling. The simplest solution is to ensure that each device is triggered to begin recording at the same time. This way all of the data being collected is synchronized without the need to record and later adjust for global timestamps.

In addition to world building, Unity3D is also used for synchronized triggering of recording. Attaching a C# script to an object in a Unity3D scene exposes functions that run at the creation and destruction of the object, which translate to the beginning and end of an application if the script is attached to an object that is always present in the scene. These methods are used to send UDP messages to Tracker and Analyze, to send serial messages to the microcontroller attached to the Mobius cameras, and to run commands to start and stop the microphone recording. This achieves the desired data synchronization between the independent recording methods.

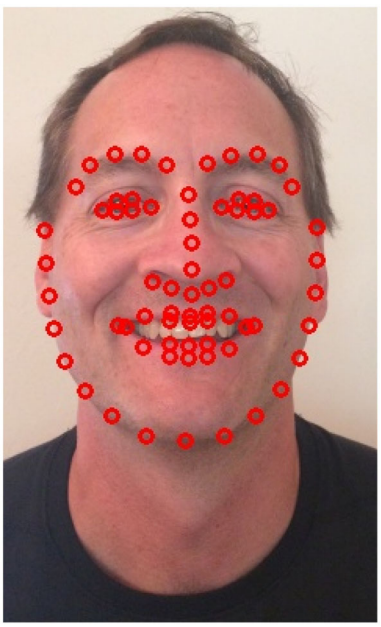

FIGURE 3 | Facial landmarks detected on a co-author.

4. RESULTS

4.1. Preliminary Results

The two most prevalent tic behaviors involve the eyes and the mouth. Close observation of the face allows for the detection of nearly half of typically expressed tics (34). Facial landmarks are detected and tracked as shown in **Figure 3**. For blinks, the relative locations and shapes of the eyes and eyebrows are computed and linear discriminants are used to distinguish blinking behavior as seen in **Figure 4**. Ongoing work is investigating the use of the complete set of facial landmarks coupled with a recurrent neural network (RNN) to further separate tic-like blinking from normal blinking.

Subjects with OCD show quantifiably different behaviors compared to a controlled population on certain common daily tasks. Experiments were designed to invoke and quantify

FIGURE 4 | Stills from video of co-author: **(A)** blinking normally, **(B)** mimicking tic behavior.

behavioral markers associated with OCD by performing common tasks within mock-up environment scenarios. We recorded video of subjects with OCD and a control group performing free-arrangement tasks and hand-washing tasks (35). A bathroom environment was constructed for subjects to perform hand-washing tasks. A soap dispenser, and drying towels were initially positioned next to a sink/faucet and toilet. The video cameras recorded top-down/frontal perspectives of 39 subjects performing the instructed tasks. Eighteen children were diagnosed with OCD, mean age 11.5 years, on the Yale-Brown Obsessive Compulsive Scale (CY-BOCS), with 21 children, mean age 10.7 years, representing the control group. Additional clinical assessment scales were completed by the child and parents. Parents completed the Child Obsessive-Compulsive Impact Scale-Revised (COIS-R) and Behavioral Assessment System for Children, Second Edition (BASC-2). The children also completed the COIS-R as well as the Multidimensional Anxiety Scale for Children-2 (MASC-2). For repetitive behavior quantification, revisiting similar objects and locations were manually counted and denoted in time by several blinded raters to address subjective bias. The total time to finish the tasks was also encoded by the raters.

Compared to the control group, OCD subjects were significantly more likely to exhibit other "extraneous" behaviors such as, the repetitive behaviors of touching/tapping, washing and drying the sink, and extensive exploration of the space ($p < 0.003$). The time duration to complete hand-washing task was also positively correlated with their CY-BOCS scores ($R = 0.54$, $p < 0.05$) and as well as with the CY-BOCS ordering/repeating dimension ($R = 0.57$, $p < 0.05$). The hand washing duration time indicated similar statistical correlation significance with the other administered test scores as well. Although the individual extraneous behaviors may be a useful behavioral marker, they were instead aggregated into a single measure due to limitations in the manual observation process to extract and quantify more detailed motion characteristics. Accordingly, we also examined the discriminative efficacy of several motion-based feature descriptors, derived from the automated extraction of the hand image-based motion trajectories (dense Farnsworth optical flow estimation) from the recorded video data (36). Generally, features based on the image pixel level trajectories of the subject hands, or their estimated velocities, correctly categorized over 80% of the subjects between the two groups. The results are considered preliminary due to the small sample size. In addition to testing more subjects to validate these findings, accurate motion trajectory information will be collected concurrently with the video to corroborate human observations and extract more complete behavioral marker characterizations. It has also been hypothesized that environment is a factor in compulsion expression. Our investigations will also test the effects of changing spatial layout, sound, and illumination on eliciting these behavioral marker differences in order to practice suppression techniques.

4.2. Anticipated Results

One anticipated result is the ability to consistently assess study participants or clinical patients at different times and locations. A difficulty in many current treatment regimens is the lack of time and resources to expertly quantify clinically relevant behaviors before, during, and after treatment. In this case, the data gathered with this instrument can be analyzed at any time and compared to other similar observations. There is also potential to utilize the instrument itself to deliver intervention. For example, a number of studies have consistently demonstrated (37) that children are better able to control tics when rewards are given for suppression. This system could "gamify" tic suppression through automatic tic detection and environment manipulation, opening up the possibility for standalone, game-like programs to train and support the ability to voluntarily suppress tics or for using such a tool to augment existing behavioral interventions that aim to improve tic control via skills training.

Additionally, there has been recent work that shows a correlation between galvanic skin response (GSR) and tic expression and frequency (38). One application of this instrument would be to connect the GSR with other external sensations (i.e., auditory or visual) to investigate strategies for tic suppression using biofeedback. GSR alone was not shown to create statistically significant reduction in tic frequency (39) but this shows the need for further research.

4.3. Use of Instrument

With the number of diverse components of the system it is important that the instrument be as streamlined as possible.

A researcher needs to simultaneously power on the CAVE projectors, VICON cameras, audio interface, and the microcontroller used to power and synchronously start Mobius camera recording. Subsequently, the researcher then needs to launch the tracker software and verify that the necessary VICON cameras are online. Following this there is a currently time-intensive calibration step for the Xsens tracking suit. Finally, any custom Unity3D applications can be started. At this point, the system is ready to be run.

5. CONCLUSION

The proposed system has broader impact than the domain of neuropsychiatric disorders. Education will be one of the primary beneficiaries of this system. Researchers in areas from computational complexity to computer vision to immersive environments will be able to design new experiments and collaborate with colleagues and students from varied backgrounds. Cross-disciplinary research encourages new ways of thinking about a problem and enhances critical thinking skills for all involved.

One of the most unique aspects of the proposed system is the ability to generate real-time, integrated feedback. The CAVE allows a participant to see the context, trigger, and resulting behavior as it occurs. This allows for more efficient symptom monitoring and training. Specific scenarios can be recreated and replayed at will until desired results are achieved.

Additionally, this immersive environment has the capability to make objective, quantifiable measurements. Most assays currently require a specifically trained human observer who nonetheless introduces bias and subjectivity. The precise digital recording feature of this system will increase consistency and objectivity of diagnostic. And quite often, human observers may not even be available for remote or rural populations (40, 41). This system and subsequent tools will be able to take the place of specialists where none are available or are not easily accessible (digital connected health). Because of the digitization process of the data, results can be easily sent for analysis to any location. This will greatly expand current treatment and diagnostic options for large segments of underserved populations.

Mental health assessment continues to make progress as diagnostic tools and treatments are refined and standardized. We hope that this progress is accelerated with the unique mix of capabilities of this system. Validated results from this tool will be used to create simpler and more streamlined new tools which can be deployed to other research partners, clinics, and perhaps even homes. With the ubiquity of personal computers and integrated cameras, there is an expectation that lessons learned with the CAVE setup can be utilized by anyone anywhere.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because, the dataset distribution requires permission by the UMN Human Subjects Committee. Requests to access the datasets should be directed to Nikolaos Papanikolopoulos, papan001@umn.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB UMN. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

RM has contributed 20% of the paper content, while the other authors have equal contributions. All authors contributed to the article and approved the submitted version.

FUNDING

This material was based upon work supported by the National Science Foundation through grants #CNS-1439728, #CNS-1338042, #CNS-1919631, and #CNS-1939033.

ACKNOWLEDGMENTS

The support of the Minnesota Robotics Institute is greatly appreciated.

REFERENCES

1. Dalsgaard S, Thorsteinsson E, Trabjerg BB, Schullehner J, Plana-Ripoll O, Brikell I, et al. Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence. *JAMA Psychiatry*. (2020) 77:155–164. doi: 10.1001/jamapsychiatry.2019.3523
2. McGuire JF, Piacentini J, Storch EA, Murphy TK, Ricketts EJ, Woods DW, et al. A multicenter examination and strategic revisions of the Yale Global Tic Severity Scale. *Neurology*. (2018) 90:e1711–9. doi: 10.1212/WNL.0000000000005474
3. Hus V, Lord C. The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *J Autism Dev Disord*. (2014) 44:1996–2012. doi: 10.1007/s10803-014-2080-3
4. Reyes ADL, Lerner MD, Keeley LM, Weber RJ, Drabick DAG, Rabinowitz J, et al. Improving interpretability of subjective assessments about psychological phenomena: a review and cross-cultural meta-analysis. *Rev Gen Psychol*. (2019) 23:293–319. doi: 10.1177/1089268019837645
5. Mol Debes N, Hjalgrim H, Skov L. Limited knowledge of tourette syndrome causes delay in diagnosis. *Neuropediatrics*. (2008) 39:101–5. doi: 10.1055/s-2008-1081457
6. Insel TR. Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Arch Gen Psychiatry*. (2009) 66:128–133. doi: 10.1001/archgenpsychiatry.2008.540
7. Hoyt WT. Rater bias in psychological research: when is it a problem and what can we do about it? *Psychol Methods*. (2000) 5:64. doi: 10.1037/1082-989X.5.1.64
8. Scarpa A, Reyes NM, Patriquin MA, Lorenzi J, Hassenfeldt TA, Desai VJ, et al. The modified checklist for autism in toddlers: Reliability in a diverse rural American sample. *J Autism Dev Disord*. (2013) 43:2269–79. doi: 10.1007/s10803-013-1779-x

9. Mazefsky C, Kao J, Oswald D. Preliminary evidence suggesting caution in the use of psychiatric self-report measures with adolescents with high-functioning autism spectrum disorders. *Res Autism Spectr Disord*. (2011) 5:164–74. doi: 10.1016/j.rasd.2010.03.006
10. Rutter M, Le Couteur A, Lord C. *Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services (2003).
11. Marín O. Developmental timing and critical windows for the treatment of psychiatric disorders. *Nat Med*. (2016) 22:1229. doi: 10.1038/nm.4225
12. De Girolamo G, Dagani J, Purcell R, Cocchi A, McGorry P. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. *Epidemiol Psychiatr Sci*. (2012) 21:47–57. doi: 10.1017/S2045796011000746
13. Guccione C, di Scalea GL, Ambrosecchia M, Terrone G, Di Cesare G, Ducci G, et al. Early signs of schizophrenia and autonomic nervous system dysregulation: a literature review. *Clin Neuropsychiatr*. (2019) 16:86–97.
14. Keane BP, Cruz LN, Paterno D, Silverstein SM. Self-reported visual perceptual abnormalities are strongly associated with core clinical features in psychotic disorders. *Front Psychiatry*. (2018) 9:69. doi: 10.3389/fpsyt.2018.00069
15. Bernabei M, Preatoni E, Mendez M, Piccini L, Porta M, Andreoni G. A novel automatic method for monitoring Tourette motor tics through a wearable device. *Mov Disord*. (2010) 25:1967–72. doi: 10.1002/mds.23188
16. Khorasani A, Daliri MR. HMM for classification of Parkinson's disease based on the raw gait data. *J Med Syst*. (2014) 38:147. doi: 10.1007/s10916-014-0147-5
17. Peker M, Sen B, Delen D. Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. *J Healthc Eng*. (2015) 6:281–302. doi: 10.1260/2040-2295.6.3.281
18. Scherer S, Stratou G, Lucas G, Mahmoud M, Boberg J, Gratch J, et al. Automatic audiovisual behavior descriptors for psychological disorder analysis. *Image Vis Comput J*. (2014) 32:648–58. doi: 10.1016/j.imavis.2014.06.001
19. Grafsgaard JF, Wiggins JB, Boyer KE, Wiebe EN, Lester JC. Automatically recognizing facial expression: predicting engagement and frustration. In: *EDM*. Raleigh, NC (2013).
20. Koumpouros Y, Kafazis T. Wearables and mobile technologies in Autism Spectrum Disorder interventions: a systematic literature review. *Res Autism Spectr Disord*. (2019) 66:101405. doi: 10.1016/j.rasd.2019.05.005
21. Rehg JM, Abowd GD, Rozga A, Romero M, Clements M, Sclaroff S, et al. Decoding children's social behavior. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2013). doi: 10.1109/CVPR.2013.438
22. Manner MD. Identifying differences in social responsiveness among preschoolers interacting with or watching social robots. In: *IJCAI*. Minneapolis, MN (2018). p. 5777–8. doi: 10.24963/ijcai.2018/832
23. Laptev I. On space-time interest points. *Int J Comput Vis*. (2005) 64:107–23. doi: 10.1007/s11263-005-1838-7
24. Wang H, Ullah MM, Klaser A, Laptev I, Schmid C. Evaluation of local spatio-temporal features for action recognition. In: *British Machine Vision Conference*. Grenoble (2009). doi: 10.5244/C.23.124
25. Zor R, Fineberg N, Eilam D, Hermesh H. Video telemetry and behavioral analysis discriminate between compulsive cleaning and compulsive checking in obsessive-compulsive disorder. *Eur Neuropsychopharmacol*. (2011) 21:814–24. doi: 10.1016/j.euroneuro.2011.03.006
26. Kim K, Kim SI, Cha KR, Park J, Rosenthal Z, Kim JJ, et al. Development of a computer-based behavioral assessment of checking behavior in obsessive-compulsive disorder. *Comprehens Psychiatry*. (2010) 51:86–93. doi: 10.1016/j.comppsych.2008.12.001
27. Parsons TD, McPherson S, Interrante V. Enhancing neurocognitive assessment using immersive virtual reality. In: *2013 1st Workshop on Virtual and Augmented Assistive Technology (VAAT)*. Orlando, FL (2013). p. 27–34. doi: 10.1109/VAAT.2013.6786190
28. Freeman D, Antley A, Ehlers A, Dunn G, Thompson C, Vorontsova N, et al. The use of immersive virtual reality (VR) to predict the occurrence 6 months later of paranoid thinking and posttraumatic stress symptoms assessed by self-report and interviewer methods: a study of individuals who have been physically assaulted. *Psychol Assess*. (2014) 26:841. doi: 10.1037/a0036240
29. Maskey M, Rodgers J, Grahame V, Glod M, Honey E, Kinnear J, et al. A randomised controlled feasibility trial of immersive virtual reality treatment with cognitive behaviour therapy for specific phobias in young people with autism spectrum disorder. *J Autism Dev Disord*. (2019) 49:1912–27. doi: 10.1007/s10803-018-3861-x
30. Riva G, Ba nos R, Botella C, Mantovani F, Gaggioli A. Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Front Psychiatry*. (2016) 7:163. doi: 10.3389/fpsyt.2016.00164
31. Oagaz H, Schoun B, Pooji M, Choi M. Neurocognitive assessment in virtual reality through behavioral response analysis. *IEEE J Biomed Health Inform*. (2018) 23:1899–910. doi: 10.1109/JBHI.2018.2881455
32. Cruz-Neira C, Sandin DJ, Defanti TA. The Design and Implementation of the CAVE. In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, SAP'17*. Ames: ACM (1993). p. 8:1–8:8. Available online at: https://dl.acm.org/doi/10.1145/166117.166134
33. Cruz-Neira C, Sandin DJ, Defanti TA, Kenyon RV, Hart JC. The CAVE: audio visual experience automatic virtual environment. *Commun ACM*. (1992) 35:64–72. doi: 10.1145/129888.129892
34. Ganos C, Bongert J, Asmuss L, Martino D, Haggard P, Münchau A. The somatotopy of tic inhibition: where and how much? *Mov Disord*. (2015) 30:1184–9. doi: 10.1002/mds.26188
35. Bernstein GA, Hadjiyanni T, Cullen KR, Robinson JW, Harris EC, Young AD, et al. Use of computer vision tools to identify behavioral markers of pediatric obsessive-compulsive disorder: a pilot study. *J Child Adolesc Psychopharmacol*. (2017) 27:140–7. doi: 10.1089/cap.2016.0067
36. Fasching J, Walczak N, Hadjiyanni T, Bernstein G, Cullen K, Mikkelsen M, et al. Exploring video descriptions for handwashing activities as part of an obsessive-compulsive disorder study. In: *Proceedings of 24th Mediterranean Conference on Control and Automation (MED)*. Athens (2016). p. 1174–9. doi: 10.1109/MED.2016.7535917
37. Greene DJ, Koller JM, Robichaux-Viehoever A, Bihun EC, Schlaggar BL, Black KJ. Reward enhances tic suppression in children within months of tic disorder onset. *Dev Cogn Neurosci*. (2015) 11:65–74. doi: 10.1016/j.dcn.2014.08.005
38. Nagai Y, Cavanna A, Critchley HD. Influence of sympathetic autonomic arousal on tics: implications for a therapeutic behavioral intervention for Tourette syndrome. *J Psychosom Res*. (2009) 67:599–605. doi: 10.1016/j.jpsychores.2009.06.004
39. Nagai Y, Cavanna AE, Critchley HD, Stern JJ, Robertson MM, Joyce EM. Biofeedback treatment for Tourette syndrome: a preliminary randomized controlled trial. *Cogn Behav Neurol*. (2014) 27:17–24. doi: 10.1097/WNN.0000000000000019
40. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. *Psychiatr Serv*. (2009) 60:1323–8. doi: 10.1176/ps.2009.60.10.1323
41. Snowden LR. Bias in mental health assessment and intervention: theory and evidence. *Am J Publ Health*. (2003) 93:239–43. doi: 10.2105/AJPH.93.2.239

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fdgth.2021.636082

Emerging and Established Trends to Support Secure Health Information Exchange

Emmanouil G. Spanakis[1], Stelios Sfakianakis[1], Silvia Bonomi[2], Claudio Ciccotelli[2], Sabina Magalini[3] and Vangelis Sakkalis[1]*

[1] Computational Biomedicine Laboratory, Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, Greece, [2] Department of Computer, Control, and Management Engineering Antonio Ruberti, Università degli Studi di Roma La Sapienza, Rome, Italy, [3] Emergency and Trauma Surgery Unit, Fondazione Policlinico Universitario Agostino Gemelli, Rome, Italy

This work aims to provide information, guidelines, established practices and standards, and an extensive evaluation on new and promising technologies for the implementation of a secure information sharing platform for health-related data. We focus strictly on the technical aspects and specifically on the sharing of health information, studying innovative techniques for secure information sharing within the health-care domain, and we describe our solution and evaluate the use of blockchain methodologically for integrating within our implementation. To do so, we analyze health information sharing within the concept of the PANACEA project that facilitates the design, implementation, and deployment of a relevant platform. The research presented in this paper provides evidence and argumentation toward advanced and novel implementation strategies for a state-of-the-art information sharing environment; a description of high-level requirements for the transfer of data between different health-care organizations or cross-border; technologies to support the secure interconnectivity and trust between information technology (IT) systems participating in a sharing-data "community"; standards, guidelines, and interoperability specifications for implementing a common understanding and integration in the sharing of clinical information; and the use of cloud computing and prospectively more advanced technologies such as blockchain. The technologies described and the possible implementation approaches are presented in the design of an innovative secure information sharing platform in the health-care domain.

Keywords: interoperability, health information exchange, eHealth, blockchain, security, patient consent

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Tsung-Ting Kuo, University of California, San Diego, United States
Michalis Zervakis, Technical University of Crete, Greece

****Correspondence:***
Emmanouil G. Spanakis
spanakis@ics.forth.gr

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *30 November 2020*
Accepted: *22 February 2021*
Published: *09 April 2021*

Citation:
Spanakis EG, Sfakianakis S, Bonomi S, Ciccotelli C, Magalini S and Sakkalis V (2021) Emerging and Established Trends to Support Secure Health Information Exchange. Front. Digit. Health 3:636082. doi: 10.3389/fdgth.2021.636082

INTRODUCTION

Information technology (IT) has long been identified as a cornerstone for the efficient, costless, timely, and reliable health-care delivery (1, 2). The availability of health-care information and patient records in digital form facilitates the persistence and posterity of valuable information and greatly support the decision-making process and even the extraction of new knowledge at both the individual and population levels. In our previous work, we have emphasized on current state of the art about cyber security in the health-care domain with emphasis on current threats and methodologies (3). This work is paraphrasing the famous words by John Donne, "no IT system is an island, entire of itself." Today, in a highly connected world where geographic boundaries

have been largely eliminated and people can freely move between cities, states, countries, or continents, the requirement for two different information systems to exchange a person's clinical data or medical history becomes vital and persistent. Sharing health information [or health information exchange (HIE)] through electronic means greatly improves the cost, quality, and patient experience of the health-care delivery.

To better secure the IT system's potential for interconnectivity and cooperation with other systems, the use of interoperable technologies and standards is needed. Depending on the extent and scope of the envisaged shared information spaces, there may be different levels of interoperability. **Figure 1** shows a proposed "maturity" model for interoperability in eHealth (4). The model consists of five levels that, incrementally, describe a more mature version of an interoperable infrastructure, starting from Level 1 for non-connected eHealth applications; Level 2 where a single eHealth application is directly linked to another application for simple data exchange (5); Level 3 for distributed systems that agree on protocols used, data formats, message exchange patterns (6), etc.; Level 4, where eHealth applications from different suppliers that serve a common goal are linked but the applications do not need to have common objectives (7); and finally, at the "universal" Level 5, where diverse eHealth applications connect to an open, interoperable infrastructure possibly spanning multiple countries (8, 9).

Interoperability and data sharing in the health-care domain is additionally challenging due to the multiplicity of the stakeholders, that is, the entities that operate (or are involved in any way) in this domain and which will be affected by any "disruption" or reform of the system. Some of the most important stakeholders or actors are therefore the following:

- The patients who are actually treated or, in general, are the recipients of the health services.
- The medical professionals (physicians and medical personnel) to provide the medical care.
- The health-care organizations (HCOs) (health-care providers) as represented by their director boards who actually administer the health delivery from a business perspective.
- The insurance companies that provide health coverage plans.
- The pharmaceutical companies that produce and market medications to be prescribed by physicians for the treatment of patients.
- The governments and other regulatory parties who control, coordinate, and set the rules, rights, and obligations of any involved party.

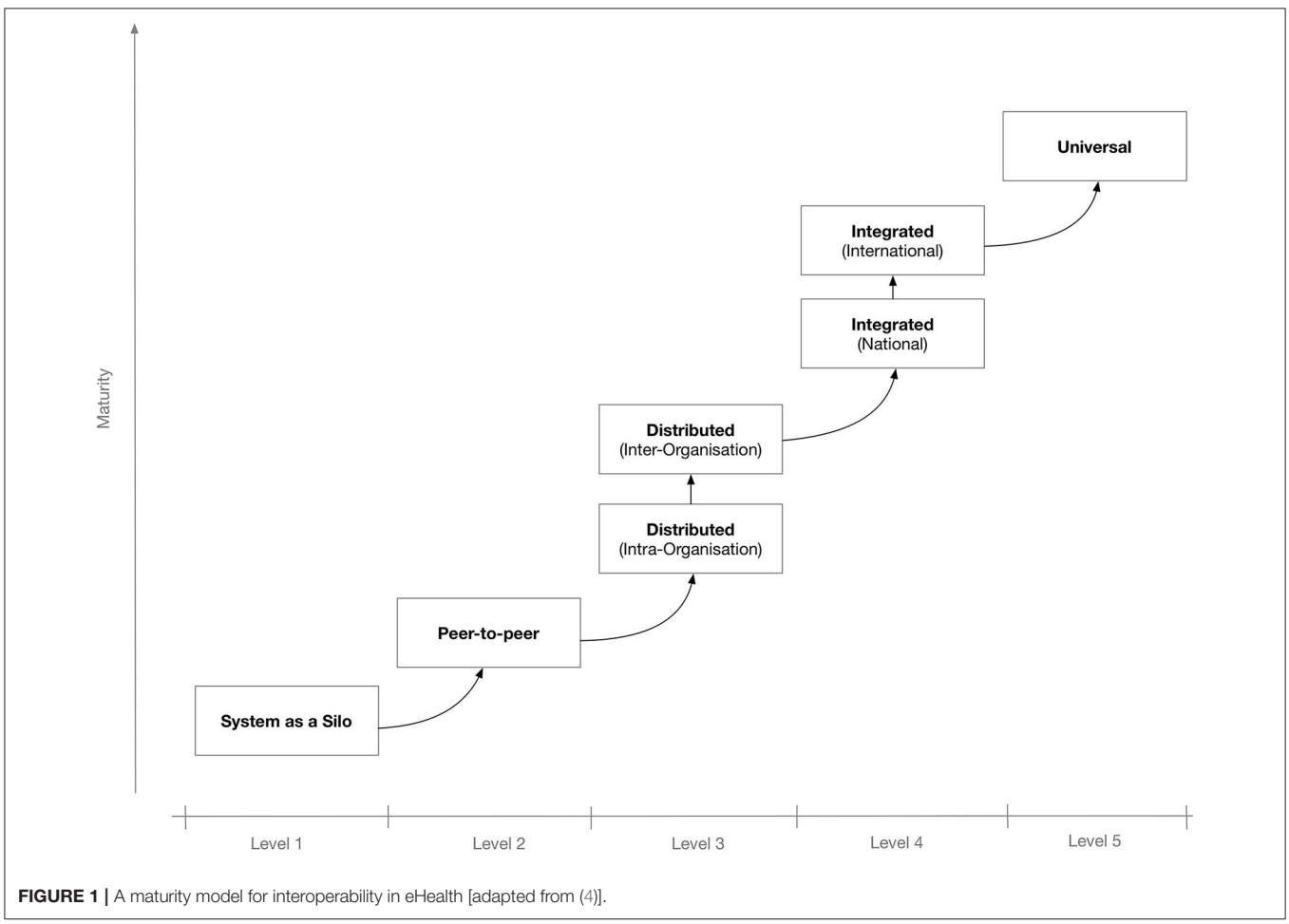

FIGURE 1 | A maturity model for interoperability in eHealth [adapted from (4)].

All these actors could have an influence in the design of a data sharing system and can also set important, and conflicting in some cases, requirements. For example, patients would like to have their medical record shared but only after their approval and only with specific authorized personnel in specific circumstances; an HCO can be extremely cautious about sharing the data of their patients with another organization because they are concerned by the security and availability of their systems; governments of EU member states (MSs) can impose strict laws about the transfer of their citizens in cross-border health-care treatment scenarios; and medical professionals require fast and effortless access to a patient's medical history in emergency situations, which cannot be the case if time-consuming authorization processes are the norm. It is imperative, therefore, even though the objective is to design a technical solution for the sharing of clinical data, that all these constraints and requirements are considered and addressed in a satisfying manner.

From a strictly technical point of view, the sharing platform may need to interoperate with a large number and diverse set of IT systems, each with their own protocols, data formats, etc., Some of the most important systems that manage patient-related data and could be used as data sources for information sharing are as follows:

- Electronic health records (EHRs): These are patient-centered systems that store and manage clinical information, such as, a patient's medical history, diagnoses, medications, immunization dates, allergies, radiology images, and lab and test results. They are managed by authorized personnel, usually in the context of a single HCO, although they can span more.
- Personal health records (PHRs): These are electronic applications that are used by people managing their own health information in a private and confidential environment. They are simpler systems than EHRs, and in some cases, they can be connected (temporarily or otherwise) to more enterprise level HCO systems (e.g., EHRs or other hospital information systems).
- Laboratory information systems (LISs): Used inside hospitals and clinics to record, manage, and store data for clinical laboratories in a patient-centric way (sending laboratory test orders to lab instruments, tracking those orders, and then recording the results in a searchable database).
- Picture archiving and communication system (PACS): These are systems used in a clinical setting for the storage and convenient access to medical images from multiple modalities (source machine types). Digital Imaging and Communications in Medicine (DICOM) is the standard format and suite of protocols for the storage and transfer of images from PACS services.

Moreover, in a health ecosystem, there may be additional systems, for example, for the management of insurance claims and clinical decision support systems (CDSSs).

In the general context, information sharing involves more than one party (health-care providers, organizations, etc.,) that needs to cooperate and agree on the way the exchange of information happens, and what the rules and policies are that govern it. Interoperability involves many different aspects, such as legislation and guidelines, contracts, and agreements between exchanging parties, governance and maintenance, shareable workflows, standardized data elements, semantic and syntactic choices, applications, technical infrastructure, and safety and privacy issues. The Refined eHealth European Interoperability Framework (EIF) is a set of recommendations that specify the standards, protocols, procedures, and policies that when deployed can improve the interoperability of eHealth applications within the EU and across its MSs by providing specific recommendations for all these aspects (10). **Figure 2** depicts how these aspects can be represented in interoperability "levels" that permit two different organizations to communicate.

This framework provides a great overview of the needed "glue" so that two or more health-care environments can collaborate and serve as a common, multilevel, and multi-perspective model on the interoperability requirements. The six different levels of the (refined) EIF are the following:

- Legal and regulatory: Legislation and regulatory guidelines that define the boundaries for interoperability across borders, but also within a country or region.
- Policy, which represents the contracts and agreements between the sharing organizations so that trust is established and responsibilities are assigned.
- Care process: Shared workflows that define how the integrated care is delivered and how these workflows are managed.
- Information, which defines the data models, the concepts and their values, the terminologies, and controlled vocabularies that cater for the common understanding of the exchanged electronic messages.
- Applications, which define how the data are extracted from and imported to the health-care information systems and how the transport of the data takes place using health-specific technologies and standards.
- IT infrastructure is at the lowest level and corresponds to general-purpose communication and network protocols.

This generic eHealth Interoperability Framework is highly relevant for the use cases of the health information sharing since sharing is greatly facilitated between interoperable systems/organizations. Here, we focus on the sharing of health-related information across HCOs and even across countries and continents. This is an important use case to improve the secure and efficient delivery of health care across Europe (11, 12). Cross-border health care in Europe has been recognized as of 2011 with Directive 2011/24/EU, which established patients' rights to access safe and high-quality health care, including across national borders within the EU, and their right to be reimbursed for such health care (13). As can be deduced by considering the Refined eHealth EIF (**Figure 2**), the cross-border sharing of clinical information is a complex scenario due to the fact that data need to be transferred between different countries and therefore, requires overcoming barriers such as, establishing a common trust framework, uniquely identifying citizens, and translating between different schemas and terminologies. This

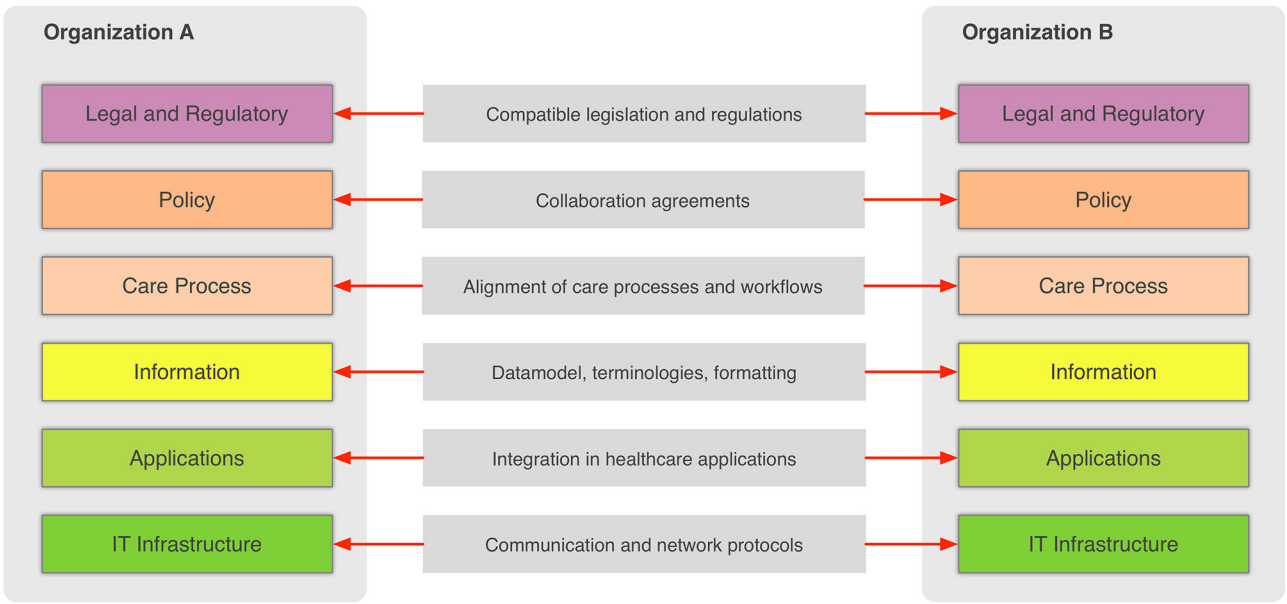

FIGURE 2 | Alignment of interoperability levels between two communicating organizations.

paper presents current approaches to address most of these issues in the European context while presenting and evaluating emerging technologies such as blockchain that have been in the limelight recently. Our objective is to take advantage of both well-established and novel technologies that complement each other in order to design an architecture for the secure exchange of clinical information in the European context.

MATERIALS AND METHODS

Current Status on Standards-Based Health Data Exchange

In the health-care industry, large standards developing organizations have defined numerous standards, data formats, terminologies, etc., in order to support the design and building of interoperable IT systems. Perhaps the most well-known and most important standards are the ones introduced by Health Level 7 (HL7) and SNOMED, which can be used as a foundation for the development of data exchange standards among eHealth systems (14). Two more standards organizations are Integrating the Health Enterprise (IHE), which focuses primarily on integration and interoperability; and the Clinical Data Interchange Standards Consortium (CDISC). CDISC produced the Operational Data Model (ODM) to "facilitate the regulatory-compliant acquisition, archive and interchange of metadata and data for clinical research studies" (15). ODM is an XML-based format that provides a number of constructs for modeling electronic case report forms (CRFs) and can also be used in sending forms data from a clinical trial system to an EHR system. In the area of medical devices, the Continua Health Alliance, a non-profit, open-industry coalition of health-care and technology companies working to establish a system of interoperable personal health solutions, develops an ecosystem of connected technologies, devices, and services that will enable the more efficient exchange of fitness, health, and wellness information (16). Among its proposed standards, Continua proposes specifications and standards such as Bluetooth, USB, medical devices (IEEE 1173), and HL7 to enable people to use home-based devices to monitor their weight, blood pressure, and glucose and blood oxygen levels and to share these data with their health-care professionals.

The exchanged information can be in multiple data formats based on the type of data, device category, etc., (17). Some of the most common formats based on the health applications using them are the following:

- For medical imaging, the use of DICOM is almost universal and defines not only the content (DICOM file format) but also communication protocols for the exchange of medical images (18).
- In the area of DNA sequencing and other -omics data formats including FASTQ file format (19), which is used to store sequence information, and the standard flowgram format (SFF), which is used to encode sequence reads (20).
- The majority of the EHR systems adopt the HL7 standard clinical document architecture (CDA) as the interoperable data format (21). CDA is part of the HL7 version 3 family, and it is based on a reference information model (RIM) that serves as a semantic model that consists of a set of structural components (e.g., classes with data types) and semantic relations that are used to represent clinical notes in the form of an extensible markup language document.

The use of controlled vocabularies and terminologies allows for the unambiguous representation of important value sets, such as,

the diagnosis and the medicines (22). The following are examples of such terminologies:

- The International Classification of Diseases (ICD) provides a common language for reporting and monitoring diseases, used throughout the world to compare and share data in a consistent standard way between hospitals, regions, and countries and over periods of time. It is used to classify diseases and other problems for payment, management, and research, as recorded on many types of health records including medical records and death certificates. ICD-11 is the latest version of it, whereas ICD-10 (released in 1993) remains widely used.
- SNOMED CT, already mentioned above, is the most comprehensive multilingual clinical health-care terminology available. It is used in EHR systems to facilitate clinical documentation and reporting and to retrieve and analyze clinical data. SNOMED CT is both a coding scheme, identifying concepts and terms, and a multidimensional classification, enabling concepts to be related to each other, grouped, and analyzed according to different criteria.
- Logical observation identifiers names and codes (LOINC) provides a set of universal identifiers for medical laboratory observations. LOINC provides codes for the observation names (e.g., eye color), not the observation finding (e.g., blue eyes). LOINC therefore provides codes for questions; and where needed, other vocabularies, such as, SNOMED CT, provide codes for the answers.
- The Unified Medical Language System (UMLS) is an important terminology resource, intended for use mainly by developers of health information systems. The UMLS "Metathesaurus" uses several different source vocabularies and seeks to reflect and preserve the meanings of concept names and relationships from these sources. It is therefore a valuable resource for the translation between the different source vocabularies.

Application-level interfaces are also needed to support the communication and exchange of the standards-based encoded information. The role of HL7 is principal on this front: HL7's name comes from "Level Seven," which, according to the Open Systems Interconnection (OSI) model that standardizes communication functionality in IT, corresponds to *application layer*. From its establishment in the late 1980s, HL7 was therefore focused on exchanging information within hospitals. The focus remains almost the same today, but HL7 has progressed from different paradigms over the years, in order to describe the structure, semantics, and management of the exchanged information. The development of HL7 version 3 (HL7v3) started around 1995 in order to introduce more consistency between the implementations of version 2 following an object-oriented development methodology. The most recent proposal by HL7 is Fast Healthcare Interoperability Resources (FHIR), which leverages web technologies to overcome the complexity of HL7v3 (23).

On the other hand, the IHE initiative has defined a number of "integration profiles," which are detailed specifications for communication among systems to address key clinical use cases, all based on established standards. IHE profiles organize and leverage the integration capabilities that can be achieved by coordinated implementation of communication standards, such as, DICOM, HL7, W3C, and security standards (24). Some of the IHE integration profiles that might be interesting in the context of interfacing health information systems and sharing of clinical information are as follows:

- Cross-Enterprise Document Sharing (XDS): Share and discover EHR documents between health-care enterprises, physician offices, and clinics, acute care in-patient facilities, and PHRs.
- Patient Demographics Query (PDQ): Enables applications to query by patient demographics (e.g., name) for patient identity from a central patient information server.
- Patient Identifier Cross Referencing (PIX): Allows applications to query for patient identity cross-references between hospitals, sites, HIE networks, etc.,
- PDQ HL7 v3 (PDQv3): Extends the PDQ profile leveraging HL7 version 3.
- PIX: Extends the Patient Identifier Cross-Reference profile leveraging HL7 version 3.
- Cross-Community Access (XCA): Allows to query and retrieve patients' EHRs held by other communities.
- Cross-Enterprise Document Reliable Interchange (XDR): Exchanges health documents between health enterprises using a web service-based point-to-point push network communication.
- And many others.

There are two main architectural approaches for the implementation of an information sharing platform: centralized and federated (25). In the centralized approach, a central data warehouse and accompanied services act as middlemen for the exchange of information and a single source of patient data that are shared among the participating organizations. On the other hand, in the federated architecture, a central infrastructure is also in place, but in this case, it merely acts as a facilitator for locating the data sources. An example of this case would be a common registry that stores only the links to the original patient records, medical images, etc., while the linked data are not transferred outside their primary premises unless explicitly requested by any interested client system. In addition to these opposite approaches for designing a distributed information sharing platform, there are also various hybrid options, such as using messaging with "publish-subscribe" communication that can be introduced to complement either the centralized or federated architectures.

There are advantages and disadvantages in all of the abovementioned deployment options. For example, in the federated approach, there are more strong concerns about the privacy, security, and availability of the data shared and their original sources (26). The operation of a mission-critical radiological information system (RIS) in a hospital can be severely affected if multiple peers request DICOM images from its PACS, and this poses an additional burden and cost for the acquisition and management of adequate infrastructure in the source organization. Instead, a centralized strategy allows for

easy access to the whole information shared but also leads to a concentration of the costs for maintaining the infrastructure needed and can be problematic at the operation level (a "single point of failure"). Furthermore, there are more costs on integrating the different data sets under a common "schema," resolving conflicts or even supporting the timely update of the persisted information when a source system acquires new or modified data.

Emerging Supportive Technologies: Blockchain

Blockchain is a *decentralized*, *distributed* data structure used to store *transactions* (aggregated in blocks) across many computers (27, 28). Blockchain has been extensively used for Bitcoin (29). For health care, we have seen the work of Kuo et al. (30) where they performed a systematic review on how blockchain can be used in health-care applications. Blockchain core is the embedded distributed ledger technology able to support for data integrity, authenticity, and origin. In blockchain, each block is linked to the previous one through a cryptographic hash, and it is a data structure that allows to store a list of transactions (31). In the blockchain, a transaction abstracts and allows to keep track of an exchange or interaction between two entities. Transactions are created and exchanged by peers of the blockchain network and modify the state of the blockchain data structure. An efficient categorization and a comprehensive overview of the latest privacy-preserving mechanisms and policies regarding privacy-preserving methods and characteristics, in smart electric grids, focusing on the use of the blockchain technology and the multi-authority access control paradigm is studied in (32, 33).

Concerning data access, we can have the following:

- *Public blockchain*: There are no restrictions on reading blockchain data and submitting transactions for inclusion into the blockchain.
- *Private blockchain*: Direct access to blockchain data and submitting transactions is limited to a predefined list of entities.

Concerning data management, we can have the following:

- Permissioned blockchain: Transaction processing is performed by a predefined list of peers with known identities.
- Permission-less blockchain: No restrictions on identities of transaction processors (i.e., blocks creators).

Combining the two perspectives, we can have four categories as depicted in **Figure 3**.

Blockchain supports auditability and transparency, as any reader is able to verify the correctness of the state of the system. Indeed, by storing all the transactions, it is possible to re-play (starting from a correct checkpoint) the entire history and check that the current state is consistent with the set of recorded transactions. It is important to note that, when using a blockchain to store data, there exists an inherent trade-off between transparency and privacy. Indeed, if the primary requirement is to have a fully transparent system, we need to accept that anyone is allowed to see any piece of information (sacrificing privacy). Conversely, if the primary requirement is to have a private system, it will not provide transparency. A trade-off between transparency and privacy is however possible, but it will come at the cost of efficiency, as it would require employing complex cryptographic primitives.

RESULTS AND DISCUSSION

A Novel View on Information Sharing

Nowadays, the sharing of a patient's clinical information between two HCOs (e.g., hospitals) usually requires a great amount of manual work in order to check and validate patient's consent (by consulting signed papers) or, at worst, results in privacy loss by extending the trust circle, e.g., to all physicians from the requesting organization. The main limitation of the current sharing pattern can be summarized in the following:

1. Sharing medical data may require time and possibly multiple interactions, also involving the patient in the loop.
2. Sharing medical data is currently a physical point-to-point interaction. If the same set of data needs to be shared with multiple parties, it would require multiple sharing patterns to be in place.
3. Sharing is currently asymmetric. Organization A may have the consent to share patient's data with organization B, but the inverse may not be true.

In order to overcome these deficiencies, we aim to design a platform—the Innovative Secure Information Sharing Platform (InSISP)—that is able to support a fast and efficient medical information sharing at both national and cross-national levels, taking into account sharing constraints, included those imposed by the General Data Protection Regulation (GDPR). It is imperative that patient's consent is central in this framework, and one of the challenges is to make the consent management robust, simple, and secure. A sequence diagram of the possible interactions to support the data sharing is shown **Figure 4**.

From an abstract point of view, InSISP can be seen as a data repository that can be accessed by HCOs (i.e., clients) to store and retrieve shared data by using a common interface and format (e.g., CDA Release 2 using standard vocabularies such as SNOMED and LOINC). The shared data repository is surrounded by a federation of collaborating entities (organizations/clients) that is dynamic and evolving. Once the federation is established, participating entities can start sharing data according to the data processing consent provided by patients. To this aim, we can identify two additional functionalities: (i) data sharing and (ii) data processing consent management. **Figure 5** summarizes a possible decomposition of the InSISP and highlights the three storage components.

Federation Management

The federation management functionality has the aim to manage the federation life cycle, and in particular, it should allow new HCOs to join and HCOs no longer interested in participating in

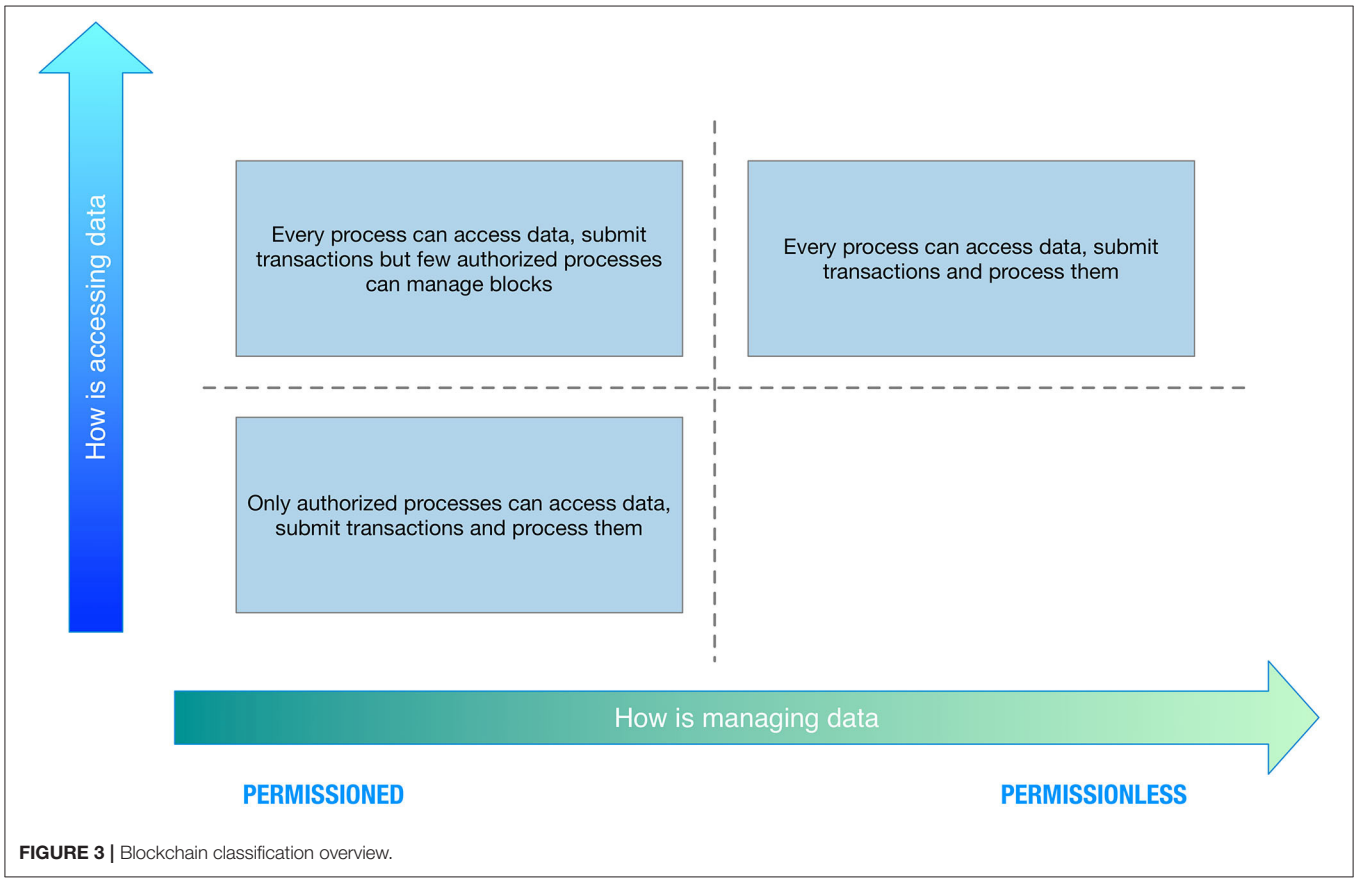

FIGURE 3 | Blockchain classification overview.

the federation to leave. In particular, this functionality supports the following operations:

1. *Join the federation*: HCOs should be able to become part of the federation at any time. When considering a membership service, all the members are assumed to be uniquely identified. So from the perspective of the InSISP development, there should be an external service that is able to support the identification of members and to provide them with a digital identity.
2. *Leave the federation*: HCOs may decide to leave the federation at their will, and the InSISP should support the removal of the entity from the membership and should notify the end of the sharing to connected entities.
3. *Get the federation membership*: Allows an HCO participating in the federation to get the current membership and know the set of HCOs potentially involved in the sharing, including their identification, public keys, and the categories of data that are currently available for the sharing.

The federation management intrinsically relies on the execution of a distributed protocol running, and thus, there are two main options to implement a federation membership service: (i) *client/server* or (ii) *peer to peer*.

In the client/server case, the current membership of the system is maintained by a trusted third party (TTP). When a new HCO wants to join the federation, it simply needs to contact the TTP and identify and authenticate itself with the TTP that will proceed by adding it to the current view. Similarly, when an HCO in the federation wants to leave, it simply needs to notify the TTP that will remove it from the current view. The current membership can be obtained again by querying the TTP. The main advantage of this option is that all the complexities of the membership management are delegated to the TTP. However, this also implies that the TTP is clearly a single point of failure for the system as well as its main bottleneck.

In the peer-to-peer approach, HCOs collaborate to maintain a consistent view of the system by exchanging messages and trying to reach a consensus on the sequence of views generated to include new members and to remove old ones.

Chockler et al. (34) discussed in detail the formalization and the specification of the group membership service, while more recently, Aguilera et al. (35) considered the problem of building a reconfiguration service to support the development of a distributed shared storage.

Let us note that in all these cases, the emphasis is on how to provide a consistent view to all the members. To the best of our knowledge, there does not exist any approach investigating the cost of realizing a membership service using blockchain technologies.

The main advantage of the peer-to-peer approach is its intrinsic resilience. In addition, in peer-to-peer settings, it is also possible to consider a blockchain-based approach to construct the sequence of consistent views providing the view auditability property for free. The main drawback is the cost

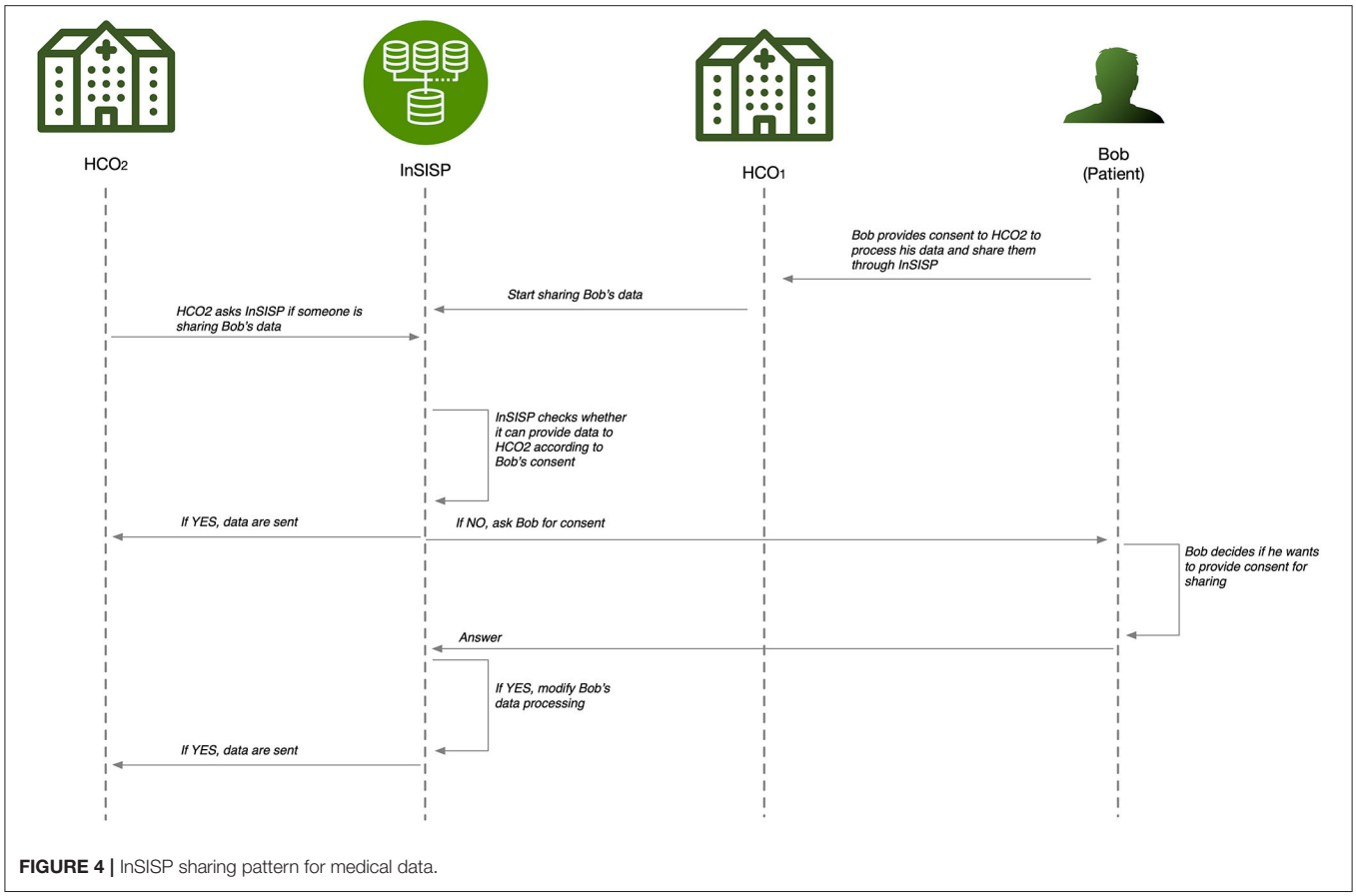

FIGURE 4 | InSISP sharing pattern for medical data.

imposed by the management of consistent views, as it requires to run coordination and synchronization protocols among all the participants that would bring poor scalability in case of a highly dynamic federation.

Data Processing Consent Management

The data processing consent management function has the aim to support the development of a digital data processing activity registry to store and access patients' data processing consent. The data processing consent is granted by a patient for a specific set of data to a specific set of entities and for a specific purpose and period of time. In order to support the automatic verification of patients' consent, such information must be stored and managed by the HCO in an electronic form (e-Consent) (36–38). Also, every patient has the right to modify his/her consent and has the right to be forgotten; i.e., at any point in time, he/she may ask to revoke all his/her previous consent (while also erasing any data identifying him/her).

To this aim, the data processing consent management functionality should offer the following operations:

1. *Provide new consent*: It allows to add a new entry to the table and to specify the beneficiary entities and to set the expiration time of the consent.
2. *Update consent*: It allows to modify existing consent by allowing the sharing with a new beneficiary or by removing a beneficiary.
3. *Remove (all) consent*: It basically implements the right to be forgotten by deleting all the consent previously provided by a specific patient.

Discussion About Possible Design and Deployment Options

Let us note that the data processing consent management function supports every HCO in managing its own data processing activity registry. Thus, from this point of view, we can say that it is local to every HCO.

As a consequence, the most appropriate choice is to design it as a local data store managed and accessed only by one HCO. Of course, in order to increase the resiliency and security of the storage, it can be also replicated, but all the replicas will still be managed by the same HCO.

A distributed design raises a privacy issue in patients' information. Indeed, even if data in the registry are not sensible by themselves, they could be easily correlated to infer sensitive information about patients and would result in a privacy violation; e.g., by looking to the list of HCOs where Bob did his analysis, you may infer that Bob is affected by a specific disease. To solve this issue, it is necessary to employ anonymization scheme generation an extra cost without any particular advantage in terms of reliability or security.

Data Sharing Management

Data sharing is the core functionality of the InSISP, as it manages the real transfer of medical data between parties. It offers just

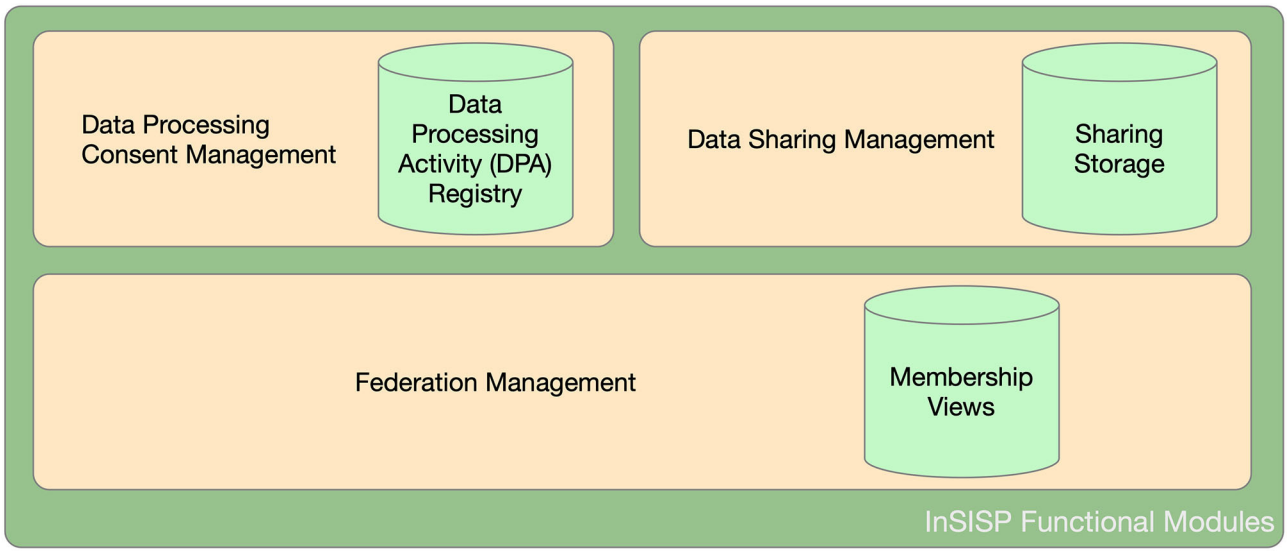

FIGURE 5 | InSISP functional decomposition.

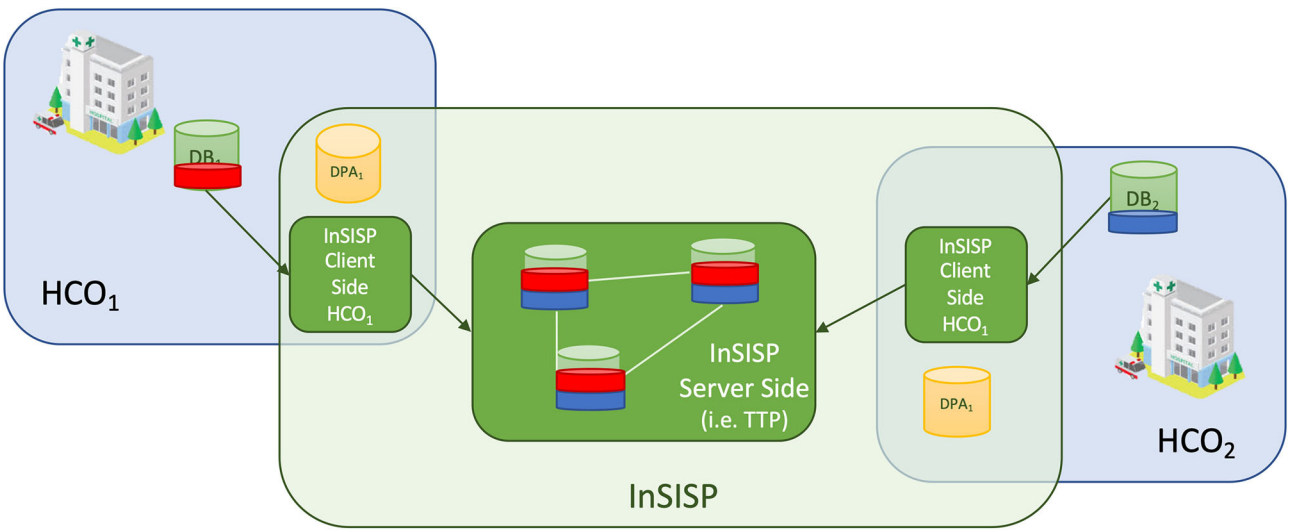

FIGURE 6 | Overview of the client/server design with a trusted third party (TTP).

one main operation, i.e., the Get Data, which is used to retrieve a specific piece of data for a specific patient and transfer it according with the patient's consent.

We can consider three main options to design and deploy this functionality:

1. Client/server.
2. Peer to peer with message exchanges.
3. Peer to peer with shared memories.

In the client/server case, data available for the sharing are copied and pushed toward a centralized TTP that will take care of satisfying the sharing request. In order to be GDPR compliant, a specific consent to move data to the TTP must be signed as well as the consent to share data with all the federation members[1]. **Figure 6** shows an overview of a possible client/server design.

In the peer-to-peer case, the idea is that the sharing is realized by letting HCOs in the federation cooperate with each other. We can distinguish two cases: cooperation realized through message exchange and cooperation realized through shared memory.

In the message exchange case, the sharing is realized using an *ad hoc* request–reply communication pattern as shown in **Figure 7**.

[1]Let us note that this second set of consent could be removed if every HCO provides the TTP a copy of its data processing activity registry. However, as mentioned above, this would add the complexity of finding a good anonymization scheme that allows to preserve patients' privacy still allowing the TTP to check consent.

In the peer to peer with shared memory design (**Figure 8**), each HCO creates a shared memory space where it stores all the pieces of data that can be shared according to the data processing activity registry. As an example, let us consider the case where Bob provided the consent to HCO_1 to share his X-ray images with HCO_2. This means that HCO_1 and HCO_2 create locally a shared

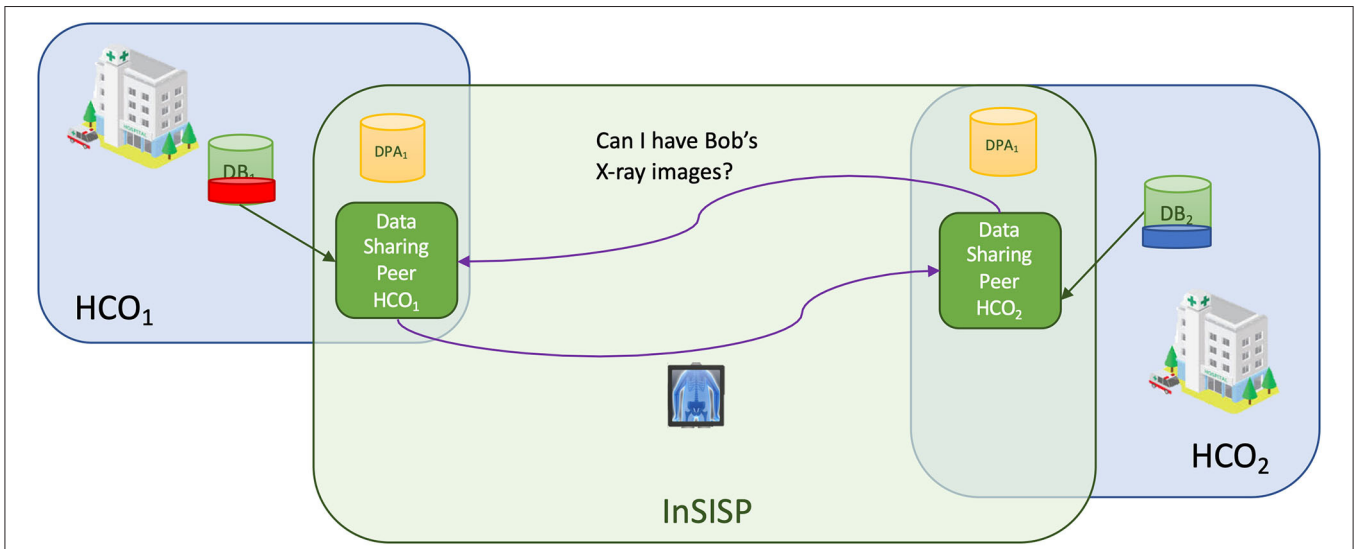

FIGURE 7 | Overview of the peer to peer with message exchange design.

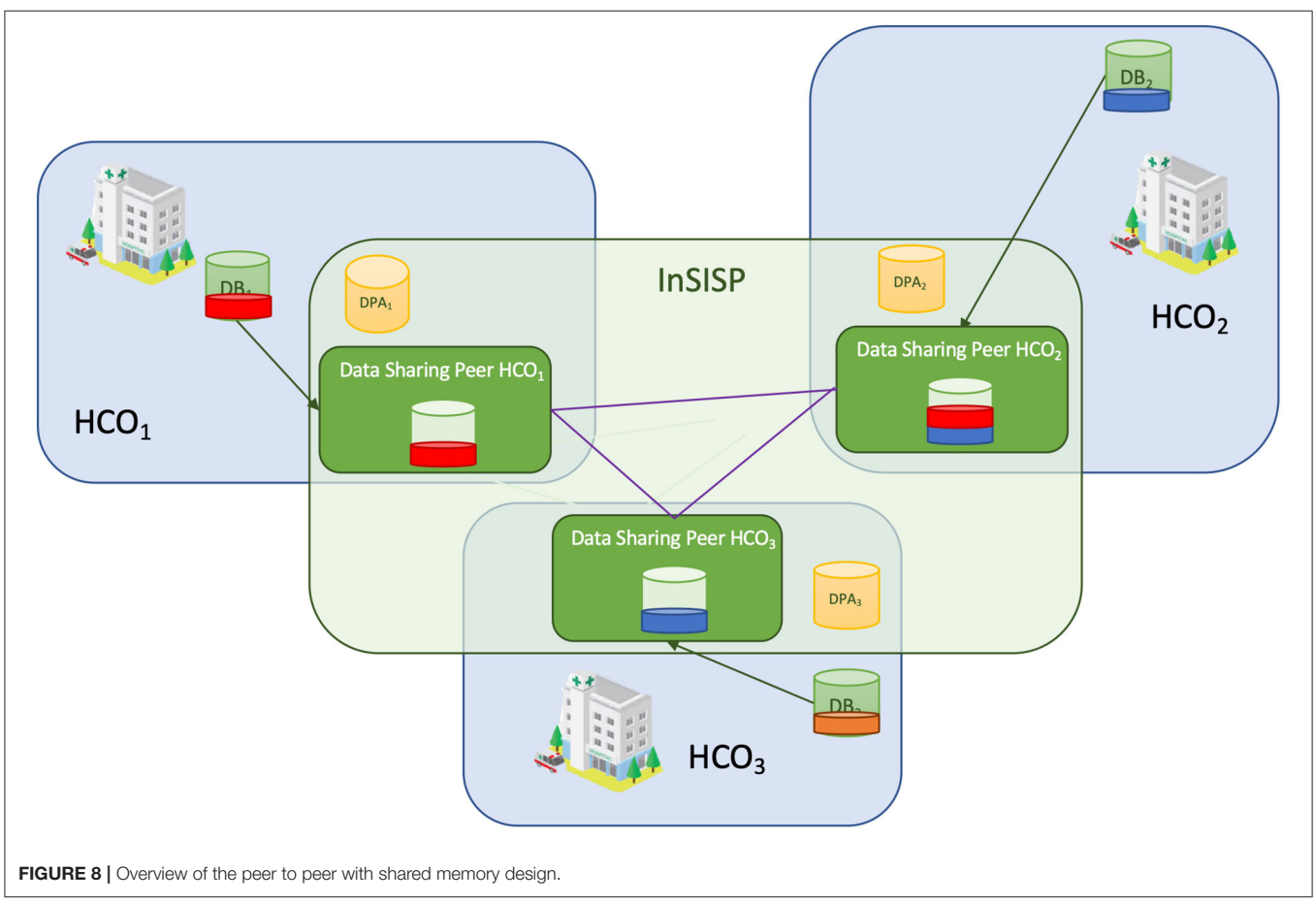

FIGURE 8 | Overview of the peer to peer with shared memory design.

space where they will store a copy of all Bob's X-ray images (the red slice in **Figure 8**).

Discussion About Possible Design and Deployment Options

Technically, all the three considered designs are feasible. However, as anticipated in the previous section, the client/server scenario poses several challenges from the point of view of the consent needed in order to make it compliant, and in particular, the main issue is the large set of consent that is necessary and that patients may be reluctant to provide. Instead, the two point-to-point designs solve this issue, as they exploit locally the consent information and move data only toward authorized HCOs.

Summary of Recommendations

According to the considerations done in the previous sections, **Table 1** summarizes the viable options for the design of each functionality of the InSISP.

Evaluation of Emerging Technical Solutions: Blockchain

In order to evaluate if blockchain is a valid option to support the InSISP deployment, the methodology presented in (39) considers the following three steps:

1. Requirement analysis to assess blockchain benefit.
2. Evaluation of the most appropriate blockchain solution where the designer is guided on the choice of the most suitable blockchain category, based on blockchain-specific criteria depending on who are readers and writers of the data and who is allowed to generate data.
3. Blockchain configuration selection, which assists the designer throughout the decision-making process for the configuration of the blockchain compliantly with the chosen category and the given project requirements.

Let us remark that blockchain is intrinsically a distributed system, and it makes no sense trying to use it when a distributed setting is not appropriate, which means that it can be used for the federation membership and for the data sharing functionalities of InSISP. In the following, we will evaluate the suitability of blockchain-based solution by adoption of the methodology described in (35).

In fact, trying to evaluate the blockchain technologies for the data sharing scenarios and design solutions described above, the first step is the analysis of requirements related to the component under analysis in order to understand the benefit of adopting a blockchain-based solution for its low-level design and development.

The factors that are considered in this step are listed in the following:

- Data or state storage. The first element to consider is to check if the module under analysis needs to store data or system state. If no information needs to be stored, clearly no blockchain is needed.
- Immutability and data integrity. With immutability, we refer to the property of a data to never change (i.e., a constant value that is never updated). If immutability is a requirement, then blockchain is certainly an option, as this is probably the most distinctive property of any blockchain. Integrity is strictly related to immutability, and this is why they are analyzed together, also considering that they are both closely related to cryptography. If a component requires data protection from unauthorized modifications, then this requirement can be met with a blockchain.
- Non-repudiation. Non-repudiation means that the author of some message/data cannot deny that it produced the message. This is another fundamental property that can be easily satisfied using blockchains.
- Multiple writers. This criterion considers the multiplicity of entities in charge of writing data in the storage. If only one entity is a writer, thus a common database is probably most appropriate than a blockchain especially from the performance perspective, i.e., in terms of throughput and latency.
- TTP always online. A TTP is an entity that facilitates interactions between mutually mistrusting entities. If in the system a TTP is required and it is planned to be always online, entities can delegate to it write operations as transactions, or state changes. Therefore, the TTP plays the role of a trusted deliverer and verifier. In this case, a blockchain, known for being a trust less technology, becomes useless, and the methodology brings to the related output. Otherwise, it can happen that the involvement of a TTP is planned but not for being always online: in this case, it could play the role of an authority giving authorizations for permissioned blockchains. Alternatively, a TTP may not exist at all. In the latter situations, it is not possible to exclude the recommendation of using a blockchain.
- Writers are known and trusted. If all the entities interested in writing know and mutually trust each other, a blockchain is superfluous and not recommended (again mainly for performance issues).

The flowchart of this analysis step is shown in **Figure 9**. According to this decision flowchart, **Tables 2**, **3** answer the relevant questions for the assessment of blockchain in the

TABLE 1 | Summary of design options and recommendations.

		Is it a valid design and deployment option?	
		Yes	**No**
DPA registry supporting the data processing consent management	Local	X	
	Centralized		X
	Distributed		X
Sharing storage supporting the data sharing	Local	X (peer to peer with message exchange)	
	Centralized		X
	Distributed	X	
Membership view supporting the federation management	Local		NA
	Centralized	X	
	Distributed	X	

DPA, data processing activity.

federation management and data sharing operations. Following the flowchart in **Figure 1** at a first glance, we get that blockchain is not recommended to support the implementation of the federation membership mainly because there exists a basic level of trust between members of the federation. In addition, we are also assuming that the federation membership is also relying on a trusted external service providing digital and secure identities to participants. However, we also highlighted the opportunity to consider a non-repudiation requirement, and we considered the importance of preserving data integrity (i.e., to ensure that the current membership cannot be altered). Thus, if these two requirements become more relevant or if assumptions on the identity platform or about trustworthiness of participant cannot be met, then blockchain becomes immediately a viable solution.

Using the same process, at first, we get that blockchain is recommended to support the implementation of the data sharing mainly because it is supporting efficiently the data integrity requirement; i.e., it allows to trace data accesses and verify their authorship and integrity. We should keep in mind that medical data are mostly read-only, and in our context, they are shared between trusted parties. Thus, the main benefit we can get by adopting a blockchain-based solution is the support for data integrity verification and data auditability. However, this feature must be carefully balanced with the "right to be forgotten" requirement (i.e., a MUST requirement imposed by the GDPR regulation) and its implication on the adoption of a blockchain-based solution. In order to support the implementation of the "right to be forgotten," we need to guarantee that data can be deleted from the blockchain when the data owner asks to do it. Currently, deleting data efficiently from a blockchain is still

TABLE 2 | Assessment of blockchain in the federation management.

FEDERATION MANAGEMENT	
Do you need to store data or state?	YES. The data to be stored are represented by the identifiers of federation members and their related information (e.g., keys used to verify the integrity of the messages).
Do you have immutability or data integrity requirements?	NO. HCO identifiers should not change. However, the set of identifiers may change in time, as well as some of the additional information stored may need to be updated (e.g., public/private keys may need to be refreshed as well as digital certificates). Data integrity YES. The federation membership should contain only the identifier of effective members. Identifiers should not be tampered or created.
Do you have non-repudiation requirements?	NO. Non-repudiation is desirable, but it is not currently a requirement.
Do you need to support multiple writers for the same data?	YES. Let us recall that the data stored are a set with all the identities of participating members. As a consequence, this set can be updated by any member that wants to leave or by a new member that is trying to join the federation.
Is there a TTP and is it always online?	NO. A TTP need to be assumed as bootstrap node to be accessed by members that want to join the federation. However, this node is not guaranteed to be always online.
Are writers known and trusted?	YES. In our context, we are considering the creation of a federation among collaboration entities. Thus, participating members must be known and trusted.

HCO, health-care organization; TTP, trusted third party.

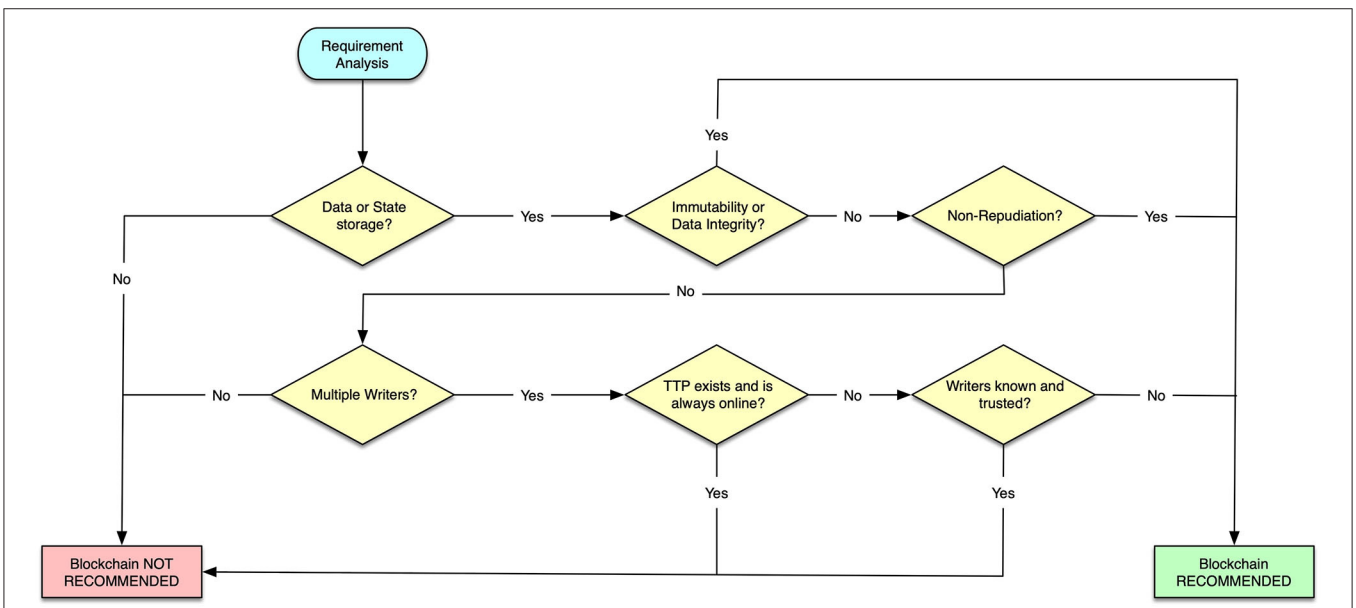

FIGURE 9 | Flowchart: blockchain, yes or no? adapted from (39).

TABLE 3 | Assessment of blockchain for data sharing operations.

DATA SHARING	
Do you need to store data or state?	YES. As the name suggest, the data sharing block is going to store medical data that need to be shared among the federation.
Do you have immutability or data integrity requirements?	Immutability NO. Medical data may need to be removed from the storage to guarantee the "right to forgotten" ruled out in GDPR, or specific fields may need to be updated as stated in the "right to rectification" (contact information associated to medical data). Integrity YES. Medical data have a strong integrity requirement (also coming from GDPR). In addition, immutability is also a highly desirable property.
Do you have non-repudiation requirements?	NO. Non-repudiation is desirable, but it is not currently a requirement.
Do you need to support multiple writers for the same data?	NO. When dealing with medical data, we are considering a type of data that are produced by a data producer and then become typically read-only (e.g., blood exam reports or X-ray images are produced, and then they can just be accessed in read mode).
Is there a TTP and is it always on-line?	YES. Without loss of generality, we can assume that the HCO that produced the medical is trusted and is always available.
Are writers known and trusted?	YES. Medical data can be produced only by HCOs, and they are assumed to be trusted and known in the federation.

GDPR, General Data Protection Regulation; TTP, trusted third party; HCO, health-care organization.

an open research problem, and the few existing solutions are currently based on the adoption of computationally expensive cryptographic techniques. Furthermore, when dealing with medical data, there is also the additional complexity following the huge heterogeneity of data to be considered (i.e., text, images, and images/sounds). Blockchain technologies have been originally designed to deal with transactional data, of small size and in the form of numbers or strings. Currently, it is not clear how to extend the paradigm to work with heterogeneous data. A possible solution to this issue could be to keep such heterogeneous medical data stored locally in a classical database and store in the blockchain only its hash. However, it is still not clear how much privacy lawyers consider such metadata as an expression of a personal data, and thus, the issue may remain.

CONCLUSIONS

Information sharing in the health domain is a complex and challenging process, since there are many stakeholders involved; different and sometimes competing standards and solutions to choose from; and important security, ethics, and regulation-related constraints for any proposed solution to comply with Markakis et al. (40). HIE is a key building block for the realization of Connected Health in Europe, which "speaks to the health journey of the person, through the entire lifespan, leveraging a variety of technologies to do so" (41). Based on the information and content of this document, it is important to consider the following aspects when building a new platform for sharing medical information:

- Patient consent is of utmost importance, and infrastructure should be in place for its registration, enforcement, and withdrawal.
- Compliance with GDPR and national laws, and implementing solutions to address significant requirements, such as the "right to erasure."
- Produce interoperable solutions by linking and interoperating with well-established standards such as document and data formats (e.g., CDA and DICOM) and metadata and value lists (e.g., SNOMED and LOINC) in order to support common understanding and integration.
- Handle the whole security spectrum: authentication and authorization of users, data privacy, auditing for "post-mortem" analysis and non-repudiation, data integrity, and machine-enforced trust among the sharing organizations.
- Enable the unique identification of patients while at the same time exposing the minimal set of personal information in order to protect their privacy.
- Performance, scalability, and availability of the whole platform should be high in order to support the health-related processes efficiently.
- Be part of the health-care ecosystem, which means allow easy integration with existing infrastructure by featuring interoperable "ports and adapters" interfaces.

Additionally, the IHE profiles should not be neglected. In the 2015/1302 Commission Decision, after consulting the European multi-stakeholder platform on information and communications technology (ICT) standardization and sectoral experts, 27 IHE profiles have been identified for referencing in public procurement, such as XDS and XDS-I, PIX, and PDQ (42).

Cloud computing is now used everywhere and provides an important set of features, such as, adaptive scalability, performance, and benefits from the business perspective. Especially for the sharing of clinical information, cloud can be very advantageous especially in cases where central repositories or central coordination are needed. But organizations should also be wary about the data protection, privacy, and access control mechanisms that should be in place (43–45), either offered by the cloud provider or built in house, in order to properly handle sensitive data and comply with regulations such as GDPR.

Blockchain is a highly interesting technology that can be put in good use in information sharing, more specifically for supporting decentralization, data integrity verification, and data auditability. These inherent features of blockchain have been praised and discussed in the context of health care as valuable tools (41, 46). Nevertheless, there are some major issues to be resolved, such as the compliance with GDPR's "right to be forgotten" requirement, which, unless the blockchain implementation is adapted, requires the deletion of data from the blockchain when the data owner asks to do it and this is not feasible, by design, in the "traditional" blockchain implementations. Furthermore, there is also the additional complexity followed by the huge

heterogeneity of medical data (i.e., text and images) that do not fit exactly to the original design of blockchain. It is evident that such requirements imposed by GDPR and the application domain present challenges and introduce additional trade-offs related to the management of data, administration, and overall governance. For example, storing the health data "off-chain" (i.e., external to the blockchain network) and only metadata "on-chain" may introduce problems of availability, performance, data protection, and integrity (47). Therefore, careful considerations of the available options should be made before committing to such cutting-edge technologies.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

ES, SS, SB, and CC made core contribution to the design of the study, collection of information, guidelines, established practices, and standards and made an extensive evaluation on new and promising technologies for the implementation of a secure information sharing platform for health-related data including blockchain evaluation. SM and VS coordinated the work in respect of the project PANACEA. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been supported by the PANACEA project that has received funding from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 826293.

REFERENCES

1. U. S. Congress. *Bringing Health Care Online: The Role of Information Technologies*. Washington, DC: Office of Technology Assessment (1995). p. 232.
2. Kolodner RM, Cohn SP, Friedman CP. Health information technology: strategic initiatives, real progress: there is nothing "magical" about the strategic thinking behind health IT adoption in the United States. *Health Aff.* (2008) 27:w391–5. doi: 10.1377/hlthaff.27.5.w391
3. Spanakis EG, Bonomi S, Sfakianakis S, Santucci G, Lenti S, Sorella M, et al. Cyber-attacks and threats for healthcare – a multi-layer thread analysis. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Montreal, QC (2020). p. 5705–8. doi: 10.1109/EMBC44109.2020.9176698
4. Velsen LV, Hermens H, d'Hollosy WO. A maturity model for interoperability in eHealth. In: *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. Munich (2016). p. 1–6.
5. Chronaki CE, Chiarugi F, Mavrogiannaki E, Demou C, Lelis P, Trypakis D, et al. An eHealth platform for instant interaction among health professionals. In: *Computers in Cardiology, 2003*. Thessaloniki Chalkidiki (2003). p. 101–4. doi: 10.1109/CIC.2003.1291100
6. Spanakis EG, Lelis P, Chiarugi F, Chronaki C. R&D challenges in developing an ambient intelligence eHealth platform. In: *EMBEC 2005, Prague, Czech Republic, November 2005*. Prague (2006). vol. 11; p. 1727–983.
7. Tsiknakis M. Spanakis M. Adoption of innovative eHealth services in prehospital emergency management: a case study. *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*. Corfu (2010). p. 1–5. doi: 10.1109/ITAB.2010.5687752
8. Spanakis EG, Psaraki M, Sakkalis V. Congestive heart failure risk assessment monitoring through internet of things and mobile personal health systems. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Honolulu, HI (2018). p. 2925–8. doi: 10.1109/EMBC.2018.8513024
9. Spanakis M, Sfakianakis S, Sakkalis V, Spanakis EG. PharmActa: empowering patients to avoid clinical significant drug–herb interactions. *Medicines*. (2019) 6:26. doi: 10.3390/medicines6010026
10. European Commission. *New European Interoperability Framework, Promoting Seamless Services and Data Flows for European Public Administrations*. (2017). Available online at: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20151123_co03_en.pdf
11. Katehakis DG, Pangalos G, Prentza A. Security improvements for better and safer cross-border ePrescription and patient summary services. *Int J Reliab Qual E-Healthc.* (2017) 6:18–28. doi: 10.4018/IJRQEH.2017010102
12. Tinholt D, Carrara W, Tol T, Foley P, Graux H, Erdogan E. Study on analysis of the needs for cross-border services and assessment of the organisational, legal, technical and semantic barriers. *Eur Comm-Gen Commun Netw Content Technol.* (2013) 10:10003. Available online at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=2310
13. Peeters M. Free movement of patients: directive 2011/24 on the application of patients' rights in cross-border healthcare. *Eur J Health Law.* (2012) 19:29–60. doi: 10.1163/157180912X615158
14. Benson T. *Principles of Health Interoperability HL7 and SNOMED*. 2nd ed. London: Springer-Verlag (2012). doi: 10.1007/978-1-4471-2801-4
15. Kuchinke W, Aerts J, Semler SC, Ohmann C. CDISC standard-based electronic archiving of clinical trials. *Methods Inf Med.* (2009) 48:408–13. doi: 10.3414/ME9236
16. Carroll R, Cnossen R, Schnell M, Simons D. Continua: an interoperable personal healthcare ecosystem. *IEEE Pervasive Comput.* (2007) 6:90–4. doi: 10.1109/MPRV.2007.72
17. Marias K, Sakkalis V, Roniotis A, Farmaki C, Stamatakos G, Dionysiou D, et al. Clinically oriented translational cancer multilevel modeling: the contracancrum project. In: Dössel O, Schlegel WC, editors. *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany. IFMBE Proceedings, vol 25/4*. Berlin: Springer (2009).
18. Larobina M, Murino L. Medical image file formats. *J Digit Imaging.* (2014) 27:200–6. doi: 10.1007/s10278-013-9657-9
19. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* (2010) 38:1767–71. doi: 10.1093/nar/gkp1137
20. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* (2010) 39:D19–21. doi: 10.1093/nar/gkq1019
21. Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D, et al. The HL7 clinical document architecture. *J Am Med Inform Assoc.* (2001) 8:552–69. doi: 10.1136/jamia.2001.0080552
22. Freitas F, Schulz S, Moraes E. Survey of current terminologies and ontologies in biology and medicine. *RECIIS.* (2009) 3:239–49. doi: 10.3395/reciis.v3i1.239en
23. Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. Porto. (2013). p. 326–31.
24. Kondylakis H, Petrakis Y, Leivadaros S, Iatraki G, Katehakis D. Using XDS and FHIR to support mobile access to EHR information through personal health apps. In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. Cordoba. (2019). p. 241–4. doi: 10.1109/CBMS.2019.00058
25. Eckman BA, Bennett CA, Kaufman JH, Tenner JW. Varieties of interoperability in the transformation of the health-care information infrastructure. *IBM Syst J.* (2007) 46:19–41. doi: 10.1147/sj.461.0019

26. Nikoloudakis Y, Pallis E, Mastorakis G, Mavromoustakis CX, Skianis C, Markakis EK. Vulnerability assessment as a service for fog-centric ICT ecosystems: a healthcare use case. *Peer Peer Netw Appl.* (2019) 12:1216–24 doi: 10.1007/s12083-019-0716-y
27. Nofer M, Gomber P, Hinz O, Schiereck D. Blockchain. *Bus Inf Syst Eng.* (2017) 59:183–7. doi: 10.1007/s12599-017-0467-3
28. Vergne J. Decentralized vs. distributed organization: blockchain, machine learning and the future of the digital platform. *Organ Theory.* (2020) 1. doi: 10.1177/2631787720977052
29. Macdonald M, Liu-Thorrold L, Julien R. The blockchain: a comparison of platforms and their uses beyond bitcoin. *Work Pap*. (2017) 3:1–8. doi: 10.13140/RG.2.2.23274.52164
30. Kuo T-T, Rojas HZ, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. *J Am Med Inform Assoc*. (2019) 26:462–78. doi: 10.1093/jamia/ocy185
31. Chowdhury MJM, Ferdous MS, Biswas K, Chowdhury N, Kayes ASM, Alazab M, et al. A comparative analysis of distributed ledger technology platforms. *IEEE Access*. (2019) 7:167930–43. doi: 10.1109/ACCESS.2019.2953729
32. Triantafyllou A, Jimenez JAP, Torres ADR, Lagkas T, Rantos K, Sarigiannidis P. The challenges of privacy and access control as key perspectives for the future electric smart grid. *IEEE Open J Commun Soc*. (2020) 1:1934–60. doi: 10.1109/OJCOMS.2020.3037517
33. Radoglou Grammatikis P, Sarigiannidis P, Efstathopoulos G, Panaousis E. ARIES: a novel multivariate intrusion detection system for smart grid. *Sensors*. (2020) 20:5305. doi: 10.3390/s20185305
34. Chockler GV, Keidar I, Vitenberg R. Group communication specifications: a comprehensive study. *ACM Comput Surv*. (2001) 33:427–69. doi: 10.1145/503112.503113
35. Aguilera MK, Keidar I, Malkhi D, Shraer A. Dynamic atomic storage without consensus. *J ACM*. (2011) 58:1–32. doi: 10.1145/1944345.1944348
36. Coiera E, Clarke R. e-Consent: the design and implementation of consumer consent mechanisms in an electronic environment. *J Am Med Inform Assoc*. (2004) 11:129–40. doi: 10.1197/jamia.M1480
37. Bergmann J, Bott OJ, Hoffmann I, Pretschner DP. An eConsent-based system architecture supporting cooperation in integrated healthcare networks. *Stud Health Technol Inform*. (2005) 116:961–6.
38. Bergmann J, Bott OJ, Pretschner DP, Haux R. An e-consent-based shared EHR system architecture for integrated healthcare networks. *Int J Med Inf*. (2007) 76:130–6. doi: 10.1016/j.ijmedinf.2006.07.013
39. Staderini M, Schiavone E, Bondavalli A. A requirements-driven methodology for the proper selection and configuration of blockchains. In: *2018 IEEE 37th Symposium on Reliable Distributed Systems (SRDS)*. Salvador: IEEE (2018). p. 201–6. doi: 10.1109/SRDS.2018.00031
40. Markakis E, Nikoloudakis Y, Pallis E, Manso M. Security assessment as a service cross-layered system for the adoption of digital, personalised trusted healthcare. In: *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*. Limerick: IEEE (2019). p. 91–4. doi: 10.1109/WF-IoT.2019.8767249
41. Angraal S, Krumholz HM, Schulz WL. Blockchain technology applications in health care. *Circ Cardiovasc Qual Outcomes*. (2017) 10:e003800. doi: 10.1161/CIRCOUTCOMES.117.003800
42. European Commission. *Commission Decision (EU) 2015/1302 on the identification of 'Integrating the Healthcare Enterprise' profiles for referencing in public procurement*. (2015). Available online at: https://data.europa.eu/eli/dec/2015/1302/oj
43. Spanakis EG, Spanakis M, Karantanas A, Marias K. Secure access to patient's health records using SpeechXRays a mutli-channel biometrics platform for user authentication. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL (2016). p. 2541–4. doi: 10.1109/EMBC.2016.7591248
44. Radoglou Grammatikis PI, Sarigiannidis PG, Moscholios ID. Securing the internet of things: challenges, threats and solutions. *Internet Things*. (2019) 5:41–70. doi: 10.1016/j.iot.2018.11.003
45. Pasquale F, Ragone T. *Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing*. Faculty Scholarship (2014). Available online at: https://digitalcommons.law.umaryland.edu/fac_pubs/1542
46. Linn LA, Koo MB. Blockchain for health data and its potential use in health it and health care related research. In: *ONC/NIST Use of Blockchain for Healthcare and Research Workshop*. Gaithersburg, MD: ONC/NIST (2016). p. 1–10.
47. Miyachi K, Mackey TK. hOCBS: a privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design. *Inf Process Manage*. (2021) 58:102535. doi: 10.1016/j.ipm.2021.102535

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.

frontiers in Digital Health

METHODS
published: 14 June 2021
doi: 10.3389/fdgth.2021.628646

Integration of Biobanks in National eHealth Ecosystems Facilitating Long-Term Longitudinal Clinical-Omics Studies and Citizens' Engagement in Research Through eHealthBioR

Athos Antoniades[1], Maria Papaioannou[1], Apostolos Malatras[2], Gregory Papagregoriou[2], Heimo Müller[3,4], Petr Holub[4], Constantinos Deltas[2] and Christos N. Schizas[1]*

[1] eHealth Lab, Department of Computer Science, University of Cyprus, Nicosia, Cyprus, [2] biobank.cy Center of Excellence in Biobanking and Biomedical Research, University of Cyprus, Nicosia, Cyprus, [3] Institute of Pathology, Medical University of Graz, Graz, Austria, [4] Biobanking and Biomolecular Resources Research Infrastructure - European Research Infrastructure Consortium, Biobanks and Biomolecular Resources Research Infrastructure Consortium, Graz, Austria

OPEN ACCESS

Edited by:
Chris Nugent,
Ulster University, United Kingdom

Reviewed by:
Sabine Müller,
Fraunhofer Institute for Industrial Mathematics (FHG), Germany
Anne-Christin Hauschild,
University of Marburg, Germany

***Correspondence:**
Athos Antoniades
cspgaa2@cs.ucy.ac.cy

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *12 November 2020*
Accepted: *11 May 2021*
Published: *14 June 2021*

Citation:
Antoniades A, Papaioannou M, Malatras A, Papagregoriou G, Müller H, Holub P, Deltas C and Schizas CN (2021) Integration of Biobanks in National eHealth Ecosystems Facilitating Long-Term Longitudinal Clinical-Omics Studies and Citizens' Engagement in Research Through eHealthBioR. Front. Digit. Health 3:628646. doi: 10.3389/fdgth.2021.628646

Biobanks have long existed to support research activities with BBMRI-ERIC formed as a European research infrastructure supporting the coordination for biobanking with 20 country members and one international organization. Although the benefits of biobanks to the research community are well-established, the direct benefit to citizens is limited to the generic benefit of promoting future research. Furthermore, the advent of General Data Protection Regulation (GDPR) legislation raised a series of challenges for scientific research especially related to biobanking associate activities and longitudinal research studies. Electronic health record (EHR) registries have long existed in healthcare providers. In some countries, even at the national level, these record the state of the health of citizens through time for the purposes of healthcare and data portability between different providers. The potential of EHRs in research is great and has been demonstrated in many projects that have transformed EHR data into retrospective medical history information on participating subjects directly from their physician's collected records; many key challenges, however, remain. In this paper, we present a citizen-centric framework called eHealthBioR, which would enable biobanks to link to EHR systems, thus enabling not just retrospective but also lifelong prospective longitudinal studies of participating citizens. It will also ensure strict adherence to legal and ethical requirements, enabling greater control that encourages participation. Citizens would benefit from the real and direct control of their data and samples, utilizing technology, to empower them to make informed decisions about providing consent and practicing their rights related to the use of their data, as well as by having access to knowledge and data generated from samples they provided to biobanks. This is expected to motivate patient engagement in future research and even leads to participatory design methodologies with citizen/patient-centric

designed studies. The development of platforms based on the eHealthBioR framework would need to overcome significant challenges. However, it would shift the burden of addressing these to experts in the field while providing solutions enabling in the long term the lower monetary and time cost of longitudinal studies coupled with the option of lifelong monitoring through EHRs.

Keywords: biobanks, ethical, national health, legal and social issues (ELSI), electronic health record

INTRODUCTION

"It's more important to know what sort of person has a disease than to know what sort of disease a person has" is a phrase attributed to the ancient Greek physician Hippocrates. It is a phrase that was expressed almost 2,500 years ago and is one of the most powerful justifications for the necessity and the value of designing and developing complete, functional, and reliable electronic data and medical sample capture systems today that support medical research. It is clear, nowadays, that different types of persons require a different type of health management, either proactive or reactive (1). Being able to understand the type of person paves the way to select the most suitable and beneficial health strategy for the citizen's benefit. Thus, well-designed and implemented systems at the national level that include a well-integrated electronic health record (EHR) system to biobanks may be regarded as a tool for painting the full picture of a citizen's health status. This, in turn, will enable the discovery of specific disease subtypes that may be associated with specific treatment outcomes, empowering personalized healthcare and adapting preventive or therapeutic recommendations on the specialized health needs of every citizen.

By serving the citizen-centricity idealism, the central actor of the eHealth ecosystem is the citizen who is responsible for making the proper selections to improve his/her own health quality. Accordingly, it is crucial to allow the citizen to understand thoroughly the impact of choices at both the personal and societal levels, in terms of health and financial status. For example, citizens should become educated about the meaning of pharmacogenomics testing and how it can be used to minimize the ineffective medications and the ineffective doses as well as to control drugs' adverse effects (2–4). Human genome sequencing is an important biomedical finding, which generates an explosion of genetic data obtained for clinical purposes [e.g., spot medically actionable diseases or variants for which preventive measures are available (5, 6)]. The power of the analysis of this kind of genetic data can be further strengthened if it is incorporated into national integrated EHR systems (7, 8). The continuous dramatic drop in price of the genome sequencing escalates the need to find a way to accomplish this integration (9), and thus a mass adoption of sequencing-based technologies can be utilized for clinical care improvement (10). Notably, the genetic/genomics-based medicine is reported to reduce costs and improve outcomes mainly because of its preventive character, which allows the identification of potential or early-stage health problems (2).

Biobanks collect, catalog, and store biological samples, acting as a biorepository tool with great impact on medical research. Biobanks also act as databanks that maintain the data generated from the analyses of samples, providing researchers access to large numbers of digitized data across many subjects for often cross-purpose research studies. However, biobanks have provoked questions on privacy, research, and medical ethics. To address these challenges, biobanks adhere to governing principles and policies that ensure legal and ethical adherence of national-level (National Bioethics committees approvals) and European-level requirements [General Data Protection Regulation (GDPR)]. Digital means to support these activities have been created to minimize privacy risks and support research data sharing, which is part of the primary objective of biobanks.

EHRs are a systematized collection of patients' electronically stored health information in a digital format. EHR registries can exist in the institutional (hospitals and clinics) as well as national level. These are designed to store data accurately and to capture the state of the health of each person across time. The focus is not on the population but on the single individual patient; therefore, their direct application in research, although possible, is coupled with significant challenges. These are both technical (lack of homogeneity, missing data, etc.) and policy challenges (privacy legal/ethical) (11). They provide real-time digital citizen's health information records designed following the principles of citizen centricity. Consequently, all the actors of the healthcare community (healthcare providers, medical systems, medical organizations, etc.) should work as one around the citizen facilitating healthcare and improving treatment outcomes (4). To achieve that, the citizens have their own health data ownership, and they are given the power to control the access to their individual EHRs. Accordingly, healthcare providers will be given access to the most relative medical information by the citizen and the ability to create or maintain a citizen's EHR content data. In that way, the citizen will be able to dictate that his/her medical data follow him/her across the continuum of his/her care without any information discontinuities, loss of data, or communication problems between systems or people in a systematic fashion (8). Thus, EHR can be considered the backbone of a national eHealth ecosystem, which essentially joins the pieces of a citizen's health puzzle together.

An integrated national EHR system designed to cover the whole population and to be interoperable gives the right to the citizens to enjoy portable health, since EHR of a citizen will

be remotely accessible by every healthcare provider across the country (12). To achieve that, a strong collaboration between IT and healthcare professionals is required (13) to define precisely and adequately EHR contents that will sufficiently embed workflows and automations reflecting the traditional applied clinical practice. As stated in (14), the process of rebuilding the EHR system from the ground up will be painful, but doing it properly "will make apparently insolvable problems solvable." Furthermore, implementing such a system under the umbrella of the National eHealth Authority (NeHA) as explained in (12) will assure the standardization of the data elements and the architecture of the records (15) to be compliant with the EU directives and guidelines. Thus, accomplishing the primary target to develop a national eHealth ecosystem embracing all the healthcare actors under one standardized system will use a basic healthcare instrument in the integrated national EHR.

Many successful attempts have been made to capitalize on the wealth of data in EHRs for research purposes. These rely on focusing on a set number of specific attributes of interest to the research topic and on policies that need to be put in place to gain access to the data. The data attributes selected undergo quality control and transformations to meet homogeneity requirements before they are used in research. A common approach used to address privacy concerns is the pre-processing at the EHR registry side of the data to anonymize it, providing metadata that meet the research requirements while maintaining anonymity (k-anonymity approaches) (16, 17) and limiting the impact of privacy concerns. Others, instead, focus on federated analysis approaches where the analyses are run at each EHR registry with only the research outcomes made available to researchers. Even with high-dimensional genetic data, a data type inherently difficult to anonymize k-anonymity approaches can be applied to transform the data to be usable for research anonymized metadata (18).

METHODOLOGY

The target of the proposed eHealthBioR framework is to facilitate the implementation of biomedical research by allowing the citizen to provide authorization to specific research studies that link the citizen's data sourced from a biobank and an EHR system. Citizens will be able to monitor the access to their data, and they will be able to withdraw their consent at any moment for any individual project. The eHealthBioR takes for granted that the biobank and the national integrated EHR system have collected and stored data in the corresponding databanks from the same individuals. The legal and ethical framework in which biobanks and the national integrated EHR system work internally is out of the scope of this paper. Rather, the target is to provide a framework schema that allows the citizens to have an active role in their participation in research studies combining data sourcing from a national biobank and EHR while providing the infrastructure to allow for the mechanisms to be compliant with GDPR, as well as with other national or international legal or ethical requirements.

Current State-of-the-Art in Electronic Health Record–Biobank Integration

Biobanks can benefit the most from their integration with the national EHR by utilizing the wealth of detailed and longitudinal EHR data sources, which are enriched for clinically relevant phenotypes and outcomes to study genetics at a population scale (5, 19, 20). A nationwide EHR-coupled biobank also permits the reduction of the demographically distinct group bias, which underlies biomedical research studies allowing the easy and fast creation of large, inclusive patient/citizen cohorts that foster investigation of a biomedical hypothesis (3). Additionally, gaining access to nationalized pharmaceutical and cause of death registries can provide useful information for phenotype curation (19). Nevertheless, the integration of national EHR with biobanks will allow the latter to carry out long-term follow-up research studies, whereas the results will not be limited to diseases for which the participants were originally assessed (21) but also to the updated and current health status of the participants. Not only that, but in (21), cost and time efficiency of EHR-linked biobanks are reported in multiple ways using the BioVU, an EHR-based biobank paradigm. Although large in scale, BioVU is not linked to a nationwide EHR. Thus, we can safely conclude that the cost-saving infrastructure of an EHR-integrated biobank grows dynamically as the EHR population reflects the whole country's citizens. This effect is even larger when the EHR and the biobank are designed from scratch to serve each other and synergistically improving citizens' health quality and standards.

GDPR requires that data subjects should be able to determine in advance what the scope and the consequences of their data processing are and should not be taken by surprise at a later point about the ways in which their personal and health data have been used (22). Thus, the transparency of the processing can be characterized as forming the underlying basis for any exercise of the rights of the data subject. The eHealthBioR allows the citizen to grant access to both biobanking and clinical data to specific entities (e.g., researchers) in a transparent framework under which the citizen has the control to view, monitor, and (whenever desired) withdraw the access rights that he/she provides for specific research studies. The citizen authorizes researchers to process his/her data for very specific purpose(s) in the context of a research study. The authorization is granted for a very specific study, which solely allows the researcher to use it for this purpose. Any other project requires another separate authorization.

The efficiency of the aforementioned EHR-based biobank services will be further leveraged by the development of computational methods, which will accurately extract data from clinical databases and link the data to DNA repositories (20), implementing preventive and predictive medicine.

All these tasks should be performed fulfilling the legal, technical, and financial frameworks of the national eHealth ecosystem, without diverting from the citizen-centered objective (12). To manage that, citizens should be able to give their consent to any biomedical study that they would like to participate in and to state the period of participation [e.g., broad consent forms (23)]. It is encouraging though that citizens are willing to enroll in such processes when they feel trust and that their privacy

is protected as well as when they see that related issues that may arise are successfully addressed (24). An integrated national eHealth ecosystem built with these principles and bridged with an effective communication strategy about a citizen's benefits that arise from his/her participation in such studies can increase the overall population's enrolment of the studies (4).

An Overview of Electronic Health Record Linked Biobanks

Recently, there is a plethora of biobanks, mentioned in the current bibliography, that work in partnership with national or private EHR systems. Comprehensive studies examining the established EHR-linked biobanks can be found in (19, 25, 26). A summary of the most popular biobank projects working with EHR is presented in **Table 1**.

The UK (including England, Wales, Scotland, and Northern Ireland) and Estonia seem to be two of the few countries that offer a single-payer-and-provider comprehensive healthcare system (43) covering the total of their population. As observed from **Table 1**, both of them have population-based biobanks coupled with a national EHR system. However, the coupling process is a continuously evolving ongoing process that is not completed for any of them. All other biobanks presented in **Table 1** are not linked with a national EHR system.

Yet all EHR-based biobanks of **Table 1** have made remarkable progress in EHR linking processes, which are worth to study and adopt/extend the best practices created and used by these leading initiatives and consortia. Moreover, the different challenges that these studies have faced in genomic test interpretation, understanding, and communication must be examined. For instance, issues related to missing data and lack of quality have been reported (26). This is a common problem in EHR-based biobanks and can happen for a variety of reasons. For example, specific tests are ordered only for specific disease-diagnosed citizens, while healthy people's EHR lack this information. Statistical and machine learning mechanisms are employed to deal with this problem (44–46). Actually, the rapid acceleration of statistical, computational, and machine learning method development urges the need to define their utilization perspective in genomics-aware EHRs (19). Furthermore, the development of a robust sustainable ethical/legal/social framework of the EHR-integrated biobanks must be formalized (5). Another open topic of discussion is about the potential bias inherited in EHR health data and how this bias confounds the analysis of biomedical studies (8). Bias related to EHR can be expressed by the loss to follow-up or the absence of clinical information for a study participant (21). EHR information bias may also apply for different data collection methods that can be used to record a specific health data measurement amongst different health providers (47). A critical challenge, which applies for the whole spectrum of biobanks, especially the EHR-based biobanks, at the international level is the establishment of data harmonization processes amongst biobanks (8). The need for harmonizing clinical sequencing and interpretation has been identified by the eMERGE Network and set as target for the network's phase III. By the end of 2019, they managed to harmonize two sequencing centers toward the technical and interpretive aspects of the clinical sequencing tests (48). Another effort for the establishment of globally scalable technology, policy, and procedures regarding the sharing of biospecimen and phenotypic data on wide consented cohorts has been proposed by Mandl et al. (49). Finally, a research topic of broad and current interest is the creation of interoperability standards describing management and sharing of genomic and clinical data amongst EHR-paired biobanks (2, 32). Although the importance of introducing interoperability in EHR (12) and in biobank systems (50) is recognized by the main stakeholders of healthcare systems, the actual progress toward true semantic interoperability has been slow, even for well-developed national healthcare information systems (51). "Enabling genomic data sharing for the benefit of human health" is the motto of the Global Alliance for Genomics and Health (GA4GH) network whose strategic plan involves progress acceleration of standards and frameworks for genomic data sharing aiming in responsible sharing of clinical-grade genomic data by 2022 (52). Ultimately, the employment of current and emerging interoperability standards that will enable EHRs' understanding of genetic/genomic data and biobanks understanding of clinical and phenotypic data might be the only way to reach a complete integration between EHR systems and biobanks.

The Impact of the Proposed Integrated eHealth Architecture Linking Electronic Health Records to Biobanks

Dealing with the challenges of using EHR data in research and linking them to biobanks will enable the systematic pairing of clinical studies (both observational and interventional) and healthcare provider-initiated examinations with the biobank specimens data for

- cost-effective longitudinal studies in genomic medicine through the utilization of the longitudinal character of the EHR data;
- theoretically lifetime duration of clinical trials and studies, which could observe the outcome of participants through EHRs for as long as the participants provide consent, without the need or cost of follow-up visits; and
- improvement of personalized clinical care by considering genetic/genomic implications to patient care throughout their clinical workflow to reach a clinical decision.

The novelty of this proposal is that both the proposed national integrated EHR and the population-based integrated biobank are based on two research projects, eHealth4U and CY-Biobank, which have recently initiated and they will have a nationwide impact in Cyprus. The eHealth4U project has initiated the design and the development of a national integrated EHR system where the CY-Biobank is designed to be a major national health resource aiming in supporting applied and basic research for improving health quality. The fact that the two projects run in parallel gives them the potential to incorporate, from the very beginning, the required EU and national standards and protocols in their structural design and to find the most suitable integration

TABLE 1 | A selection of major biobank schemes that work with EHRs.

Biobank	Country	Population-based biobank	Link to national EHR	Data types	References
eMerge Network https://emerge-network.org/	USA	False	False	Genotype/WES/WGS	(3)
Million Veteran Program (MVP) https://www.research.va.gov/mvp/	USA	False	False	Genotype/WES/WGS	(27)
DeCODE Genetics https://www.decode.com	Iceland	True	False	WGS	(28)
BioVU https://victr.vumc.org/biovu-description/	USA	False	False	Genotype	(29)
Michigan Genomics Initiative https://precisionhealth.umich.edu/michigangenomics/	USA	False	False	Genotype	(26, 30)
BioMe™ https://icahn.mssm.edu/research/ipm/programs/biome-biobank	USA	False	False	Genotype	(31)
Estonian Biobank (by Estonian Genome Center) https://genomics.ut.ee/en/about-us/estonian-genome-centre	Estonia	True	True	Genotype/WES/WGS	(23, 32)
Biobank Japan http://www.ims.riken.jp/english/projects/pj02.php	Japan	True	False	Genotype	(33)
China Kadoorie http://www.ckbiobank.org/site/	China	True	False	Genotype	(34)
DiscovEHR http://www.discovehrshare.com	USA	False	False	WES	(35, 36)
UK Biobank (UKBB) http://www.ukbiobank.ac.uk	UK	True	True	Genotype/WES	(37, 38)
Genomics England https://www.genomicsengland.co.uk/	UK	True for rare disease and their families, and patients with cancer	True	WGS	(39)
Generation Scotland: The Scottish Family Health Study (GS:SFHS) https://www.ed.ac.uk/generation-scotland	Scotland	True for families cohort	True	Genotype	(40–42)

WES, whole-exome sequencing; WGS, whole-genome sequencing.

architecture in the context of the eHealth national ecosystem. To do that, technological, legal, and interoperability elements must be crystallized in deep detail. A very first approach is presented in the following sections.

PROPOSED INTEGRATION FRAMEWORK EHEALTHBIOR

The objectives of eHealthBioR based on the policy modules of the framework are as follows:

- to support the efficient and near-real-time calculation of statistical power for prospective experiments, including providing incidence and prevalence of conditions utilizing privacy-preserving aggregate data approaches;
- to support monitoring linked to policy enforcement through centralized dashboards to ensure data security, ethical use of data, and support efficient policy evolution;
- to enable the automated long-term monitoring of consenting citizens' EHRs to research projects they have approved and tonsure national, international (GDPR) privacy legislation;
- to ensure secure and transparent data communication, storage, and analyses;
- to generate reports available to participating citizens on the use of their data across multiple studies, while enabling them to exercise their rights at any point in a way that builds confidence and supports engagement of citizens/patients in research; and
- to support in the long term the creation of participatory design studies that start from the main stakeholders, the patients, and to ensure that researchers focus on objectives prioritized by the patients and that the research maintains its focus on reaching conclusions maximizing the positive impact to patients.

Figure 1 presents how researchers in the biomedical domain are currently working regarding the design and implementation of their research studies in contrast to the way that they are expected to handle their research studies in the eHealthBioR

framework. More precisely, nowadays in Cyprus, an authorized researcher has access only to the existing biobanking data of a citizen who participates to a research study (gray line linking the researcher and biobank in **Figure 1**). With the implementation of eHealth4U (expected to finish in 2022), all the clinical healthcare data of the citizen shall be kept in the national EHR repository of Cyprus and will be under the control of the citizen (gray line linking the citizen and national EHR repository). However, with the current schema, if clinical healthcare data are required for the study, the researcher must contact the citizen in person to extract the extra clinical information needed (gray line linking the researcher and citizen). A different approach is proposed by the eHealthBioR framework, which will allow the registration and implementation of research studies (green arrow between the researcher and the eHealthBioR rectangle). The citizen can join the system and select to participate to a research study by giving an informed consent and then can monitor how the research study progresses through time (green arrows between the citizen and the eHealthBioR rectangle). The informed consent of a citizen dictates which citizen's data can be pulled from the two connected repositories, the population-based biobank and the national EHR Registry presented with green arrow between the two repositories and the eHealthBioR rectangle in **Figure 1**. Pulled citizen's data will then be provided to the selected research study for analysis and further processing using the technical modules while addressing legal and ethical challenges, enabling research, and empowering citizens to participate.

In the current model, researchers can apply to gain access to data through a biobank, and biobanks enroll patients through their healthcare providers or directly. Researchers may also collaborate with healthcare providers and also gain access to data or subjects for prospective studies through that interaction. However, in both cases, privacy-related legal and ethical challenges must be met, and this poses a significant threshold, that to overcome, and enable efficient sharing; in many cases, anonymization methodologies are chosen to limit or eliminate the possibility for the researcher to gain access to the patient's personal information. This introduces hurdles in the process, such as the need for the healthcare provider to de-anonymize data and evaluate legal and ethical aspects including risks increasing the cost. But most importantly, this approach works for collecting prospective data defined as part of clinical research forms, samples collected, etc., but it does not enable easy direct access to prospective data, anyway collected by healthcare professionals in EHRs. Furthermore, there is no universal way of enrolling citizens across multiple healthcare providers or gaining access to all data referring to subjects that are provided with healthcare by different carers (such as the case of comorbidities and the need to access different providers for specific conditions).

eHealthBioR sits in the middle of researchers, citizens, biobanks, and the National Electronic Health Registries, providing the relevant modules to both integrate the data sources, clean, and quality control, as well as the tools needed to ensure legal and ethical adherence and data access to both researchers and citizens.

The Case of Cyprus Biobank and Electronic Health Record Integration

Cyprus, being a small country (population 1 million) that was at the same time initiating a planned large biobank, as well as designing a national integrated EHR system, provided a unique opportunity to achieve the implementation of the proposed integration framework to act as a case study. The implementation of the prototype of the national integrated EHR system was initiated in October of 2019 (eHealth4U). At the same time, the design and the implementation of a population-based biobank (Cy-Biobank) in Cyprus also began. These two systems are expected to have a strong impact on Cypriot citizens' healthcare services. Additionally, researchers can also benefit from the deployment of these two systems, since they will be able to access biobanking and clinical information for the same citizen avoiding to ask the citizen to repeat questions, examinations, measurements, or any other healthcare information that is already included in the citizen's national integrated EHR or biobank profile. The proposed framework is the result of this work, and although it is designed as a generic framework able to be re-purposed across different national and international platforms, it is currently undergoing development based on the specific requirements of Cyprus. The ultimate objective is the production of reports demonstrating the specific cost, benefits, and challenges faced to support the development of platforms in other countries and settings.

The fact that the proposed framework involves the national integrated EHR, which is not just another commercial EHR system, should be considered as a very important factor for the success of the proposed framework. The national integrated EHR system will contain the minimum set of healthcare data required by the national medical associations in Cyprus to describe the health status of a citizen. The healthcare data will be given in an interoperable format, allowing other healthcare systems to communicate and exchange the total of the citizen's healthcare information. The contents of this core set of health data shall be determined by a decree of the Minister of Health, and all healthcare providers serving in Cyprus will be obligated to update in a responsible manner the EHR profiles of their patients.

The holders and operators of the healthcare systems and databanks will be regulated by the NeHA, which was established by law in 2019.[1] The national eHealth law defines the legal framework under which the healthcare providers will work. As stated in the relevant law, the NeHA will be responsible to monitor their compliance with the relevant national and European Union laws and relevant standards but also with the national eHealth objectives.

Consequently, the proposal of this paper is about the development of the underlying framework (eHealthBioR), which takes for granted the existence of the following:

(a) a national integrated EHR system with a Single e-Health Records Bank[1];
(b) a population-based biobank with a Healthcare Provider's Databank[1]; and

[1]No. 59 (I)/2019 Electronic Health Law of Cyprus.

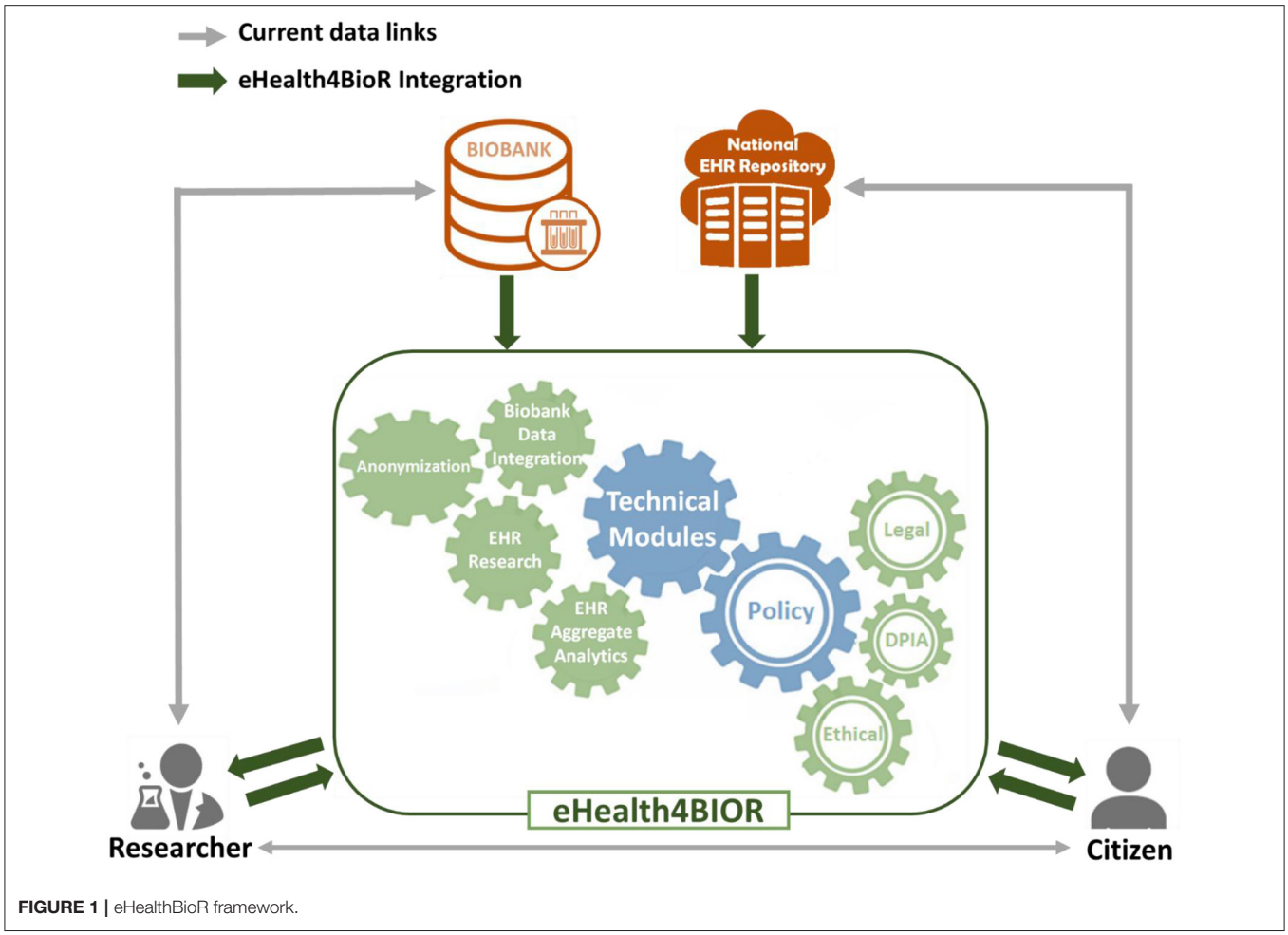

FIGURE 1 | eHealthBioR framework.

(c) a regulator monitoring all processes regarding healthcare data and databanks[1].

The fashion in which data are collected, stored, deleted, or retrieved for the Single e-Health Records Bank or Healthcare Provider's Databank is out of the scope of this paper.

The target of the eHealthBioR is to facilitate the implementation of research studies by providing access to healthcare data describing the same citizen pulled from both systems, the national integrated EHR Single e-Health Records Bank and the population-based biobank databank. Besides, as reported in (22), the requirement of providing (by the citizen) a specific consent for every research study that uses biological materials or personal data obtained during a medical intervention is stated in the Protocol on Biomedical research states[2] and confirmed in the EU Clinical Trials Regulation (No. 536/2014 EU)[3].

Based on this requirement, the eHealthBioR provides the following services to the citizen:

(a) View ongoing and future research studies that ask for participants.
(b) Become informed about each suggested research study at any time from any place for as long as he/she wishes. Per each research study, the following information will be available (53):

(i) purpose(s) of the research study,
(ii) a description of any reasonably foreseeable risks or discomforts to the subject or the society,
(iii) a description of any benefits to the subject or to others that may reasonably be expected from the research,
(iv) a statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained,
(v) whether the results of the research study might be used for commercial purposes,
(vi) entities involved in the research study (e.g., universities, laboratories, and pharmaceutical companies),
(vii) relevant legislations, EU directives, or any other legal or ethical binding regulations that characterize the research study,

[2]Explanatory Report, CETS 195, 78.
[3]Art. 28.2, EU Clinical Trials Regulation EU No. 536/2014.

(viii) research study phases indicating the steps taken in the context of each phase and the time points that a citizen can join the study, and
(ix) report explicitly the healthcare data that the research study asks for authorization to gain access (from both sources—EHR and biobank).

The eHealthBioR framework will allow the employment of multiple electronic media, including text, graphics, audio, video, podcasts, passive and interactive Web sites, biological recognition devices, and card readers, to convey information related to the study.

(c) Provide informed consent to a research study. The informed consent must be created following the guidelines given by European Commission in Ethics Review in FP7: Guidance for Applicants.[4] The effective and expiry dates of the consent will be clearly indicated. The citizen will be also informed about the right to revoke the consent at any time.
(d) View the studies that he/she participates in and the granted access rights.
(e) Monitor the progress of the research studies that he/she participates in. The citizen will have access to the research studies results, publications, media and press, commercialization of the results, etc.
(f) Revoke the consent from a research study that he/she participates in.

The citizen can do all the above from the comfort of his/her home or any other desired place, at any time, without feeling pressured by any person, place, and time limitations. The citizen becomes the key component of the research leading to the production of citizen-centered research with self-administered participants being informed about the research before their registration and continue learning about how the research develops through the whole lifetime of the study (given that the citizen does not withdraw the consent). The proposed framework perceives the citizen as the only controller of his/her own healthcare data (either stored in a biobank or the national integrated EHR) and the decision whether any other entity (e.g., researcher, research organization, and hospital) will be provided with access rights to a subset of his/her data belongs exclusively to the citizen. This is fundamental for the eHealthBioR framework in order to empower the citizens and thus create the potential to become a leading tool designed to optimize health research and citizen-centered care in practice. In that way, it is expected to increase citizens' enrollment rates to research studies as well as their recruitment and retention improving at the same time the communication of the research progress and development to the wider public educating citizens toward new scientific information on health and biomedical topics (e.g., treatments, cures, and prognosis). Therefore, while the main objective of the eHealthBioR is to create a wide range of capabilities for better understanding, increased motivation, and high engagement for the citizen, the researcher also benefits from it by increasing the efficiency of research implementation and improving the research quality.

The researcher entity (human or organization) is capable to do the following when using the eHealthBioR:

(a) Register a new research study in the eHealthBioR framework by providing the context of the research study specific informed consent that will be provided by the citizens that wish to participate:
   a. provide all the information required as presented in the eHealthBioR services to the citizen, point (b, i–ix);
   b. specify if the study requires identifiable subjects and to what extent;
   c. set minimum number of participants;
   d. set maximum number of participants; and
   e. set the duration of the study.
(b) For each of the registered research studies:
   a. view the list with the participants who provided the study-specific informed consent. If the study-specific informed consent defines anonymized subjects, then the list will be given in a format that will retain the anonymization feature; and
   b. access (read only) the healthcare data as explicitly declared in the study-specific informed consent provided by the citizen-participant.
(c) Close a research study, and, therefore, all the study-specific informed consent given by the participants will be revoked.

Acknowledging the complexity of designing and implementing a research study that combines biobanking data with clinical healthcare data, the challenge of the eHealthBioR is to provide a trustful environment for the citizen in which he/she will feel safe and confident to provide the researcher entity with access rights to his/her data by an intentional choice. The key factor to success is to provide the right tools to build the research study specific informed consent and the proper function of the framework ensuring that all processes using citizen's data are compliant with the context of the provided informed consent.

The citizen can join a research study following an opt-in model, as illustrated in **Figure 2**. In the eHealthBioR framework, each citizen has access to view information about all the active research studies, which are registered in the system by researchers at that particular time period. The citizen shall be able to select a research study he/she is interested in and study the information that is important for the citizen to know in order to provide an informed consent. Then, voluntarily, the citizen can provide the informed consent for the selected research study, which will enable his/her participation to the study.

The last step leads to the creation of an active consent, which will define explicitly what citizen's data can be drawn from the biobank and the national integrated EHR Repository to be used for the referenced research study for which the researcher is the owner and the citizen is the participant. **Figure 3** exhibits the importance of the active consent provided by the citizen to a researcher for the purposes of a specific research study.

[4]EUROPEAN COMMISSION, Ethics Review in FP7: Guidance for Applicants: Informed Consent. ftp://ftp.cordis.europa.eu/pub/fp7/docs/informed-consent_en.pdf.

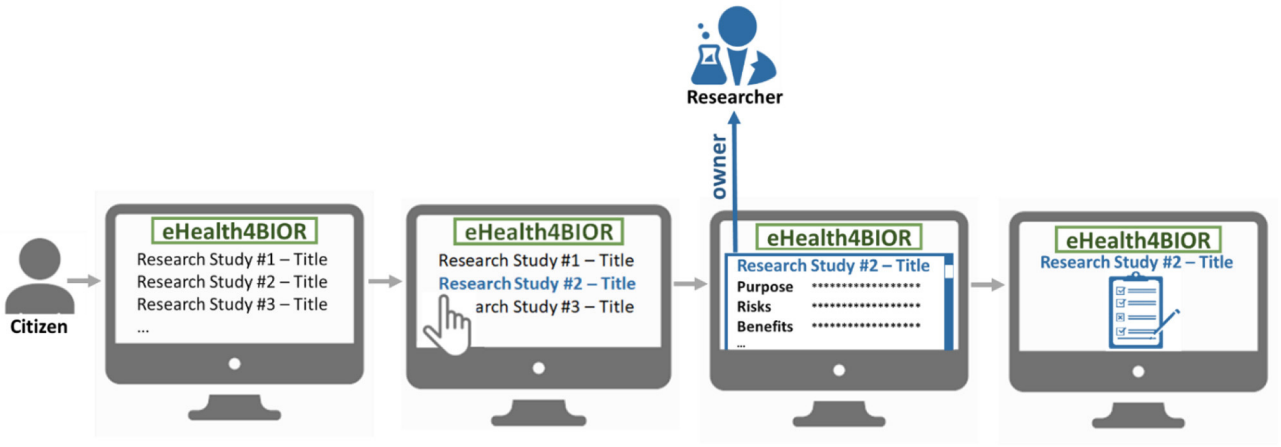

FIGURE 2 | The process of the voluntary participation of a citizen to a research study under the eHealthBioR framework.

Whenever the researcher wishes to process the citizen's data with the technical modules provided by the eHealthBioR framework (**Figure 1**), the active consent will be used to form the final set of the biobanking and clinical citizen's data retrieved from the two repositories as presented in **Figure 3**. In the same figure, the researcher appears to interact with eHealthBioR since he/she will be able not only to view but also to process the citizen's dataset by using the technical modules provided by the eHealthBioR framework. Accordingly, the citizen is presented to receive information from the eHealthBioR, which allows a participant of a particular research study to follow and get updates on the results, the outcomes, and any other related activity.

The implementation of the eHealthBioR e-consent must be designed in such a way as to promote the participating researchers to fulfill the general requirements of a study-specific informed consent (53) using understandable language to provide the relevant information. At the same time, they should be promoted to avoid the use of exculpatory statements, which waive or appear to waive the participant's legal rights.

The Electronic Health Record Research Module

EHR-sourced data represent a national pool of citizens' phenotyping information that represents the state of the health of citizens through time. Biobanks on the other hand act as biospecimens warehouse coupled with research quality data, which can be used to implement biomedical studies toward preventive, predictive, and therapeutic knowledge discovery, or intervention/diagnostic testing. Still, clinical information is very important for biobanking studies, and therefore, biobanking data should be linked to citizen's EHR data. In that way, biobanks gain access to a huge amount of up-to-date information about the participants to their studies, creating holistic datasets to address complex diseases and conditions and enabling long-term monitoring in longitudinal studies with minimal cost. The EHR and biobank integration model offers a tremendous opportunity for the biobanks to explore health information for the citizens of their cohorts for their life beginning 9 months before their birth until their current age or their death.

The EHR research module will enable the direct access of relevant data of subjects who have provided an informed consent for an approved study longitudinally, designed as a method to monitor patients in the long term. Nevertheless, the patients should be informed about how and what their data are used for and enable them to withdraw their consent at any point through the national EHR platform. This module can work as a catalyst supporting research activities such as biobanking, by both enabling the longitudinal monitoring of the patient through their EHRs and by enabling the user to have knowledge of when, how, and why their data are used with the ability to withdraw consent at any point.

This will include an API that will allow the long-term monitoring of the patients through the EHRs for research purposes while empowering the patient with the ability to withdraw consent or exercise any of their other rights as a subject in the study at any time through the system.

The API will authenticate the research entity requesting the data and only provide it access to citizens who have completed all policy requirements to participate in that study and only data that have been approved as part of the relevant informed consent and bioethics approval.

The Electronic Health Record Aggregate Analytics Module

The aggregate analytics module will perform in near-real-time aggregation and preliminary statistical analyses of all data in the EHR system focusing on disease prevalence and incidence. As this module will only provide aggregate data, the generated outcomes will not constitute personal data, and therefore, the requirements for access to data generated by this module will be limited. Care will be taken to ensure de-anonymization is not possible by utilizing well-established methods including perturbation and data suppression (16, 18).

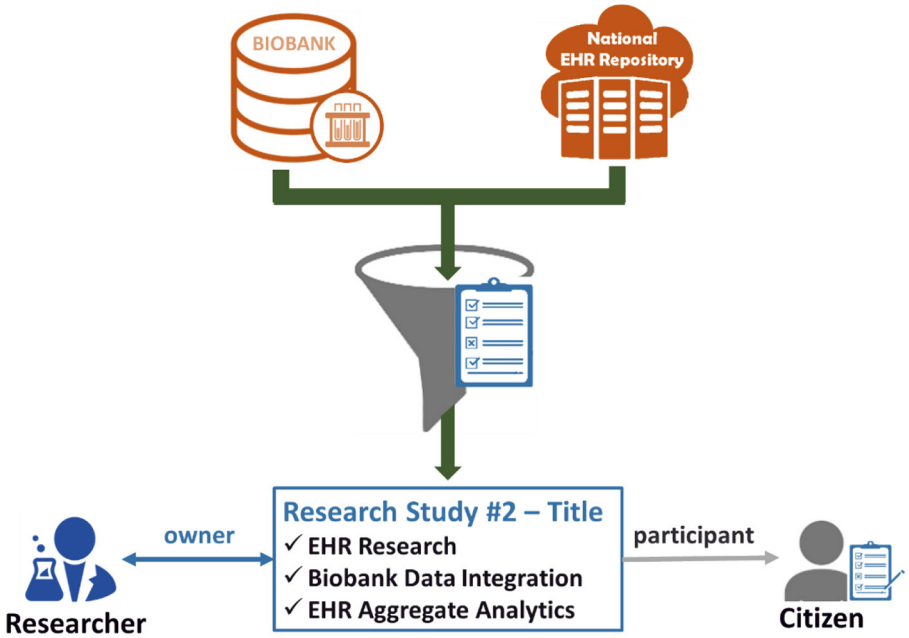

FIGURE 3 | Citizen's biobanking and national electronic health record (EHR) data flow controlled by the citizen's informed consent.

This will support governmental organizations at the national (Ministry of Health) as well as the international (World Health Organization) to have direct access enhancing their activities, and early identification of issues, such as changes in the incidence of infectious diseases. Furthermore, it will help guide research by the identification of the most prominent research challenges (diseases with sudden recent increases or high overall incidence). Additionally, it is easier to identify the subjects that could be informed of a future research and asked to provide their consent to participate, provided that the legal and ethical challenges are addressed through proper channels.

The Biobank Data Integration Module

Currently, there are independent legacy EHR systems across different healthcare providers; however, the number of these that support international standards for encoding and communication is increasing with support mandated at the policy level across many countries.

To achieve integration with EHRs on a biobank's side, bi-directional communication protocols need to be developed through secure APIs. This module will support both the pulling of data from existing external to the biobank EHR systems and the integration of these data with the biobank's own sample databases. This process will enable the long-term monitoring of the patient and will provide bi-directional communication informing the EHR system (if it supports this functionality) with what data have been produced from the biological sample provided to the biobank, focusing on data that are or may become in the near future relevant to the healthcare provider's team. Examples include somatic or cell line mutations in cancer and patients detected through sequencing of samples provided to the biobank that, although performed as part of research projects, may help identify optimal secondary treatments. This will enable the biobank to enrich the data of patients who provide samples.

DISCUSSION

There is great research potential in EHR records, but there are also key challenges. These challenges can all be traced back to the purpose of EHRs, that is, to support the individual citizens' healthcare and to some extent health insurance provision not to enable research. Already multiple research initiatives have successfully demonstrated that EHRs can and should be used in research, as they provide a low-cost unique longitudinal data source spanning potentially the entire life of the subjects in the studies. It is also been demonstrated that it is possible to address to a sufficient extent the challenges that EHRs pose to be reused in research, related to missing data, quality of data, or the diverse potential sources of data. However, to ensure legal and ethical compliance, without any data loss, and to simultaneously enable the integration of data from different sources of the same subjects (such as the case of patients providing samples and data to biobanks), following techniques of anonymization is the wrong approach. By definition these rely on destroying the link between the data and the individual, but that is what we most want to preserve in order to link the EHR to the structured electronic case report forms (eCRFs) and data collected through the biobank.

Thus, what is proposed in this paper is an approach that puts the citizen-patient in the driver's seat, enabling them to quickly review and take informed decisions about providing consent for specific research experiments, monitor the use, and empower them to practice their legal rights. This way, research participants

are empowered to join, maintain their interest in the study, and understand how, when, and why their data are used. Most importantly, they can engage with the research in a meaningful way, including observing potential research outcomes relevant to their care in the long term. We suggest that this eHealthBioR or similar integrations of biobanks to EHR registries may act as a key enablers in engaging more citizens, healthy or patients, in research, while also increasing adherence to the study directions.

CONCLUSION

In this paper, we outlined the state of the art in data-reuse from EHRs and analyzed that although challenging to use in research, EHRs pose a tremendous opportunity, especially for long-term longitudinal studies. However, current approaches that attempt to solve the privacy aspects through anonymity or federated executions, suffer from a major disadvantage in that they do not allow for external to the EHR registry data silos that contain information on the same subjects to be integrated in the EHRs. Furthermore, the legal and ethical challenges associated with access health data, especially in the case of molecular level data such as genetics that are inherently personal data, pose a significant hurdle to research. We proposed a methodology that addresses these challenges, and we suggest a framework that integrates a national integrated EHR system with biobanks within a legal and ethical context to support the citizens to provide their data for research purposes. In this way, the strict adherence to legal and ethical requirements is ensured as well as the empowerment sense of the participants since they are enabled to exercise their legal rights. We provide a first draft of how the integration of such modules could be achieved through the proposed eHealthBioR.

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

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work presented in this paper is partially funded by the Cy-Biobank project, which receives funding from the European Union's, Horizon 2020 research and Innovation program, under Grant Agreement No. 857122, and the eHealth4U project which is co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: INTEGRATED/0916/0030).

REFERENCES

1. Williams MS, Buchanan AH, Davis FD, Faucett WA, Hallquist MLG, Leader JB, et al. Patient-centered precision health in a learning health care system: Geisinger's genomic medicine experience. *Health Aff (Millwood)*. (2018) 37:757–64. doi: 10.1377/hlthaff.2017.1557
2. Ullman-Cullere M, Mathew JP. Emerging landscape of genomics in the electronic health record for personalized medicine. *Hum Mutat*. (2011) 32:512–6. doi: 10.1002/humu.21456
3. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The electronic medical records and genomics (eMERGE) network: past, present, and future. *Genet Med*. (2013) 15:761–71. doi: 10.1038/gim.2013.72
4. Hall JL, Ryan JJ, Bray BE, Candice B, David L, Kristin NL, et al. Merging electronic health record data and genomics for cardiovascular research. *Circ Cardiovasc Genet*. (2016) 9:193–202. doi: 10.1161/HCG.0000000000000029
5. Smoller JW, Karlson EW, Green RC, Kathiresan S, Macarthur DG, Talkowski ME, et al. An eMERGE clinical center at partners personalized medicine. *J Pers Med*. (2016) 6:1–10. doi: 10.3390/jpm6010005
6. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. (2017) 19:249–55. doi: 10.1038/gim.2016.190
7. Shoenbill K, Fost N, Tachinardi U, Mendonca EA. Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. *J Am Med Inform Assoc*. (2013) 21:171–80. doi: 10.1136/amiajnl-2013-001694
8. Abul-Husn NS, Kenny EE. Personalized medicine and the power of electronic health records. *Cell*. (2019) 177:58–69. doi: 10.1016/j.cell.2019.02.039
9. Davies K. The era of genomic medicine. *Clin Med*. (2013) 13:594–601. doi: 10.7861/clinmedicine.13-6-594
10. Tasa T, Krebs K, Kals M, Mägi R, Lauschke VM, Haller T, et al. Genetic variation in the Estonian population: pharmacogenomics study of adverse drug effects using electronic health records. *Eur J Hum Genet*. (2019) 27:442–54. doi: 10.1038/s41431-018-0300-6
11. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. (2017) 106:1–9. doi: 10.1007/s00392-016-1025-6
12. Schiza EC, Kyprianou TC, Petkov N, Schizas CN. Proposal for an eHealth based ecosystem serving national healthcare. *IEEE J Biomed Health Inf*. (2019) 23:1346–57. doi: 10.1109/JBHI.2018.2834230
13. MD G, Linda. *Why Doctors Hate Their EHRs*. (2017). Available online at: http://www.physicianspractice.com/ehr/why-doctors-hate-their-ehrs, 2019
14. Persaud N. A national electronic health record for primary care. *CMAJ*. (2019) 191:E28–9. doi: 10.1503/cmaj.181647
15. Simkus R. Physicians need to participate in developing standards for electronic health records. *CMAJ*. (2019) 191:E588. doi: 10.1503/cmaj.71920
16. Antoniades A, Aristodimou A, Georgousopoulos C, Forgó N, Gledson A, Hasapis P, et al. Advancing clinical research by semantically interconnecting aggregated medical data information in a secure context. *Health Technol*. (2017) 7:223–40. doi: 10.1007/s12553-017-0188-0
17. Aristodimou A, Antoniades A, Pattichis CS. Privacy preserving data publishing of categorical data through k-anonymity and feature selection. *Healthcare Technol Lett*. (2016) 3:16–21. doi: 10.1049/htl.2015.0050
18. Antoniades A, Keane J, Aristodimou A, Philipou C, Constantinou A, Tozzi F, et al. The effects of applying cell-suppression and perturbation to aggregated genetic data. In: *IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)* (2012). doi: 10.1109/BIBE.2012.6399777
19. Wolford BN, Willer CJ, Surakka I. Electronic health records: the next wave of complex disease genetics. *Hum Mol Genet*. (2018) 27:R14–21. doi: 10.1093/hmg/ddy081
20. Wells QS, Gupta DK, Smith JG, Collins SP, Storrow AB, Ferguson J, et al. Accelerating biomarker discovery through electronic health records,

automated biobanking, and proteomics. *J Am Coll Cardiol*. (2019) 73:2195–205. doi: 10.1016/j.jacc.2019.01.074
21. Bowton E, Field JR, Wang S, Schildcrout JS, Van Driest SL, Delaney JT, et al. Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med*. (2014) 6:234m3. doi: 10.1126/scitranslmed.3008604
22. Salokannel M, Tarkkala H, Snell K. Legacy samples in Finnish biobanks: social and legal issues related to the transfer of old sample collections into biobanks. *Hum Genet*. (2019) 138:1287–99. doi: 10.1007/s00439-019-02070-0
23. Leitsalu L, Alavere H, Tammesoo M, Leego E, Metspalu A. Linking a population biobank with national health registries-the estonian experience. *J Pers Med*. (2015) 5:96–106. doi: 10.3390/jpm5020096
24. Antommaria AHM, Brothers KB, Myers JA, Feygin YB, Aufox SA, Brilliant MH, et al. Parents' attitudes toward consent and data sharing in biobanks: a multisite experimental survey. *AJOB Empir Bioeth*. (2018) 9:128–42. doi: 10.1080/23294515.2018.1505783
25. Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J*. (2017) 39:1481–95. doi: 10.1093/eurheartj/ehx487
26. Beesley LJ, Salvatore M, Fritsche LG, Pandit A, Rao A, Brummett C, et al. The emerging landscape of health research based on biobanks linked to electronic health records: existing resources, statistical challenges, and potential opportunities. *Stat Med*. (2020) 39:773–800. doi: 10.1002/sim.8445
27. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million veteran program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. (2016) 70:214–23. doi: 10.1016/j.jclinepi.2015.09.016
28. Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, et al. Whole genome characterization of sequence diversity of 15,220 Icelanders. *Sci Data*. (2017) 4:170115. doi: 10.1038/sdata.2017.115
29. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Trans Sci*. (2010) 3:42–8. doi: 10.1111/j.1752-8062.2010.00175.x
30. Fritsche LG, Gruber SB, Wu Z, Schmidt EM, Zawistowski M, Moser SE, et al. Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from the Michigan genomics initiative. *Am J Hum Genet*. (2018) 102:1048–61. doi: 10.1016/j.ajhg.2018.04.001
31. Belbin GM, Wenric S, Cullina S, Glicksberg BS, Moscati A, Wojcik GL, et al. Towards a fine-scale population health monitoring system. *bioRxiv*. (2019) 184:780668. doi: 10.1101/780668
32. Milani L, Leitsalu L, Metspalu A. An epidemiological perspective of personalized medicine: the Estonian experience. *J Intern Med*. (2015) 277:188–200. doi: 10.1111/joim.12320
33. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, et al. Overview of the BioBank Japan project: study design and profile. *J Epidemiol*. (2017) 27:S2–8. doi: 10.1016/j.je.2016.12.005
34. Chen Y, Wright N, Guo Y, Turnbull I, Kartsonaki C, Yang L, et al. Mortality and recurrent vascular events after first incident stroke: a 9-year community-based study of 0.5 million Chinese adults. *Lancet Global Health*. (2020) 8:e580–90. doi: 10.1016/S2214-109X(20)30069-3
35. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science*. (2016) 354:aaf6814. doi: 10.1126/science.aaf6814
36. Staples J, Maxwell EK, Gosalia N, Gonzaga-Jauregui C, Snyder C, Hawes A, et al. Profiling and leveraging relatedness in a precision medicine cohort of 92,455 exomes. *Am J Hum Genet*. (2018) 102:874–89. doi: 10.1016/j.ajhg.2018.03.012
37. Atkinson MD, Kennedy JI, John A, Lewis KE, Lyons RA, Brophy ST, et al. Development of an algorithm for determining smoking status and behaviour over the life course from UK electronic primary care records. *BMC Med Inf Decis Making*. (2017) 17:2. doi: 10.1186/s12911-016-0400-6
38. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JX, Ye B, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*. (2020) 586:749–56. doi: 10.1038/s41586-020-2853-0
39. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ*. (2018) 361:k1687. doi: 10.1136/bmj.k1687
40. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet*. (2006) 7:74. doi: 10.1186/1471-2350-7-74
41. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort profile: generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. (2012) 42:689–700. doi: 10.1093/ije/dys084
42. Kerr S, Campbell A, Marten J, Vitart V, McIntosh A, Porteous D, et al. Electronic health record and genome-wide genetic data in Generation Scotland participants [version 1; peer review: 2 approved, 1 approved with reservations]. *Wellcome Open Res*. (2017) 2:85. doi: 10.12688/wellcomeopenres.12600.1
43. Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, et al. UK phenomics platform for developing and validating EHR phenotypes: CALIBER. *J Am Med Informatics Association*. (2019) 26:1545–59. doi: 10.1093/jamia/ocz105
44. Beaulieu-Jones B, Moore Jh. Missing data imputation in the electronic health record using deeply learned autoencoders. *Biocomputing*. (2017) 22:207–18. doi: 10.1142/9789813207813_0021
45. Beaulieu-Jones B, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and managing missing structured data in electronic health records: data analysis. *JMIR Med Inf*. (2018) 6:e11. doi: 10.2196/medinform.8960
46. Basile AO, Ritchie MD. Informatics and machine learning to define the phenotype. *Exp Rev Mol Diagn*. (2018) 18:219–26. doi: 10.1080/14737159.2018.1439380
47. Bower JK, Patel S, Rudy JE, Felix AS. Addressing bias in electronic health record-based surveillance of cardiovascular disease risk: finding the signal through the noise. *Curr Epidemiol Rep*. (2017) 4:346–52. doi: 10.1007/s40471-017-0130-z
48. Zouk H, Venner E, Lennon NJ, Muzny DM, Abrams D, Adunyah S, et al. Harmonizing clinical sequencing and interpretation for the eMERGE III network. *Am J Hum Genet*. (2019) 105:588–605. doi: 10.1016/j.ajhg.2019.07.018
49. Mandl KD, Glauser T, Krantz ID, Avillach P, Bartels A, Beggs AH, et al. The genomics research and innovation network: creating aninteroperable, federated, genomics learning system. *Genet Med*. (2020) 22:371–80. doi: 10.1038/s41436-019-0646-3
50. Alterovitz G, Heale B, Jones J, Kreda D, Lin F, Liu L, et al. FHIR Genomics: enabling standardization for precision medicine use cases. *NPJ Genom Med*. (2020) 5:13. doi: 10.1038/s41525-020-0115-6
51. Warren LR, Clarke J, Arora S, Darzi A. Improving data sharing between acute hospitals in England: an overview of health record system distribution and retrospective observational analysis of inter-hospital transitions of care. *BMJ Open*. (2019) 9:e031637. doi: 10.1136/bmjopen-2019-031637
52. Global Alliance. *Global Alliance for Genomics and Health*. Available online at: https://www.ga4gh.org/about-us/ (accessed June, 2020).
53. Maloy JW, Bass PF, III. Understanding broad consent. *Ochsner J*. (2020) 20:81–6. doi: 10.31486/toj.19.0088

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership