frontiers | Research Topics

Mobile and wearable systems for health monitoring

Edited by
Mohamed Elgendi, Richard Ribon Fletcher, Derek Abbott, Dingchang Zheng, Panicos Kyriacou and Carlo Menon

Published in
Frontiers in Digital Health

ISSN 1664-8714
ISBN 978-2-8325-2341-4
DOI 10.3389/978-2-8325-2341-4

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

Mobile and wearable systems for health monitoring

Topic editors

Mohamed Elgendi — ETH Zürich, Switzerland
Richard Ribon Fletcher — Massachusetts Institute of Technology, United States
Derek Abbott — University of Adelaide, Australia
Dingchang Zheng — Coventry University, United Kingdom
Panicos Kyriacou — City University of London, United Kingdom
Carlo Menon — ETH Zürich, Switzerland

Citation

Elgendi, M., Fletcher, R. R., Abbott, D., Zheng, D., Kyriacou, P., Menon, C., eds. (2023). *Mobile and wearable systems for health monitoring*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2341-4

Table of contents

frontiers | Frontiers in Digital Health

TYPE Editorial
PUBLISHED 20 April 2023
DOI 10.3389/fdgth.2023.1196103

OPEN ACCESS

EDITED AND REVEIWED BY
Uwe Aickelin,
The University of Melbourne, Australia

*CORRESPONDENCE
Mohamed Elgendi
✉ moe.elgendi@hest.ethz.ch

Carlo Menon
✉ carlo.menon@hest.ethz.ch

RECEIVED 29 March 2023
ACCEPTED 04 April 2023
PUBLISHED 20 April 2023

CITATION
Elgendi M, Fletcher RR, Abbott D, Zheng D, Kyriacou P and Menon C (2023) Editorial: Mobile and wearable systems for health monitoring.
Front. Digit. Health 5:1196103.
doi: 10.3389/fdgth.2023.1196103

Editorial: Mobile and wearable systems for health monitoring

Mohamed Elgendi[1]*, Richard Ribon Fletcher[2], Derek Abbott[3], Dingchang Zheng[4], Panicos Kyriacou[5] and Carlo Menon[1]*

[1]Biomedical and Mobile Health Technology Lab, ETH Zurich, Zurich, Switzerland, [2]Mobile Technology Group, Department of Mechanical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, United States, [3]School of Electrical and Mechanical Engineering, University of Adelaide, Adelaide, NSW, Australia, [4]Research Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom, [5]School of Engineering, City University of London, London, United Kingdom

KEYWORDS

mHealth, wearables, remote health monitoring, health data science, health analytics, digital health, mobile computing, Artificial intelligence in healthcare

Editorial on the Research Topic
Mobile and wearable systems for health monitoring

Over the past decade, there has been a growing interest in using mobile and wearable systems for healthcare applications. To contribute to this field, our Research Topic on *Mobile and Wearable Systems for Health Monitoring* has collected a variety of contributions, ranging from original research to reviews and perspectives, all focused on exploring new or existing methods, protocols, and models for health monitoring. These contributions cover various topics, including machine learning algorithms, data quality, and using smartphones and wearable devices to assess COVID and mental health. Additionally, our initiative encouraged multidisciplinary approaches to explore innovative health applications, resulting in exciting studies that examine new wearable and mobile devices in a hospital setting. Our Research Topic aims to provide insights into the potential of mobile and wearable systems for healthcare applications and inspire further research in this field.

The COVID-19 pandemic emphasized the importance of leveraging digital infrastructure for remote patient monitoring, which could be facilitated through the use of wearable sensors. The review article by Seshadri et al. showed that by using predictive platforms, these devices could potentially aid in disease detection and monitoring on an individual and population level. Public health officials and researchers could also use anonymous data to track and mitigate the spread of the virus. Their manuscript highlights the potential of clinically relevant physiological metrics from commercial devices in monitoring the health, stability, and recovery of COVID-19+ individuals and front-line workers. This paper helps encourage front-line workers and engineers to develop digital health platforms for monitoring and managing the pandemic.

It is challenging to evaluate the reliability of a wearable device in healthcare. The work by Santos et al. discusses how a wearable ambulatory monitoring system was optimized to monitor COVID-19 patients in isolation wards in the United Kingdom. The system used a chest patch and pulse oximeter to estimate and transmit continuous vital sign data from patients to remote nurse bays, minimizing the risk of infection for nursing staff. The system operated through a secure web-based architecture and fault-tolerant software strategies, allowing for remote monitoring of patients. The plan was used for almost half

of the patients in the isolation ward during the peak of hospital admissions in the local area and was found to be effective in monitoring patients. An updated version of the system has also been used in subsequent waves of the pandemic in the UK. The implementation of wearable ambulatory monitoring systems represents a crucial step towards a safer and more effective way to monitor COVID-19 patients in isolation wards, significantly minimizing the risk of infection for nursing staff and opening a promising path for the future of healthcare.

As the use of wearable sensors grows, researchers have developed a range of wearable devices that collect data on our daily activities, such as movement, sleep duration, heart rate, skin temperature, and more, to monitor and analyze our mental health. The data collected by these devices are translated into patterns that can indicate symptoms of mental health disorders, such as depression, anxiety, and stress. Machine learning helps identify behavioral markers and the relationship between the raw sensor data and mental health conditions. The review by Sheikh et al. discusses the available smartphone-based, wearable, and environmental sensors that can be used to detect mental health conditions and provide a helpful tool for managing and treating them.

Cardiovascular disease, which is the leading cause of mortality worldwide, remains a popular field of research for new wearable sensor algorithms. Electrocardiogram (ECG) signals continue to be used throughout medicine and researchers have been enhancing the accuracy of automated heart disease diagnosis by exploring mathematical feature transformations for ECG signal segments. Liang et al.'s team tested six different mathematical transformation methods using 10-second ECG segments and found that applying the reciprocal transformation resulted in consistently better classification performance for normal and abnormal heartbeats. The cubic transformation was the second-best in terms of heartbeat detection accuracy. Surprisingly, the commonly used logarithmic transformation did not perform as well as the reciprocal or cubic transformations. By using the optimal transformation method, the reciprocal transformation, doctors can improve the accuracy of detecting normal and abnormal heartbeats by 35.6%. The researchers concluded that adding a simple data transformation step, such as the reciprocal or cubic, to the extracted features can significantly enhance current automated heartbeat classification, leading to better and faster diagnoses of heart diseases.

In addition to advances in algorithms, new data collection methods have also begun to emerge for cardiovascular disease, such as the regular monitoring of blood pressure for detecting hypertension and reducing the risk of cardiovascular disease. Since the traditional cuff-based method can be inconvenient and discourage regular monitoring. Smartphone sensors, such as video cameras, can detect arterial pulsations and be used to assess cardiovascular health. Researchers have developed advanced image processing and machine learning techniques to predict blood pressure using smartphones or video cameras. The review by Steinman et al. discusses the challenges associated with using smartphones in homes and clinics, but further testing is necessary under different conditions. This research shows that smartphones and video cameras have the potential to measure multiple cardiovascular metrics beyond blood pressure, which could significantly reduce the risk of cardiovascular disease.

The traditional functions of wearable sensors to measure physical activity and sleep (actigraphy) has now been extended to mobile and portable platforms that can be used in the home environment. Despite the recognized importance of sleep and circadian rhythm, conventional polysomnography (PSG) monitoring requires specialized equipment, making it unrealistic to assess the sleep stage in the home environment. The use of camera-based methods for sleep monitoring has increased in recent years; however, the published studies have mainly focused on adults. Kamon et al. developed an infrared camera-based monitoring system for children between 0–6 years old. Extremely randomized decision trees ("Extra Trees"), an ensemble machine learning algorithm, were used to estimate the sleep stages from various information extracted from body movements. Comparable performance was achieved compared to simple PSG scoring, suggesting that their system could be potentially used as a non-contact sleep monitoring system for children at home.

Photoplethysmography (PPG) continues to be popular technique widely used in Digital and Wearable Health monitoring. However, there has yet to be a published consensus on signal quality expectations, especially for morphological PPG pulse analysis. Huthart et al. conducted a signal quality expectation survey with fellow international researchers in skin contact PPG measurements. They determined a consensus regarding the minimum recording length, the minimum number of undistorted pulses required, and the threshold proportion of noisy beats needed for the recording rejection. Their study provided initial recommendations to support the need to move toward improved standardization in measurement protocol and morphological pulse wave analysis and the need to gather repeatable and meaningful PPG data for implementing PPG sensing technology in Digital and Wearable Health devices.

In critical care settings, patients require continuous monitoring, but current methods do not capture important functional and behavioral indices. Advances in non-invasive sensing technology and deep learning techniques can transform patient monitoring by enabling continuous and granular monitoring of critical care measures. The paper by Davoudi et al. highlights current approaches to pervasive sensing in acute care, identifies limitations and opportunities, and emphasizes the potential of pervasive sensing technology to improve patient outcomes by enabling real-time adaptation of pain medications, personalizing analgesia choice, and improving assessments of delirium and mobility.

As another example of home-based health monitoring, Laidig et al. proposed a set of algorithms that use two foot-worn IMUs to accurately determine spatiotemporal gait parameters essential for clinical gait assessment without requiring magnetometers or precise sensor mounting. The proposed methods offer a calibration-free and unsupervised approach to gait assessment in daily-life environments. The algorithms are validated on a broad dataset of healthy subjects and orthopedic and neurological patients walking on a treadmill, showing a

strong correlation and high accuracy for walking speeds and pathologies.

Folkvord et al. described a research study investigating the effectiveness of self-tracking apps on psychological outcomes related to self-empowerment and better-informed lifestyle decision-making. The study will consist of three parts: a systematic review of experimental evidence, a longitudinal field experiment comparing exercise programs with and without the aid of the self-tracking app Strava, and interviews with a subset of participants to gather qualitative data. The authors hope this study will provide a better understanding of the effects of self-tracking apps on psychological outcomes and inform the development of more effective health and activity monitoring tools.

A study by Kaye et al. examined changes in Chronic obstructive pulmonary disease (COPD) patients' COPD Assessment Test (CAT) scores and short-acting beta-agonists (SABA) inhaler use over six months in a digital health program that provided electronic medication monitors (EMMs) and a smartphone app. The results showed that CAT scores improved by a mean of −0.9 points, and SABA use decreased by −0.6 puffs per day, with more remarkable improvement observed in patients with a higher disease burden. These findings suggest that passively collected data from EMMs can be used to monitor disease burden and treatment outcomes in COPD patients.

In another application of mobile technology to pulmonary medicine, an article by Dave and Gupta discussed machine learning (ML) in contact tracing apps during the COVID-19 pandemic. ML can use data from these apps to forecast virus spread and identify vulnerable groups. Still, it's essential to ensure the accuracy, reliability, and lack of biases in the dataset to make reliable predictions. The article presents two requirements to meet international data quality standards for ML. It identifies where these requirements can be met, given varying contact tracing apps and smartphone usage in different countries. Lastly, this approach's advantages, limitations, and ethical considerations are discussed.

As we see the use of wearable and digital technologies being applied to the treatment and monitoring of mental health disorders—particularly for outpatient care—we are starting to see more research exploring how wearable data can be used to inform mental health treatment. A study by Abbas et al. aimed to examine whether machine learning-based visual and auditory digital markers can quantify response to antidepressant treatment (ADT) with selective serotonin reuptake inhibitors (SSRIs) and serotonin–norepinephrine uptake inhibitors (SNRIs). The researchers used automated smartphone tasks to measure facial, vocal, and head movement characteristics across four weeks of treatment for MDD patients. Results show that digital markers associated with MDD demonstrate validity as measures of treatment response, with significant changes in the MADRS and multiple digital markers, including facial expressivity, head movement, and amount of speech.

In conclusion, the recent advancements in mobile and wearable devices have opened tremendous opportunities to improve practical health applications and enable more accurate clinical decision-making. These technologies have greatly enhanced clinical capabilities, from gait assessment to blood pressure monitoring and COVID signs detection. Applying machine learning and statistical models to different health applications has shown promising results. The progress in mobile and wearable devices highlights the central role that model development will play in shaping the future of healthcare. This multidisciplinary research topic covers a range of areas, including machine learning, signal quality, patient reporting, mobile contact tracing, virtual assessment, and computational approaches for health monitoring.

Author contributions

ME led the editorial. ME, RRF, DA, DZ, PK, and CM provided valuable feedback. All authors have consented to the publication of the editorial. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

frontiers in Digital Health

REVIEW
published: 23 June 2020
doi: 10.3389/fdgth.2020.00008

Wearable Sensors for COVID-19: A Call to Action to Harness Our Digital Infrastructure for Remote Patient Monitoring and Virtual Assessments

Dhruv R. Seshadri[1], Evan V. Davies[2], Ethan R. Harlow[3], Jeffrey J. Hsu[4], Shanina C. Knighton[1,5], Timothy A. Walker[1], James E. Voos[3] and Colin K. Drummond[1]*

[1] Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, United States, [2] Department of Electrical Engineering, Case Western Reserve University, Cleveland, OH, United States, [3] Department of Orthopaedics, University Hospitals of Cleveland Medical Center, Cleveland, OH, United States, [4] Division of Cardiology, Department of Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, United States, [5] Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, United States

The COVID-19 pandemic has brought into sharp focus the need to harness and leverage our digital infrastructure for remote patient monitoring. As current viral tests and vaccines are slow to emerge, we see a need for more robust disease detection and monitoring of individual and population health, which could be aided by wearable sensors. While the utility of this technology has been used to correlate physiological metrics to daily living and human performance, the translation of such technology toward predicting the incidence of COVID-19 remains a necessity. When used in conjunction with predictive platforms, users of wearable devices could be alerted when changes in their metrics match those associated with COVID-19. Anonymous data localized to regions such as neighborhoods or zip codes could provide public health officials and researchers a valuable tool to track and mitigate the spread of the virus, particularly during a second wave. Identifiable data, for example remote monitoring of cohorts (family, businesses, and facilities) associated with individuals diagnosed with COVID-19, can provide valuable data such as acceleration of transmission and symptom onset. This manuscript describes clinically relevant physiological metrics which can be measured from commercial devices today and highlights their role in tracking the health, stability, and recovery of COVID-19+ individuals and front-line workers. Our goal disseminating from this paper is to initiate a call to action among front-line workers and engineers toward developing digital health platforms for monitoring and managing this pandemic.

Keywords: wearable sensors, COVID-19, pandemic, predictive analytics, remote patient monitoring

OPEN ACCESS

Edited by:
Mohamed Elgendi,
University of British Columbia, Canada

Reviewed by:
Maria Vittoria Bulgheroni,
Ab.Acus, Italy
Carlo Massaroni,
Campus Bio-Medico University, Italy

****Correspondence:***
Dhruv R. Seshadri
dhruv.seshadri@case.edu

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

Received: *03 May 2020*
Accepted: *11 June 2020*
Published: *23 June 2020*

Citation:
Seshadri DR, Davies EV, Harlow ER, Hsu JJ, Knighton SC, Walker TA, Voos JE and Drummond CK (2020) Wearable Sensors for COVID-19: A Call to Action to Harness Our Digital Infrastructure for Remote Patient Monitoring and Virtual Assessments. Front. Digit. Health 2:8. doi: 10.3389/fdgth.2020.00008

INTRODUCTION

Overview of COVID-19

The Coronavirus Disease 2019 (COVID-19), first recognized in December 2019 in Wuhan, China, is the latest respiratory disease pandemic currently plaguing global health. It has been shown to be caused by a novel coronavirus, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), that is structurally related to the virus that causes SARS. Li et al.

defined a suspected COVID-19 case as pneumonia that matched the following four criteria: (1) fever, with or without a recorded temperature; (2) radiographic evidence of pneumonia; (3) low or normal white-cell count or low lymphocyte count; and (4) no reduction in symptoms after antimicrobial treatment for 3 days (1). As its name suggests, the leading cause of fatality from COVID-19 is hypoxic respiratory failure (2–4). COVID-19 has posed significant challenges for the medical and civilian communities analogous to what was experienced in two preceding instances of the SARS-CoV virus outbreak in 2002 and 2003 and the Middle East Respiratory Syndrome (MERS) in 2012 (1, 5, 6). Importantly, Li et al. studied 425 patients with confirmed COVID-19 in Wuhan and estimated that the basic reproduction number (R_0) for SARS-CoV-2, at the time, to be 2.2 (1). This suggests that each infected person, on average, can spread the infection to an average of 2.2 other people. The virus will likely continue to spread unless this number falls below 1.0 (5). Moreover, timely and effective containment strategies have been a cornerstone of managing the COVID-19 outbreak and reducing viral transmission.

Return to Daily Living Post-COVID-19: From Testing to Digital Health Implementation

Most plans for recovery and the return to "normal," everyday life are centered on testing—namely determining those who currently have an infection and those who have developed antibodies against the virus, indicating a possible recovery. With any test, there may be false positive or false negative results (7). Of note, an antibody test, while useful in quantifying the number of cases that have occurred in a population, is typically not suitable for *early* disease detection and its association with immunity to the virus has been put into question (8). Additionally, there is considerable cross-reactivity between SARS-CoV-2 and four other coronaviruses, including those associated with the common cold (9). Polymerase chain reaction (PCR)-based tests are highly sensitive and specific in the laboratory setting; however, high costs and limited availability make these tests difficult to suit population health needs. In the face of a pandemic, time is of the essence and researchers must think of new ways to improve disease diagnosis and monitoring of disease progression.

With new tests in clinical trials, we believe there is an opportunity to leverage advances in remote patient monitoring technology to assist in early disease detection and monitoring by analyzing systemic infection precursors (**Figure 1**). Wearable sensor data may enable providers and patients to be alerted of a potential SARS-CoV-2 infection before symptoms become severe (**Figure 1A**). Importantly, a recent study showed that individuals with pre-existing hypertension, heart disease, or diabetes, which makes up nearly half of the United States population, had higher rates of intensive care hospitalization and death when diagnosed with COVID-19 (12). Additionally, data suggests that this vulnerable patient population also typically underreport their symptoms (13–15), making remote detection of disease through objective measures a possible way to improve timely escalation of care. On a larger scale, hospitals could use localized, de-identified data to track the spread and severity of the outbreak without violation of users' privacy to provide population-level care (**Figures 1B,C**). This becomes more relevant when one considers that the asymptomatic carrier rate is estimated to be between 25 and 50% of the entire United States population (16, 17). With such a large population potentially carrying the virus, digital health technologies that measure physiologic parameters can be leveraged to help identify population clusters to identify an emerging COVID-19 outbreak. Harnessing this information is feasible as ∼16% of the United States population (∼52.8 million people) currently have a smartwatch (18). Such technology may enable a more precise approach for subsequent more advanced testing (e.g., physiological testing), contact tracing, and quarantining. To further incentivize the adoption of such technologies, we envision companies that produce wearables will continue to work with insurance providers and other governing bodies to make these devices more accessible to the public (13). Most recently, in Germany, the Robert Koch Institute (equivalent of the United States Center for Disease Control and Prevention, CDC) supported the adoption of a smartphone app (Corona-Datenspende) which tracked temperature, pulse, and sleep from a minimum of 10,000 volunteers wearing smartwatches or fitness trackers with the aim of understanding how much of the population is clinically symptomatic from an influenza-like illness (ILI) (14). To date, more than 160,000 people have already enrolled (15). Results from the app will be displayed on an interactive online map, enabling both health authorities and the general public to better assess the prevalence and community distributions of infections (14). In the United States, a study published in early 2020 from the Scripps Research Institute demonstrated the ability to predict "hot spots" for influenza utilizing resting heart rate and sleep data from a smartwatch or fitness band (16). The team analyzed data from more than 47,000 consistent Fitbit users in five states (California, Texas, New York, Illinois, and Pennsylvania) over a 2-years period and found that when a cluster of individuals in one-region presented with increased heart rate, a subsequent rise in ILIs was detected. These models to map the prevalence of ILIs have correlated well with CDC data in the range of 0.84–0.9 (16). These studies highlight the clinical applications of wearable sensor technology and in the case of a pandemic, where "flattening the curve" is critical to limiting disease morbidity and mortality, such tools have the potential to improve health at the population level.

MEASUREMENT OF PHYSIOLOGICAL METRICS FROM WEARABLE SENSORS FOR COVID-19 MONITORING

COVID-19, along with other viral illnesses, is associated with several physiological changes that can be monitored using wearable sensors (**Table 1**). Many metrics derived from heart rhythm such as heart rate (HR), heart rate variability (HRV), resting heart rate (RHR), and respiration rate (RR) could serve as potential markers of COVID-19 infection and are already measured by wearable devices such as the Apple Watch, WHOOP

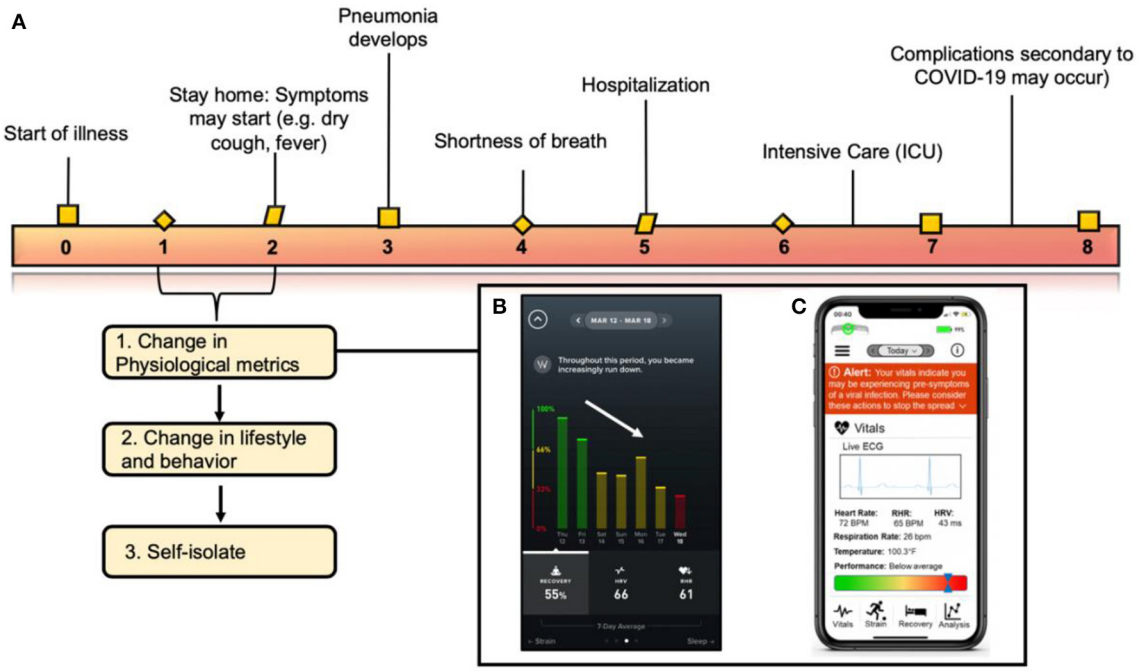


FIGURE 1 | Correlating clinical symptoms to the immune response of COVID-19 and when the implementation of wearable technology fits into the virus timeline. **(A)** Correlating symptoms to the immunological pathway of COVID-19 described in this figure. Wearable sensors can be used to detect changes in physiological metrics before a formal diagnosis. **(B)** Image from the WHOOP application correlating a decrease in recovery with a diagnosis by an individual diagnosed with COVID-19 on Wednesday the 18th (red bar). **(C)** Schematic detailing an example of an iPhone application collecting physiological data from a wearable sensor and translating those metrics to alert an individual on his/her overall health status. Figure reproduced and modified from Azuravesta Design (10), **(A)** and Team (11), **(B)**.

Strap, Fitbit, Zephyr BioHarness, or VivaLNK Vital Scout (**Table 2**). Additionally, changes in electrocardiogram (ECG) waveforms could contain information indicative of an infection. Many wearables report more complex metrics such as stress, recovery, activity, and sleep, which are typically calculated using a combination of cardiac and accelerometer-derived metrics. Due to the integration of multiple measurements, these metrics should exhibit an aggregate higher signal to noise ratio (SNR) than individual raw signals alone and, therefore, have higher predictive value. Core body temperature and arterial oxygen saturation (SpO_2) are also of clinical value due to the high prevalence of fever and respiratory symptoms in COVID-19; however, such measurements are not routinely measured by the commercial wearables today. Furthermore, as patient-centered quality metrics are considered, we hypothesize that wearable devices, once validated via rigorous longitudinal randomized controlled trials, can decrease invasive metrics derived from arterial blood gas procedures (intended to detect how well lungs move oxygen into the blood) or from obtaining cardiac troponins (indicative of myocardial injury) (17).

The upcoming sub-sections in this paper will focus on the current role wearable sensors in providing remote patient monitoring for COVID-19. Our goal in each of these sub-sections is to (1) summarize the clinical relevance of each physiological metric as it relates to COVID-19, (2) provide a brief technical overview of each parameter detection modality, and (3) provide a brief overview of patient implications as it relates to quality of care. Discussion of current clinical trials utilizing commercially available, off-the-shelf (COTS) wearable devices pertinent sensors to COVID-19 is included to highlight the current work in this domain (**Table 3**).

Cardiovascular Monitoring

There are several metrics related to cardiac function such as HR, HRV, and heart rhythm wherein changes in these metrics may be indicative of COVID-19 infection. Viral illness increases physiological stress on the body which typically manifests as an overall increase in HR. In many cases of viral infection, an elevated HR can be detected hours or days before the onset of symptoms (20). Elevation in HR is also a typical physiological response during fever as the body begins to mount a defense to infection (21). An increase in RHR can be indicative of systemic illness, and thus RHR data, on a population scale, has been proven to accurately model the outbreak of influenza (as previously described) (16). HRV, measured as the average time difference between heart beats, provides insight into overall health, performance, and stress of an individual. High HRV is associated with fitness and health (22). A significant decrease in HRV indicates inadequate recovery and is indicative of increased physiological stress (23). While there is a lack of clinical evidence on the predictive value of HRV for viral illness detection, there is a large amount of self-reported and anecdotal evidence which leads us to postulate that HRV trends can be used to predict the onset of illness (23). Researchers at Scripps recently launched

TABLE 1 | Sensor modalities for monitoring physiological metrics relevant to COVID-19.

	HR	Heart rhythm	HRV	RHR	RR	SpO_2	Skin temp	Core temp*	Sleep
ECG	x	x	x	x	x			x	
PPG	x	x	x	x	x	x		x	
Accelerometer					x				x
Temperature							x	x	

HR, Heart Rate; HRV, Heart rate variability; RHR, resting heart rate; RR, respiration rate, SpO_2, Blood Oxygen Saturation; ECG, Electrocardiogram; PPG, photoplethysmography; Temp, Temperature.
**Core Temperature is measured based on HR and skin temperature and cannot be measured as a stand-alone metric.*

TABLE 2 | Sampling of commercial wearable sensors used to monitor physiological parameters necessary for COVID-19 detection.

Company and device	Form factor	CT	ST	SpO_2	RR	HR	HRV	EDA	Other	Price ($)	FDA
AIO Sleeve 2.0	Arm sleeve			y		y	y		Act, ECG	$169	n
Apple Watch Series 4/5	Wrist monitor				y	y	y		Act, ECG	$399	y
Beddit	Contactless In-bed sensor				y	y			Slp	$150	n
Beurer SE80	Contactless In-bed sensor				y	y			Slp	$500	n
Biobeat	Wrist monitor, Chest Patch		y	y	y	y	y		BP, ECG	NA	y
BioIntellisense	Epidermal patch		y		y	y			Coughing, sneezing, freq	NA	y
Biostrap	Wrist monitor			y	y	y	y		Slp	$175-320	n
Biovotion Everion	Armband		y	y	y	y	y		Slp	NA	n
Cosinuss Two	In-ear	y		y		y	y		Act	$330	n
Empatica Embrace	Wrist monitor		y			y	y		Act, EDA	NA	y
Equivital LifeMonitor	Chest belt	y	y	y	y	y	y		GSR	NA	y
Fitbit Charge 4	Wrist monitor			y		y			Act, Slp	$150	n
Fitbit Ionic	Wrist monitor			y		y			Act, Slp	$250	n
Fitbit Versa 2	Wrist monitor			y		y			Act, Slp	$200	n
Garmin Fenix 5	Wrist monitor			y		y			Act, Slp	$500	n
Garmin Forerunner 945	Wrist monitor			y	y	y			Act, Slp	$550	n
Garmin Venu	Wrist monitor			y	y	y			Act, Slp	$300	n
Garmin Vivoactive 4	Wrist monitor			y	y	y			Act, Slp	$270	n
Hexoskin	Compression shirt			y	y	y	y		Act, Slp	$579	n
Kinsa	Smart thermometer		y							$50	n
Oura	Ring		y		y	y	y		Act, Slp	$299	n
Spire Health Tag	Tag attached to clothing				y	y			Act, Slp	$399	n
VivaLNK Fever Scout	Epidermal patch		y							$60	y
VivaLNK Vital Scout	Epidermal patch				y	y	y		Act	$150	y
WHOOP	Wrist monitor				y	y	y	y	Recovery, Slp	$30	n

Act, activity; BP, blood pressure; CT, core temperature; EDA, Early Detection Algorithm for viral illness or wellness prediction; ECG, electrocardiogram; EDA, electrodermal activity; Freq, Frequency; GSR, galvanic skin response; HR, heart rate; HRV, heart rate variability; NA, price not available online; RR, respiratory rate; Slp, sleep measures; SpO_2, oxygen saturation; ST, skin temperature. Table, Data used for table gathered from news reports, social media sites, and from Google Docs (19).

the Digital Engagement and Tracking for Early Control and Treatment (DETECT) study which seeks to correlate changes in HR to the incidence of acquiring a viral infection such as COVID-19 (24, 25). While other viral illnesses are being studied as well, the primary objective of this study is to assess HR, activity, and sleep data in 100,000 individuals to identify ILIs via the CareEvolution's myDataHelps application (24). The study, which commenced this past March, will utilize the Apple Watch, Garmin watch, and Fitbit, which are connected to Apple Health, Amazefit, or Google Fit platforms, respectively. Another study by the team at Scripps Research Institute, in collaboration with Stanford University and Fitbit, is assessing whether changes in HR, skin temperature, and SpO_2 can be used to predict the onset of COVID-19 before symptoms even start (26). These studies build upon the work published earlier this year by Scripps in correlating changes in HR to influenza (16).

TABLE 3 | Current clinical trials utilizing commercial wearable sensor devices to diagnose and monitor COVID-19.

Study name	Institution(s)/companies	Data source(s)	Focus of study	Clinical trials registry/ref
N/A	Central Queensland Univ; Cleveland Clinic	WHOOP Strap 3.0	Correlating changes in respiration rate to predicting COVID-19	N/A
COVIDENTIFY	Duke	AWs, Fitbits, Garmin	Predicting and assessing severity of contracting Covid-19 or influenza from wearable sensors and wellness surveys	N/A
DETECT study	Scripps Research Institute Stanford Univ	Fitbit, Apple Watch, Garmin, Amazefit, OURA, Beddit, etc	Determining whether changes in heart rate, activity, sleep, or other metrics might be an early indicator for COVID-19 or other viral infections	NCT04336020
COVID-19 detection study	Stanford Univ	Fitbit, Garmin, Apple Watch, and Oura	Enrolling subjects who are at higher risk of exposure.	N/A
TeamPredict	University California San Francisco	Oura Ring	Correlating changes in skin temperature and heart rate to COVID-19	N/A
Kinsa	N/A	Smart Thermometer	Correlating changes in skin temperature and social distancing guidelines to COVID-19	N/A

Table is put together based on press releases found on social media platforms and in the news. Aws, Apple Watch; Univ, University; Ref, References; N/A, Not applicable.

Electrocardiogram (ECG) and photoplethysmography (PPG) are widely used in wearable technology to monitor cardiac function (27–30). ECG is a measurement of the electrical activity in the heart, and PPG uses light (at specific nanometer wavelengths) to measure changes in blood volume (27, 31). While ECG sensors are typically implemented in the form of an epidermal patch that adheres to the stratum corneum (e.g., Zio Patch) and/or via leads to a benchtop instrument, the commoditization of wrist-worn monitors with predictive algorithms has enabled the measurement of heart rhythms from wearable devices such as the Apple Watch 4 and 5, although this measurement is not continuous (32, 33). On the other hand, PPG can be measured continuously in many locations on the body including the wrist, fingertips, earlobes, torso, and more (31). In this sense, PPG is more versatile and can be implemented in more form factors including watches and earbuds (27). While both are viable to monitor the metrics discussed above, ECG is a more direct measurement of heart activity which could potentially provide more insight toward the onset of COVID-19. There is growing evidence suggesting that COVID-19 is burdened by a higher risk of arrhythmic events (34). A study by Driggin et al. found that in 138 hospitalized COVID-19 patients, arrhythmias such as ventricular tachycardia/fibrillation represented the leading complication (19.6%) after acute respiratory distress syndrome, particularly in those admitted to intensive care unit where the prevalence rose to 44.4% (35). Future work toward moving this field forward, leveraging data analytics and wearable sensors, could involve detecting such arrhythmias in patients with COVID-19 in a real-time manner toward improving patient outcomes.

Cardiovascular Strain, Sleep, and Activity Levels

Many currently available wearable devices provide users with calculations of advanced metrics such as stress or strain, sleep, activity, and recovery. These metrics typically rely on a combination of measurements and are calculated daily. The combined measurements and long measurement time for these metrics should yield a higher SNR and thus will likely be better predictors of COVID-19 infection than any single raw metric. Cardiovascular stress or strain (the terms are used interchangeably between analytical platforms) is expressed as a dimensionless unit derived from a combination of HR and HRV data measured over a given day. For the purposes of this paper, we will use the word strain. Devices such as WHOOP measure cardiovascular strain based on time spent in HR zones. In the context of athletic performance, a field where cardiovascular strain has been extensively studied to modulate the internal workloads of athletes (27–30, 36), an individual undergoing a strength-based workout with minimal reps and periods of rest will have a lower strain if their HR is not elevated for extended periods of time (27, 36, 37). Increasing weight and adding new strength exercises will cause muscle soreness and muscle fatigue. This microtrauma from the eccentric lengthening of the muscle fibers will cause a decrease in HRV especially in the morning. Fatigued muscles will result in higher strain as the day progresses because the body is working harder to recover due to the disturbances in the individuals' homeostatic state. Along the same lines, cardiovascular strain is also expected to increase when fighting a viral infection. A viral infection such as influenza or COVID-19 does this by increasing the stress on the cardiovascular system, indicated by increases in RHR, HR, blood pressure, and an intrinsic stress hormone called catecholamines (38). Sleep is usually detected using a combination of HR patterns and accelerometer data. Sleep quality is assessed primarily through the analysis of HR, RHR, and HRV, but accelerometry may be used to determine disturbances. Elevated sleep duration has been shown to be predictive of ILI. An increase in sleep duration paired with a decrease in sleep quality would be expected to occur in COVID-19 cases. Activity metrics are intended to report the amount of physical exertion for a day or a given timeframe. Activity scores are typically based on periods of elevated HR and accelerometry. While changes in activity

data may not be particularly useful for individual treatment or diagnosis, general trends in activity data for a large population could likely be used for pandemic modeling or to study the health effects of social distancing and isolation. Martin et al. studied the relationship between exercise and respiratory track viral infections in small animal models and concluded that moderate intensity exercise reduced inflammation and improved the immune response to respiratory viral infections (39). The use of wearable sensors toward monitoring activity levels could provide an objective means of staying physically active and healthy during the COVID-19 pandemic. Recovery assessments are based on sleep, sleep quality, and HRV. There is a growing amount of evidence showing a clear downward trend in recovery scores in the days leading up to the onset of COVID-19 symptoms. These complex metrics may prove to be reliable indicators of COVID-19, but it is important to consider that each wearable device uses different algorithms and measurements for the calculation of these scores, which severely limits the population on which analysis of these metrics can be performed.

Respiration Monitoring

Respiration rate (RR) is of critical interest in COVID-19 cases due to the severe effects the virus can have on the lungs. COVID-19 presents as a lower-respiratory tract infection in most cases which can cause inflammation of lung tissue, coughing, and shortness of breath (40). The respiratory damage caused by COVID-19 reduces the overall efficiency of the lungs which results in an increased RR to compensate (40). Significantly increased RR is not as common in cases of other viral illnesses such as influenza or the common cold because these viruses typically affect the upper-respiratory tract (40). It may be concerning, however, that by time the patient is tachypneic, the disease may already be in an advanced stage. In a person who has a high likelihood of COVID-19 exposure, a device that is able to detect subtle changes in respiratory function prior to the onset of clinical symptoms, such as shallow respirations, wheezing, and shortness of breath, has the potential to be an effective tool. Of note, findings by Luo et al. indicated that as many as 70% of frontline health care workers are testing positive for COVID-19 (41). Frontline staff who care for patients with COVID-19 could benefit from the remote use of a wearable-sensor remote monitoring mechanism to objectively monitor for pre-clinical signs of infection, as a measure to prevent spread to other colleagues or patients. Additionally, current COVID-19 guidelines suggest that measuring resting RR can be used utilized as a criterion for intensive care unit (ICU) admission (42).

A recent review by Massaroni et al. assessed the suitability of different contact-based techniques for monitoring RR in clinical settings, occupational settings, and sports performance (43). Specifically, in the context of clinical settings, the authors noted that contact-based techniques such as strain, impedance, biopotential, and light intensity measurements offer a platform to detect RR in a non-invasive and unobtrusive manner. Toward the use of biopotential measurements for RR monitoring, RR can be derived from wearable devices that measure heart activity due to a phenomenon known as Respiratory Sinus Arrhythmia (RSA) (44). RSA results in increased HR during inspiration and decreased HR during expiration. Using this information, any wearable that can accurately measure heart rhythm can be used to derive respiration rate if an appropriate algorithm is implemented. Baseline resting of respiration rate can be determined when a subject is asleep and shows very little variation from night to night (40). Therefore, a significant increase in resting respiration rate indicates a high likelihood of decreased respiratory efficiency. WHOOP has focused on correlating changes in RR and recovery levels to predicting COVID-19 in their users (45). WHOOP, which recently partnered with Central Queensland University (CQU) and the Cleveland Clinic on such a study, will utilize the data collected from WHOOP's hardware from volunteers who have self-identified as having contracted COVID-19 to study changes in their respiration rate over time (45). The data, which is currently being collected for this study utilized the WHOOP 3.0 strap, was validated externally to determine the accuracy of respiration rate during sleep when compared against polysomnography (46). Based on the study, the team from WHOOP hypothesizes that measuring respiration rate during sleep could be valuable in detecting abnormal respiratory behavior in COVID-19 patients before symptoms are present (45). Recently, researchers from Duke University launched the "CovIdentify" study which utilizes devices such as the Fitbit and data from the Apple Health app to monitor an individual's sleep schedules, oxygen levels, activity levels, and HR over a 12 months period to determine if they are risk for COVID-19 (47, 48). Once the data is collected, the team will utilize their predictive algorithms to detect respiratory infections from the COVID-19 virus. Respiration rates are typically obtained in research and clinical-related settings which may not be indicative of individual's respirations at home; however, given that COVID-19 can complicate existing chronic respiratory disease, monitoring individuals in home settings can receive a more patient-centered approach to prescribing treatment (49). There remains an unmet medical need to ensure that algorithms that correlate changes in RR to COVID-19 are sensitive enough to filter out other lower respiratory infections such as pneumonia or influenza. Toward achieving this goal, the design of clinical trials to mitigate false positive diagnosis is critical toward the application of wearable sensor technology for COVID-19 monitoring.

SpO_2

The assessment of a patient with a respiratory illness typically includes measurement of the blood oxygen saturation (SpO_2), as hypoxia in certain clinical scenarios is indicative of a pneumonia. This is of particular importance in monitoring progression and severity of disease in COVID-19, where resting SpO_2 was found to be significantly lower in patients with a severe stage of the disease as determined by clinical symptoms and CT scan. SpO_2 measurements $< 90\%$ during hospital admission is seen in COVID-19 patients with higher systemic inflammatory markers and increased disease mortality (50, 51). While validated oximeters are abundant in the inpatient setting, few patients have this technology available in their homes. Smartphone-based pulse oximetry in the form of a

camera-based app and a probe-based app, the latter using an external plug-in probe, have been developed by several companies and evaluated in two published studies (52, 53). The plug-in probes showed modest accuracy in identifying hypoxia as measured by standard pulse-oximetry; however, the camera-based technology had limited ability to accurately detect hypoxia and is not considered standard of care for this purpose in the literature. PPG technology has been utilized in wearable pulse oximetry and a large number of fingertip-type oximeters are commercially available, but few meet acceptable accuracy standards (54). Examples include the MightySat™ Rx (Masimo) and Pulsox-310 (Konica Minolta). These available technologies were designed for oxygen management in chronic diseases such as chronic obstructive pulmonary disease (COPD) and sleep apnea; moreover, little research has looked at their utility in early detection and management of disease progression in acute respiratory illnesses. At present, this technology may be particularly useful, as more patients with mild symptoms of COVID-19 are being asked to stay at home and report changes in their respiratory symptoms via telemedicine modalities in an attempt to reduce the spread of the disease. In elderly patients or those with medical co-morbidities that are known to be at higher risk of disease progression, current wearable PPG technology may have a role in identifying those patients who are self-isolating at home that need a higher level of care due to hypoxemia which may or may not be accompanied by other clinical symptoms of early respiratory distress.

Temperature

Temperature measurement is extremely important to COVID-19 detection and has already been widely used by numerous countries as an immediate test to determine if travelers or citizens may be infected with COVID-19. While quarantining individuals with fever may prevent transmission to some degree, this approach to temperature monitoring is not sufficient because COVID-19 can be transmitted before a fever develops. Continuous monitoring of skin temperature is currently being implemented using wearable devices such as the TempTraq, Oura ring, VivaLNK Fever Scout, and QardioCORE. The TempTraq skin temperature sensor adheres to the body for 72 h and is currently being used to measure the temperature of frontline workers here at University Hospitals Cleveland Medical Center (55). A study performed by Stanford University using the MOVES, Scanadu Scout, Basis B1, Basis Peak, iHealth-finger, Masimo, RadTarge, and Withings found a notable increase in skin temperature as well as HR and RHR in the period preceding and during a viral infection (20). This change in skin temperature, particularly if paired with RHR and HRV information, could be used to predict COVID-19 infection before symptoms arise. Any such temperature sensing wearable device could also be used for fever tracking during illness and alert users and medical staff to a dangerous fever or sharp change in temperature. While skin temperature measurement is easy to implement, it has been shown to deviate up to 12°F from core body temperature (20). Additionally, temporal, oral, aural, and axillary temperature measurements have all been shown to be invalid estimations of core body temperature (compared against rectal thermometry) and are more prone to change due to environmental or behavioral factors (56). Core body temperature measurement provides a much more stable baseline for assessment and could prove to be more reliably indicative of illness than skin temperature and provide more insight into fevers for remote patient monitoring. Researchers at UT Southwestern Medical Center found that fluctuations in core body temperature regulate the body's circadian rhythm (57). In the study, the researchers focused on cultured mouse cells and tissues and found that genes related to circadian functions were influenced by changes in core temperature. Clinically, analytical platforms combining core body temperature measurements with those of respiration rate, HR, or HRV, could provide a more robust platform for predicting the incidence of COVID-19 in ways not done today.

Continuous skin temperature measurement is simple to implement in both hardware and software and can easily be implemented into a wearable device. Analog solutions such as thermocouples and thermistors could be used reliably, but digital temperature sensors are likely better for wearable applications due to their small size ($\sim 1 \times 1$ mm), low power requirements, and improved control. Such a sensor could be integrated into many existing wearable form factors though adhesive patches will likely prove more reliable due to constant contact with the skin. The gold standard for core body temperature is rectal thermometry. This measurement modality is not feasible for continuous measurement where non-invasive and unobtrusive monitoring is required. A large body of research has shown that core body temperature can be reliably predicted from skin temperature and HR through the use of Kalman filters or other machine learning (ML) algorithms (58, 59). While this technology needs further research and development before clinical or diagnostic use, the application of such algorithms could provide a non-invasive method to study the response of core body temperature to illness and provide advanced remote patient monitoring capabilities for fever treatment.

While there have not been any clinical trials correlating core body temperature to incidences of COVID-19, the smart thermometer maker, Kinsa, has shown from skin temperature data where people with the flu (and more recently COVID-19 infections) are located (60). The team studied a population in Miami-Dade County, Florida and found that a spike in fevers coincided with the well-known reports of Miami residents and tourists loosely following social distancing recommendations. As beaches closed and other isolation strategies were implemented in the county, the team found a significant drop in fevers. The team also noted that the trends observed in Miami hold true for other areas of the country that they studied: as individuals adhered to social distancing guidelines, within 5 days, a downward dip in fevers was observed (60). Another start-up, Oura, has partnered with the University of California, San Francisco on a new study to see if its device, Oura Ring, can detect physiological signs that may indicate the onset of COVID-19 (61). The study includes two parts wherein part one

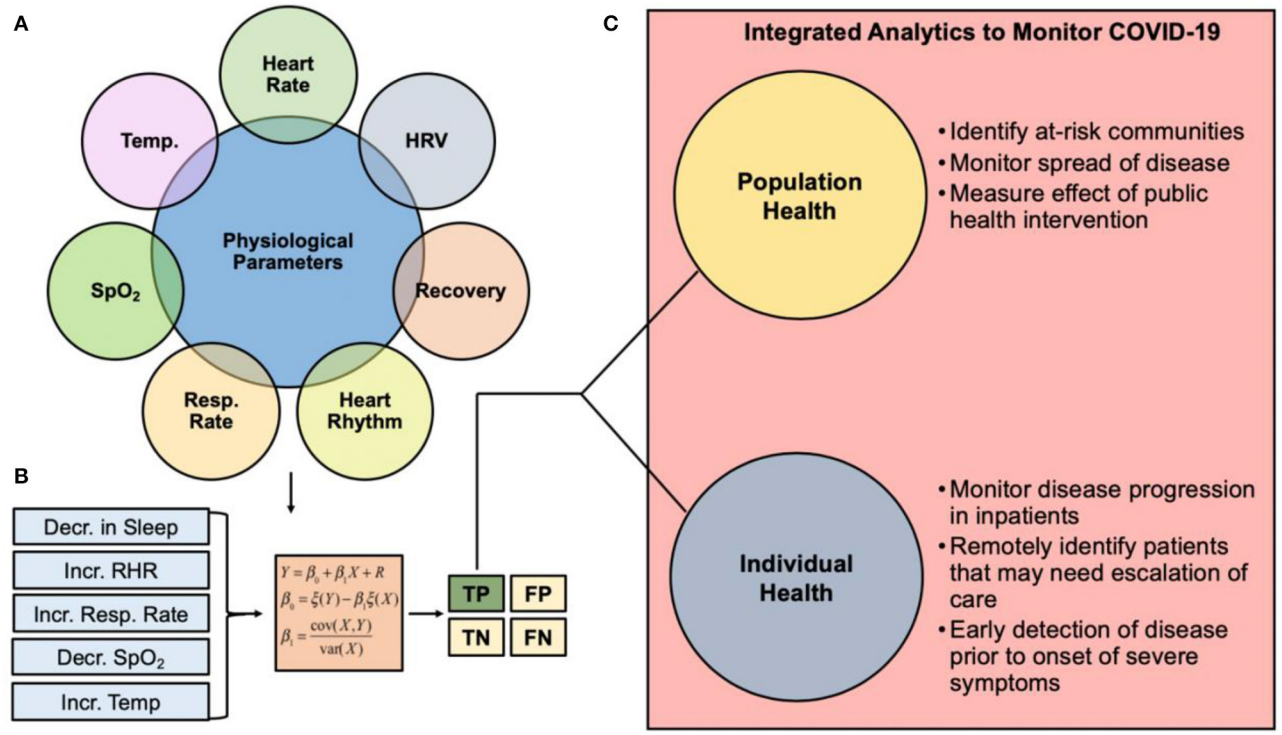


FIGURE 2 | Clinical pathway summarizing the role of wearable sensor technology and predictive analytics for monitoring COVID-19. **(A)** Physiological metrics currently capable of being measured from commercial wearable sensors. **(B)** Changes in physiological metrics can be inputted into an early detection algorithm for COVID-19 monitoring. The goal of such algorithms is to ensure the true positive rate is robust to support the use of the analytics for real-time clinical decision making. **(C)** Integrated analytics to monitor COVID-19 can be used to monitor individual or population health. HRV, heart rate variability; Resp Rate, respiration rate; SpO_2, blood oxygen saturation; Temp, temperature; TP, True Positive; FP, False Positive; TN, True Negative; FN, False Negative.

involves having 2,000 frontline healthcare professionals wear the Oura Ring to track skin temperature, sleep pattern, HR, and activity levels. Part two of the study will involve Oura's general user population wherein its 150,000 global users can opt in to participate and add to the overall pool of information with their ring's readings and daily symptom surveys (61). Recently an Oura user in Finland claimed that the ring alerted him that he was displaying symptoms of COVID-19 based on decreased recovery levels (from 80 to 90% to 54 coupled) and an increase in skin temperature of ∼1°C. These changes prompted the individual to get tested (61). The test results confirmed, that while asymptomatic, the individual had COVID-19 (61). The compilation of de-identified data sets from studies such as the two mentioned for temperature monitoring and those mentioned earlier on could lead to the development of an early detection algorithm.

EARLY DETECTION ALGORITHM TECHNOLOGY IS NEEDED FOR COVID-19 MONITORING

Many of the physiological changes measured by wearable devices discussed in the above sections can potentially be detected before a user experiences any significant clinical symptoms of illness (**Figure 2A**). We postulate that wearable devices can detect and alert users of possible infection with SARS-CoV-2 before they develop clinical symptoms through the development of an early detection algorithm (EDA) (**Figures 2A,B**). By notifying wearable device users of possible early infection, EDA could allow them the ability to self-isolate, seek care or diagnostic testing, and take other steps to mitigate transmission of the infection during a critical period of the disease process. Additionally, wearables could be used for remote patient monitoring in mild cases by allowing patients to report their vitals from home, saving critical hospital resources and reducing the risk of transmission to health care providers by avoiding in-person assessments (**Figure 2C**). A combination of the metrics listed above could result in a sufficiently high SNR to be used as a predictor of viral illness or COVID-19 risk. Developing an EDA with a high true positive and true negative rate is imperative for the translation of this technological platform for remote patient monitoring. Clinical staff such as intensive care nurses use early-warning system indicators to detect if individuals are at risk for further complications related to their care (62). Remote patient monitoring using wearable sensor technology provides an opportunity for developing more effective patient interventions, balancing nurse-patient care ratios, and decreasing costs associated with readmission rates and futile medical care.

FUTURE OUTLOOK AND RECOMMENDATIONS: ADOPTING WEARABLE SENSOR TECHNOLOGY

The development of integrated sensor technology has made it possible to remotely measure many physiologic parameters accurately, many of which are clinically useful in monitoring disease progression in a viral illness. The scope of influence of this technology is broad; it may be used to help to identify an individual under home-quarantine that needs a higher level-of-care or a community where an emerging outbreak may be imminent and requires an early intervention. We suspect that one of the largest impediments for the mass adoption of wearable sensors (and digital health technologies overall) in the United States toward remote patient monitoring is the issue of data privacy, data sharing, and underreporting. Wearable technology companies must ensure that only users who choose to participate will share their data (as done by WHOOP), and that the data will be anonymized and used for COVID-19 research only. Germany have provided us with a good example of how population health data can be handled, acknowledging their strong privacy concerns and stances on the limited collection of digital data (63). Underreporting of data by some populations may require their consent for safe data sharing and privacy agreements so that it can be used to inform better care, thus decreasing health disparities. The implementation of current trials demonstrates the convergence of wearable data, self-reported symptoms, molecular testing, and geospatial data toward developing platforms for managing COVID-19 and other outbreaks which may arise in the future. Building upon such trials, we see an opportunity to design a device that can accurately monitor many or all metrics of interest and through machine learning is able to develop an algorithm to reliably detect changes in population health status. A collaboration leveraging the expertise of clinicians, data scientists, engineers, and nurses is imperative to facilitate this advancement and may even be more acutely desired should there be a second wave of this pandemic.

AUTHOR CONTRIBUTIONS

DS, ED, EH, and SK wrote and edited the manuscript. TW, JH, JV, and CD contributed to the editing of the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors acknowledge collaboration between Case Western Reserve University, University Hospitals Cleveland Medical Center, and the University of California Los Angeles as it relates to this manuscript.

REFERENCES

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*. (2020) 382:1199–207. doi: 10.1056/NEJMoa2001316
2. PNAS. *Activation of the SARS Coronavirus Spike Protein via Sequential Proteolytic Cleavage at Two Distinct Sites*. PNAS. (2020). Available online at: https://www.pnas.org/content/106/14/5871 (accessed 19 April, 2020).
3. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. (2020) 367:1444–8. doi: 10.1126/science.abb2762
4. The New York Times. *What Does Coronavirus Do to the Body?* The New York Times. (2020). Available online at: https://www.nytimes.com/article/coronavirus-body-symptoms.html (accessed 19 April, 2020).
5. Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. *N Engl J Med*. (2020) 382:1268–9. doi: 10.1056/NEJMe2002387
6. Paules CI, Marston HD, Fauci AS. Coronavirus infections—more than just the common cold. *JAMA*. (2020) 323:707–8. doi: 10.1001/jama.2020.0757
7. The Washington Post. *A 'Negative' Coronavirus Test Result Doesn't Always Mean You Aren't Infected*. The Washington Post. (2020). Available online at: https://www.washingtonpost.com/science/2020/03/26/negative-coronavirus-test-result-doesnt-always-mean-you-arent-infected/ (accessed April 12, 2020).
8. Berkeley Lovelace WF Jr. *WHO Officials Say it's Unclear Whether Recovered Coronavirus Patients are Immune to Second Infection*. CNBC. (2020). Available online at: https://www.cnbc.com/2020/04/13/who-officials-say-its-unclear-whether-recovered-coronavirus-patients-are-immune-to-second-infection.html (accessed April 15, 2020).
9. Topol E. *Opinion: How Digital Data Collection Can Help Track Covid-19 Cases in Real Time*. Washington Post. (2020). Available online at: https://www.washingtonpost.com/opinions/2020/04/10/how-digital-data-collection-can-help-track-covid-19-cases-real-time/ (accessed April 12, 2020).
10. Azuravesta Design. *COVID-19 Pandemic*. Azuravesta Design. (2020). Available online at: https://www.azuravesta.com/covid-19-pandemic (accessed April 16, 2020).
11. Team WDS. *Heart Rate & Valuable WHOOP Data in COVID-19: Coronavirus*. WHOOP. (2020). Available online at: https://www.whoop.com/thelocker/whoop-data-coronavirus/ (accessed April 16, 2020).
12. Ferrari R, Di Pasquale G, Rapezzi C. Commentary: what is the relationship between Covid-19 and cardiovascular disease? *Int J Cardiol*. (2020) 310:167–8. doi: 10.1016/j.ijcard.2020.03.074
13. Miller C. *Aetna Launches New Program to Earn a Free Apple Watch*. (2019). Available online at: https://9to5mac.com/2019/05/02/aetna-apple-watch-attain-program/ (accessed April 12, 2020).
14. *Germany Launches Smartwatch App to Monitor Coronavirus Spread*. (2020). Available online at: https://finance.yahoo.com/news/germany-launches-smartwatch-app-monitor-123109671.html (accessed April 12, 2020).
15. The New York Times. *Germany Launches Smartwatch App to Monitor Coronavirus Spread*. The New York Times. (2020). Available online at: https://www.nytimes.com/reuters/2020/04/07/technology/07reuters-health-coronavirus-germany-tech.html (accessed April 12, 2020).
16. The Lancet Digital Health. *Harnessing Wearable Device Data to Improve State-Level Real-Time Surveillance of Influenza-Like Illness in the USA: A Population-Based Study*. (2020). The Lancet Digital Health. Available online at: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30222-5/fulltext (accessed April 12, 2020).
17. American College of Cardiology. *Troponin and BNP Use in COVID-19*. American College of Cardiology. (2020). Available online at: https://www.acc.org/latest-in-cardiology/articles/2020/03/18/15/25/troponin-and-bnp-use-in-covid19 (accessed April 20, 2020).
18. TechCrunch. Sixteen *Percent of US Adults Own a Smartwatch*. TechCrunch. (2020). Available online at: https://social.techcrunch.com/2019/02/12/sixteen-percent-of-u-s-adults-own-a-smartwatch/ (accessed April 12, 2020).
19. Google Docs. *Wearables for COVID-19 Remote Monitoring*. Google Docs. (2020). Available online at: https://docs.google.com/spreadsheets/

d/1rPvE3cBtMIfVhpSYqOpLy2WgEDaHqY1CEpyGSRVFUaI/edit?usp=sharing&usp=embed_facebook (accessed June 1, 2020).
20. Li X, Dunn J, Salins D, Zhou G, Zhou W, Rose SMS-F, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLOS Biol*. (2017) 15:e2001402. doi: 10.1371/journal.pbio.2001402
21. Karjalainen J, Viitasalo M. Fever and cardiac rhythm. *Arch Intern Med*. (1986) 146:1169–71. doi: 10.1001/archinte.146.6.1169
22. Holmes K. *5 Mental Ways to Improve Heart Rate Variability & Immunity*. WHOOP. (2020). Available online at: https://www.whoop.com/thelocker/mental-ways-to-increase-hrv-immunity/ (accessed April 13, 2020).
23. Deusen MV. *Heart Rate Variability: The Ultimate Guide to HRV*. WHOOP. (2019). Available online at: https://www.whoop.com/thelocker/heart-rate-variability-hrv/ (accessed April 12, 2020).
24. ClinicalTrials.gov. *The DETECT (Digital Engagement and Tracking for Early Control, and Treatment) Study–Full Text View*. ClinicalTrials.gov. (2020). Available online at: https://clinicaltrials.gov/ct2/show/NCT04336020 (accessed April 18, 2020).
25. *Detect*. (2020). Available onine at: https://detectstudy.org/ (accessed April 18, 2020).
26. Healthcare IT News. *Scripps, Stanford Working With Fibit to Assess Wearables' COVID-19 Tracking Abilities*. Healthcare IT News. (2020). Available online at: https://www.healthcareitnews.com/news/scripps-stanford-working-fibit-assess-wearables-covid-19-tracking-abilities (accessed April 18, 2020).
27. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfes CM, Zorman CA, et al. Wearable sensors for monitoring the physiological and biochemical profile of the athlete. *Npj Digit Med*. (2019) 2:72. doi: 10.1038/s41746-019-0150-9
28. Seshadri DR, Drummond C, Craker J, Rowbottom JR, Voos JE. Wearable devices for sports: new integrated technologies allow coaches, physicians, and trainers to better understand the physical demands of athletes in real time. *IEEE Pulse*. (2017) 8:38–43. doi: 10.1109/MPUL.2016.2627240
29. Seshadri DR, Rowbottom JR, Drummond C, Voos JE, Craker J. A review of wearable technology: moving beyond the hype: from need through sensor implementation. In: *2016 8th Cairo International Biomedical Engineering Conference (CIBEC)*. (2016). p. 52–5. doi: 10.1109/CIBEC.2016.7836118
30. Seshadri DR, Magliato S, Voos JE, Drummond C. Clinical translation of biomedical sensors for sports medicine. *J Med Eng Technol*. (2019) 43:66–81. doi: 10.1080/03091902.2019.1612474
31. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int J Biosens Bioelectron*. (2018) 4:195–202. doi: 10.15406/ijbsbe.2018.04.00125
32. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai MY, et al. Accuracy of apple watch for detection of atrial fibrillation. *Circulation*. (2020) 141:702–3. doi: 10.1161/CIRCULATIONAHA.119.044126
33. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai M, et al. Accuracy of the apple watch 4 to measure heart rate in patients with atrial fibrillation. *IEEE J Transl Eng Health Med*. (2020) 8:2700204. doi: 10.1109/JTEHM.2019.2950397
34. Lazzerini PE, Boutjdir M, Capecchi PL. COVID-19, arrhythmic risk and inflammation: mind the gap! *Circulation*. (2020). doi: 10.1161/CIRCULATIONAHA.120.047293. [Epub ahead of print].
35. Driggin E, Madhavan MV, Bikdeli B, Chuich T, Laracy J, Bondi-Zoccai G, et al. Cardiovascular considerations for patients, health care workers, and health systems during the coronavirus disease 2019 (COVID-19) pandemic. *J Am Coll Cardiol*. (2020) 75:2352–71. doi: 10.1016/j.jacc.2020.03.031
36. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfes CM, Zorman CA, et al. Wearable sensors for monitoring the internal and external workload of the athlete. *Npj Digit Med*. (2019) 2:1–18. doi: 10.1038/s41746-019-0149-2
37. Lynch A. *Ask Us Anything: WHOOP Strain*. WHOOP. (2019). Available online at: https://www.whoop.com/thelocker/ask-us-anything-strain/ (accessed June 1, 2020).
38. Torrance Memorial Medical Center. *Nov 14 P on, 2017. How Can the Flu Affect Your Heart?* Torrance Memorial Medical Center. (2020). Available online at: https://www.torrancememorial.org/News_Center/2017/November/How_Can_the_Flu_Affect_Your_Heart_.aspx (accessed June 1, 2020).
39. Martin SA, Pence BD, Woods JA. Exercise and respiratory tract viral infections. *Exerc Sport Sci Rev*. (2009) 37:157–64. doi: 10.1097/JES.0b013e3181b7b57b
40. Capodilupo E. *Tracking Respiratory Rate and the Coronavirus*. WHOOP. (2020). Available online at: https://www.whoop.com/the-locker/respiratory-rate-tracking-coronavirus/ (accessed April 4, 2020).
41. Luo M, Cao S, Wei L, Tang R, Hong S, Liu R, et al. Precautions for intubating patients with COVID-19. *Anesthesiol J Am Soc Anesthesiol*. (2020). 132:1616–8. doi: 10.1097/ALN.0000000000003288
42. Frontiers. *Remote Respiratory Monitoring in the Time of COVID-19: Physiology*. (2020). Available online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00635/full (accessed May 31, 2020).
43. Massaroni C, Nicolò A, Lo Presti D, Sacchetti M, Silvestri S, Schena E. Contact-based methods for measuring respiratory rate. *Sensors*. (2019) 19:908. doi: 10.3390/s19040908
44. Hayano J, Yasuma F, Okada A, Mukai S, Fujinami T. Respiratory sinus arrhythmia. *Circulation*. (1996) 94:842–7. doi: 10.1161/01.CIR.94.4.842
45. TechCrunch. *Researchers to Study if Startup's Wrist-Worn Wearable Can Detect Early COVID-19 Respiratory Issues*. TechCrunch. Available online at: https://social.techcrunch.com/2020/04/01/researchers-to-study-if-startups-wrist-worn-wearable-can-detect-early-covid-19-respiratory-issues/ (accessed April 18, 2020).
46. Berryhill S, Morton CJ, Dean A, Berryhill A, Provencio-Dean N, Patel SI, et al. Effect of wearables on sleep in healthy individuals: a randomized cross-over trial and validation study. *J Clin Sleep Med*. (2020) 775–83. doi: 10.5664/jcsm.8356
47. CovIdentify. *A Duke University Study*. (2020). Available online at: https://covidentify.org/ (accessed April 18, 2020).
48. Duke Pratt School of Engineering. *'CovIdentify' Pits Smartphones and Wearable Tech Against the Coronavirus*. Duke Pratt School of Engineering. (2020). Available online at: https://pratt.duke.edu/about/news/covidentify-pits-smartphones-and-wearable-tech-against-coronavirus (accessed April 18, 2020).
49. Chu M, Nguyen T, Pandey V, Zhou Y, Pham HN, Bar-Yoseph R, et al. Respiration rate and volume measurements using wearable strain sensors. *Npj Digit Med*. (2019) 2:1–9. doi: 10.1038/s41746-019-0083-3
50. Dai H, Zhang X, Xia J, Zhang T, Shang Y, Huang R, et al. High-resolution chest CT features and clinical characteristics of patients infected with COVID-19 in Jiangsu, China. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. (2020) 95:106–12. doi: 10.1016/j.ijid.2020.04.003
51. Wang Z, Yang B, Li Q, Wen L, Zhang R. Clinical features of 69 cases with coronavirus disease 2019 in Wuhan, China. *Clin Infect Dis*. (2020). doi: 10.1093/cid/ciaa272 . [Epub ahead of print].
52. Tomlinson S, Behrmann S, Cranford J, Louie M, Hashikawa A. Accuracy of smartphone-based pulse oximetry compared with hospital-grade pulse oximetry in healthy children. *Telemed J E-Health Off J Am Telemed Assoc*. (2018) 24:527–35. doi: 10.1089/tmj.2017.0166
53. Tayfur I, Afacan MA. Reliability of smartphone measurements of vital parameters: a prospective study using a reference method. *Am J Emerg Med*. (2019) 37:1527–30. doi: 10.1016/j.ajem.2019.03.021
54. Tamura T. Current progress of photoplethysmography and SPO2 for health monitoring. *Biomed Eng Lett*. (2019) 9:21–36. doi: 10.1007/s13534-019-00097-w
55. News. *Best Baby Digital Thermometer: TempTraq*. TempTraq. (2020). Available online at: https://www.temptraq.com/News/University-Hospitals-expands-use-of-TempTraq%C2%AE-syst (accessed May 3, 2020).
56. Ganio MS, Brown CM, Casa DJ, Becker SM, Yeargin SW, McDermott BP, et al. Validity and reliability of devices that assess body temperature during indoor exercise in the heat. *J Athl Train*. (2009) 44:124–35. doi: 10.4085/1062-6050-44.2.124
57. Buhr ED, Yoo ED, Takahashi JS. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science*. (2010). 330:379–385. doi: 10.1126/science.1195262

58. USARIEM. *Core Body Temperature Estimation From Heart Rate*. (2020). Available online at: https://www.usariem.army.mil/index.cfm/modeling/cbt_algorithm (accessed April 12, 2020).
59. Eggenberger P, MacRae BA, Kemp S, Bürgisser M, Rossi RM, Annaheim S. Prediction of core body temperature based on skin temperature, heat flux, and heart rate under different exercise and clothing conditions in the heat in young adult males. *Front Physiol*. (2018) 9:1780. doi: 10.3389/fphys.2018.01780
60. TechCrunch. *Kinsa's Fever Map Could Show Just How Crucial it is to Stay Home to Stop COVID-19 Spread*. TechCrunch. Available online at: https://social.techcrunch.com/2020/03/23/kinsas-fever-map-could-show-just-how-crucial-it-is-to-stay-home-to-stop-covid-19-spread/ (accessed April 12, 2020).
61. TechCrunch. *Oura Partners With UCSF to Determine if its Smart Ring Can Help Detect COVID-19 Early*. TechCrunch. (2020). Available online at: https://social.techcrunch.com/2020/03/23/oura-partners-with-ucsf-to-determine-if-its-smart-ring-can-hep-detect-covid-19-early/ (accessed April 18, 2020).
62. Kavanaugh MJ, So JD, Park PJ, Davis KL. Validation of the intensive care unit early warning dashboard: quality improvement utilizing a retrospective case-control evaluation. *Telemed J E-Health Off J Am Telemed Assoc*. (2017) 23:88–95. doi: 10.1089/tmj.2016.0073
63. HodgeMon N. *Feb 3, Pm 2020 1:22. Germany's Dual Approach to Data Regulation Under the GDPR*. Compliance Week. (2020). Available online at: https://www.complianceweek.com/data-privacy/germanys-dual-approach-to-data-regulation-under-the-gdpr/28386.article (accessed April 12, 2020).

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

BRIEF RESEARCH REPORT
published: 21 December 2020
doi: 10.3389/fdgth.2020.619692

Advancing PPG Signal Quality and Know-How Through Knowledge Translation—From Experts to Student and Researcher

Samuel Huthart[1], *Mohamed Elgendi*[2], *Dingchang Zheng*[3], *Gerard Stansby*[1,4] *and John Allen*[1,3,5]*

[1] Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [2] Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada, [3] Research Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom, [4] Northern Vascular Centre, Freeman Hospital, Newcastle upon Tyne, Unite Kingdom, [5] Northern Medical Physics and Clinical Engineering Department, Freeman Hospital, Newcastle upon Tyne, United Kingdom

OPEN ACCESS

Edited by:
Kun Qian,
The University of Tokyo, Japan

Reviewed by:
Rik Vullings,
Eindhoven University of Technology, Netherlands
Raúl Alcaraz,
University of Castilla-La Mancha, Spain

***Correspondence:**
John Allen
ad5325@coventry.ac.uk

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

Received: 20 October 2020
Accepted: 24 November 2020
Published: 21 December 2020

Citation:
Huthart S, Elgendi M, Zheng D, Stansby G and Allen J (2020) Advancing PPG Signal Quality and Know-How Through Knowledge Translation—From Experts to Student and Researcher. Front. Digit. Health 2:619692. doi: 10.3389/fdgth.2020.619692

Objective: Despite the vast number of photoplethysmography (PPG) research publications and growing demands for such sensing in Digital and Wearable Health platforms, there appears little published on signal quality expectations for morphological pulse analysis. Aim: to determine a consensus regarding the minimum number of undistorted i.e., diagnostic quality pulses required, as well as a threshold proportion of noisy beats for recording rejection.

Approach: Questionnaire distributed to international fellow researchers in skin contact PPG measurements on signal quality expectations and associated factors concerning recording length, expected artifact-free pulses ("diagnostic quality") in a trace, proportion of trace having artifact to justify excluding/repeating measurements, minimum beats required, and number of respiratory cycles.

Main Results: 18 (of 26) PPG researchers responded. Modal range estimates considered a 2-min recording time as target for morphological analysis. Respondents expected a recording to have 86–95% of diagnostic quality pulses, at least 11–20 sequential pulses of diagnostic quality and advocated a 26–50% noise threshold for recording rejection. There were broader responses found for the required number of undistorted beats (although a modal range of 51–60 beats for both finger and toe sites was indicated).

Significance: For morphological PPG pulse wave analysis recording acceptability was indicated if <50% of beats have artifact and preferably that a minimum of 50 non-distorted PPG pulses are present (with at least 11–20 sequential) to be of diagnostic quality. Estimates from this knowledge transfer exercise should help inform students and researchers as a guide in standards development for PPG study design.

Keywords: digital health, photoplethysmography, pulse, peripheral arterial disease, pulse wave analysis, signal quality, wearable sensor, artefact rejection

INTRODUCTION

Photoplethysmography (PPG) is a vascular optical measurement technique, used to detect blood volume changes in the microvascular bed of target tissue (1). Many studies have been conducted investigating various body sites as a single measurement (e.g., single PPG sensor located on a single body site) and multi-site measurements (multiple PPG sensors located across a range of body sites). The finger and toe pad sites are usually assessed. A range of features of the pulse wave have been studied, including pulse transit time, pulse interval, peak-to-peak interval, amplitude, pulse contour, as well as their natural variability (1). Subsequently, PPG has been utilized in an array of settings, from bedside physiological measurement e.g., heart rate, oxygen saturation, to hypertension assessment, and detailed peripheral vascular assessment (1–4). PPG has also become a key sensing technology in Digital and Wearable Health devices.

It is well-established that there is variability in the PPG waveform over time and that there can be differences in morphology and dynamics between different peripheral body sites (1), for example respiration as well as blood pressure changes can modulate PPG signals over periods of seconds/10's of seconds. Furthermore, artifact from sensor and limb movement and/or tremor can limit the reliable extraction of pulse features and so having a recording of sufficient length helps identify, and thus reject such episodes of noise (**Figure 1**) (5). These signal variability considerations form the rationale behind taking an average of multiple beats for representative morphological analysis of the PPG "AC" pulsatile component and the motivation for this study. However, despite the broad range of studies concerning applications of PPG especially in Digital Health (1, 3), as well as the significantly smaller number of investigations focussed specifically on quantifying signal quality (6–12), there appears little yet published on signal quality expectations, e.g., minimum length of recording or proportion of noisy beats needed to reject a recording, which consequently affects reproducibility and ultimate measurement value. We therefore carried out a consensus exercise by composing a questionnaire. We aimed to determine for morphological pulse analysis if there was agreement regarding minimum recording length, the minimum number of undistorted i.e., diagnostic quality pulses required, as well as the threshold proportion of noisy beats needed for the recording. Our wider goal is to transfer knowledge from experienced PPG workers to other (future) researchers and students internationally.

METHODS

We developed a set of questions based on a specific clinical PPG measurement scenario in our questionnaire: finger and toe pad PPG measurements carried out on a healthy adult subject, acclimatized for at least 10 min within a warm room. The individual would be relaxed, lying supine with their arms by their side, having been instructed to lie still and breathe gently throughout. We identified a group of selected fellow PPG workers established in the field and all of those contacted were known to have published/presented their research, encompassing various fields in the cardiovascular application of PPG. The questionnaire was distributed internationally to known fellow researchers working in skin contact PPG measurements.

The 6 questions on the questionnaire applicable to morphological analysis of the "AC" pulsatile component of PPG, i.e., pulse shape (13, 14), are summarized in the **Appendix**.

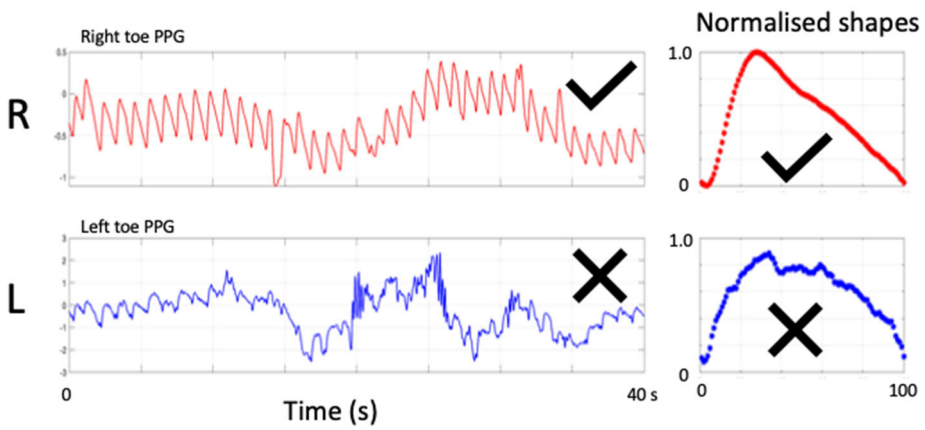


FIGURE 1 | Example of bilateral toe PPG recording from a healthy subject. "AC" pulsatile components can be seen for the toe recording, each superimposed on a slowly varying "DC" baseline. Bilateral similarity would be expected for the pulse features at the toes, including for normalized shape averaged over all beats present. The contralateral right (R) toe trace is of reasonable quality and has a normal shape characteristic. However, there is unacceptable levels of artifact present in the left (L) toe trace and appears to result in an abnormal damped normalized pulse shape. Attempts at further signal filtering and/or cluster analysis would be unlikely to salvage the left side toe trace for classification purposes.

Each question had specific numerical/range selections to choose from, there were no open responses to these questions. The questions related to the minimum length of recording time, the expected proportion of good quality beats per recording, the extent of noise needed to reject a recording, the minimum number of total and also sequential non-distorted i.e., diagnostic quality pulses required (criteria for an analysable beat outlined by Orphanidou (9): pulse requiring a clear peak, trough, anacrotic, and catacrotic phases), as well as the minimum number of respiratory cycles per recording needed, noting the low-frequency alterations in the PPG signal associated with respiration (15). The results from a subset of participants specializing in PPG and peripheral arterial occlusive disease (PAOD) diagnostics were also noted. Two authors independently checked the results from the survey.

RESULTS

In total, we approached 26 individuals *via* e-mail worldwide. Eighteen agreed to take part and give their views in the poll, coming from organizations in countries covering at least 3 continents (i.e., including the US, Europe and Asia). One questionnaire was only partially completed, and thus removed from the study—giving 17 responses in total to summarize.

With the exception of question 1, participants could select an answer relating to measurements at finger and toe sites independently. However, due to differences in experience, some opted to answer for only one measurement site. Hence, there was a discrepancy in the number of responses for the two body site locations (noting there were twelve full responses about the toe measurement site). In the few circumstances where participants had highlighted more than one answer per question for a single site, we decided to use their lowest answer, as this indicated the minimum threshold that they deemed acceptable.

The responses did tend to vary across the survey, but we were able to obtain a modal range value for each of the related questions (**Table 1**). For morphological analysis, for both the finger and toe sites, a minimum recording time of 2 min was a recommended target and participants expected 86–95% of a PPG recording to be of diagnostic quality i.e., beats with no distortion. Participants advocated a threshold for the proportion of beats with artifact required to reject a recording in the range 26–50%, and a minimum number of sequential undistorted beats of 11–20. Less clear patterns from respondents, due to spread, were for the number of respiratory cycles (modal choice was marginally for >10 cycles but with responses similar for 2 and 5 cycles selections) and the minimum number of total undistorted beats (for both finger and toe sites the modal selection was marginally for 51–60 beats). **Figure 2** shows key consensus results for the expected proportion of good quality pulses and for the proportion of beats with artifact to reject a recording. As well as such bar charts, a series of scatterplots were also produced in an attempt to identify relationships between answers to different questions, however no obvious correlations were observed and have therefore not been included.

The results from a subset of participants specializing in toe PPG and PAOD diagnostic assessments showed, however, only 4 of the 17 responding authors had published data for this application area. Nevertheless, when the answers pertaining to these specific parameters were compared to the rest of the respondents, there appeared general agreement.

TABLE 1 | Summary of modal selections for morphological PPG pulse wave analysis.

	Finger site	Toe site
Modal selection for minimum PPG recording time (minutes)	2	2
Modal selection of range for expected proportion of undistorted i.e., diagnostic quality beats (%)	86–95	86–95
Modal selection of range for proportion of beats with artifact to reject a recording (%)	26–50	26–50
Modal selection of range for minimum number of undistorted i.e., diagnostic quality beats in the recording	51–60	51–60
Modal selection of range for minimum number of successive undistorted i.e., diagnostic quality beats in the recording	11-20	11-20
Modal selection for minimum number of respiratory cycles	>10	>10

The modal selections on the questionnaire were similar overall for finger and toe measurement sites.

DISCUSSION

Responses varied across the survey, although for most questions/sites a modal value could be ascertained. This work pointed us to a potential memorable 50/50 rule of thumb for guidelines in PPG i.e., finger and toe PPG pulse traces of a few minutes in length can be acceptable for morphological pulse wave analysis provided there was <50% of the beats with artifact and that a minimum of 50 undistorted i.e., diagnostic quality beats were present, and that there is potentially an additional need for at least 10–20 successive good quality beats in the recording. Given the vastness of PPG application, there is unfortunately no exact way to formulate a "one size fits all" standard but our results should have value in helping in knowledge transfer to guide researchers and medical device developers in PPG measurement and analysis.

The broadness in survey responses we believe was in part related to a lack of agreed guidelines in PPG measurement protocol and signal quality expectations. Coupled with this, in relation to wider measurement protocol considerations, there is an absence of standardized equipment (e.g., key probe/clip attachment designs) and measurement set-up, other than the general consensus that participants should be allowed a period of time to rest (but no standardized time known), and to acclimatize to the ambient measurement conditions (but no standardized temperature and/or humidity defined) (1). Future follow-up surveys should consider seeking expert consensus in each of these areas for the benefit of wider measurement communities. There is also a large number of clinical applications covered by those surveyed and future studies could certainly look at the applications aspect in more detail with specific analysis and visualization of results in mind. Additionally, another possible

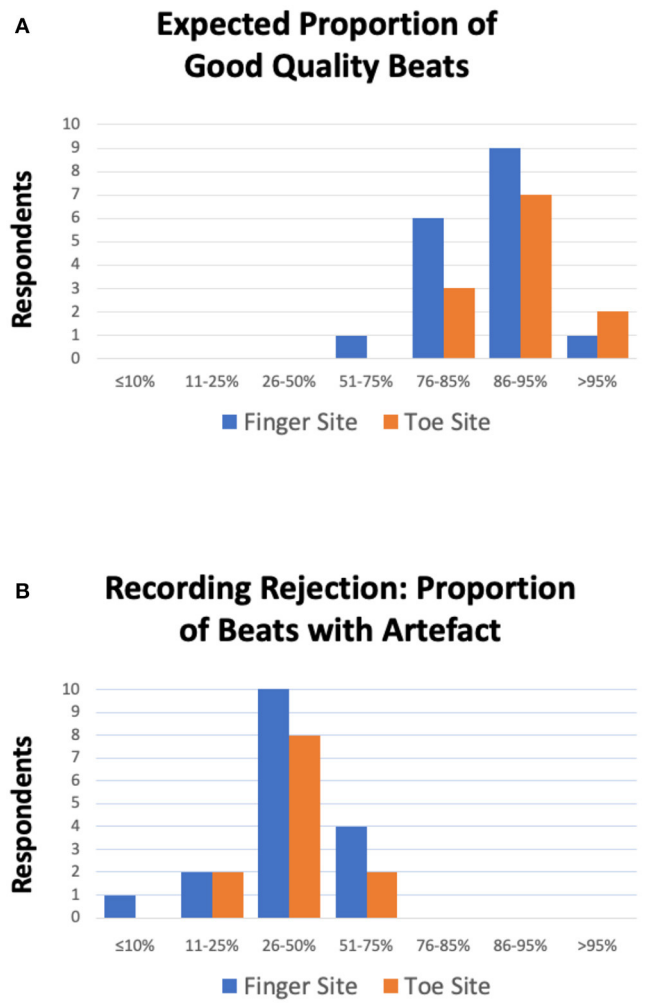


FIGURE 2 | (A) Expected proportion of good quality beats i.e., artifact free beats per PPG recording. For both finger (blue) and toe (orange) measurement sites, all participants expected the majority of the recording to be undistorted i.e., of diagnostic quality (>51%). Furthermore, the majority of respondents expected ≥86–95% of beats to be of good quality ($n = 9$ and $n = 7$ respondents for the finger and toe sites, respectively). **(B)** Proportion of beats with artifact required to reject a PPG recording for the finger (blue) and toe (orange) sites. For both sites, the majority of respondents advocated that ≥26–50% of beats having noise was sufficient to reject a recording ($n = 10$ and $n = 8$ respondents for the finger and toe sites, respectively).

reason for the lack of a general recommendation, is a current focus on noise-reducing and waveform identifying algorithms (16), an important area of study especially concerning the commercial use of PPG in wearable sensor systems for modern Digital Healthcare. If the developed algorithms can robustly eliminate noise and extract analysable pulse waves unfailingly, one might argue there would be no need for a standard, as the remaining signal should be of good quality. However, developing noise-reducing algorithms in PPG has proved to be difficult, for example with clustering methods ineffective with PPG signals dominated by noise (17). Furthermore, excessive filtering often leads to distortion of the PPG pulse shape, masking useful physiological information, whilst too little filtering may allow the quasi-static "DC" component to dominate over the "AC" pulsatile component (1). In PPG this "DC" component can include the low / very low frequency changes of the signal and not just be a value at 0 Hz (1). The high pass filter being particularly important in PPG and needs careful selection for PPG shape assessment, with 0.15–0.2 Hz proposed by Allen and Murray (18) showing no clear shape distortion for multi-site PPG measurements at ear, finger and toe sites in a group of healthy subjects (18). This work from 2004 also showed that the ratios for PPG "DC" to "AC" components were similar at finger and toe sites for this range of high pass filter settings. This "DC":"AC" measure of PPG variability also links to survey findings that responses were similar for finger and toe sites overall. To reiterate, the measurement methods and protocol used are very important and should always be clearly defined. For example, measurement device parameters, including the filtering, operating wavelength, and mode used, should be declared in research publications concerning contact PPG.

It is also important to comment on our bias assessment of the survey design and subsequent analysis. A survey may include a form of error such as sampling variability, interviewer effects, frame errors, response bias, and non-response bias (19–22). We designed the survey questions on PPG morphology and signal quality to contain only close-ended questions, which are answered by a simple selection from 7 (8 for Q1) choices. The main reason being to create data that are easily quantifiable, and straightforward to code. This also allowed us to categorize respondents into groups based on the options they have selected, thus increasing interpretability of the data. We also minimized all biases in our research with the following areas considered. Specification error—this did not occur as there was clear communication between the experts and data analysts. Frame error—this did not occur as our target population was those experienced in PPG research. Non-response error—in our survey all experts responded to questions regarding the finger site, however 12 (of 17) responded to the questions concerning the toe sites, potentially leading to a non-response error in this scenario. We suggest that it was the case that the 5 non-respondents, pertaining to toe measurements, focused their answers on their main measurement site of interest rather than guessing for a site they had little experience of. We also note that we did not add any value on behalf of any of the experts. Measurement error—our survey was carefully and manually carried out to reduce the risk of such an error. Finally, processing error—the lead (SH) and corresponding (JA) co-authors independently checked all results presented in this paper from the original set of questionnaire responses.

To our knowledge, this is the first report of its kind drawing on the knowledge and experience of published PPG workers. We obtained survey responses from 17 (12 of these responded on toes) well-respected experts in the field, noting that in Delphi exercises, a minimum of 12 respondents is generally considered to be sufficient to enable consensus to be achieved (23, 24). A possible limitation though may be the varying experience among the researchers contacted, albeit working predominantly

in morphological analysis, but with lesser involvement at the toe site compared to the finger site. All of those contacted to take part were professionally known by the senior author (JA) and were known to have published/presented their research, encompassing various fields in the cardiovascular application of PPG across the globe. We accept that such PPG researcher choice would involve bias and is potentially a limitation of this study, and future wider surveys could be designed to assess differences in opinion between researchers from different geographical regions. However, this we believe is something for the future in a carefully designed survey setting out to investigate this specific aspect. Another limitation is the proposed ranges for some questions being wide, and thus being heavily open to interpretation, specifically the minimum length of time per recording.

Our results have shown value and indicated key recommendations for contact PPG recordings in relation to noise and signal quality expectations. It is also important to note signal quality and noise rejection are very important topics in clinical physiological measurement not just for PPG but for a range of signal types—this should be a very important topic for design of noise reduction algorithms in Digital Healthcare platforms including wearable sensing applications.

A potential future study would be to distribute another questionnaire investigating a selection of these suggested ranges further with narrower limits, and a more detailed description of the purpose of the measurement i.e., it may be that a large proportion of participants recommend that a recording warrants rejection if 41–50% noise is present, compared to 31–40%. Also, morphological analysis covers a wide range of sub-studies in PPG, and therefore many researchers would put forward answers to the above questions differing entirely, depending on the pulse feature in question. The specific application to Digital Health and Wearable sensing with PPG should be an added focus. Furthermore, even though individuals were given the opportunity to provide additional details at the conclusion of the survey, written justification for their decisions would have been insightful. In future follow-on surveys, a wider range of body sites should be considered (including the ear and forehead sites for example), as well user experiences and expectations with various forms of transmission and reflection mode sensor/tissue attachment technology as well as remote i.e., imaging PPG. Lastly, we believe it would next be helpful to issue the survey to an open biosensor-based research community, for future reference.

SUMMARY

In summary, our results can be used as a guide for future studies in PPG and especially morphological pulse wave analysis, specifically in determining wider views in justifying which signals to utilize, discard and repeat. This area is very important in Digital Healthcare systems with wearable sensing and the need to gather repeatable and meaningful PPG data. Our study also provides initial recommendations available for other workers in the field of PPG—facilitating knowledge transfer to students and researchers to support the move toward improved standardization in measurement protocol, morphological pulse wave analysis, as well as address the real-world problem of artifact reduction in PPG.

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 Newcastle University Ethics Committee (Online Application: Ref 3549/2020). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SH, GS, and JA contributed to conception and design of the study. SH performed the statistical analysis and wrote the first draft of the manuscript. JA wrote sections of the manuscript and confirmed the reported analysis results. ME shaped the article's focus with formulating the knowledge translation aspect as well as supporting the bias assessments. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

In agreement with our questionnaire terms and conditions, all responses have remained anonymous. We would however like to thank all of those who contributed their opinions, to add benefit in what we believe is an important effort to try and determine a standard for PPG research looking at both measurement protocol and PPG pulse wave analysis.

REFERENCES

1. Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas.* (2007) 28:R1–39. doi: 10.1088/0967-3334/28/3/R01
2. Bentham M, Stansby G, Allen J. Innovative multi-site photoplethysmography analysis for quantifying pulse amplitude and timing variability characteristics in peripheral arterial disease. *Diseases.* (2018) 6:81. doi: 10.3390/diseases6030081
3. Sun Y, Thakor N. Photoplethysmography revisited: from contact to noncontact, from point to imaging. *IEEE Trans Biomed Eng.* (2016) 63:463–77. doi: 10.1109/TBME.2015.2476337
4. Elgendi M. On the analysis of fingertip photoplethysmogram signals. *Curr Cardiol Rev.* (2012) 8:14–25. doi: 10.2174/157340312801215782
5. Allen J. Thesis: *The Measurement and Analysis of Multi-site Photoplethysmographic Pulse Waveforms in Health and Arterial Disease.* Newcastle upon Tyne: Newcastle University (2002).

6. Elgendi M. Optimal signal quality index for photoplethysmogram signals. *Bioengineering (Basel)*. (2016) 3:21. doi: 10.3390/bioengineering3040021
7. Karlen W, Kobayashi K, Ansermino JM, Dumont GA. Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation. *Physiol Meas*. (2012) 33:1617–29. doi: 10.1088/0967-3334/33/10/1617
8. Li Q, Clifford GD. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiol Meas*. (2012) 33:1491–501. doi: 10.1088/0967-3334/33/9/1491
9. Orphanidou C. Quality *Assessment for the Photoplethysmogram (PPG). Signal Quality Assessment in Physiological Monitoring*. Nicosia: Springer Nature (2018). doi: 10.1007/978-3-319-68415-4
10. Sukor JA, Redmond SJ, Lovell NH. Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiol Meas*. (2011) 32:369–84. doi: 10.1088/0967-3334/32/3/008
11. Yongbo L, Chen Z, Liu G, Elgendi M. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. *Sci Data*. (2018) 5:180020. doi: 10.1038/sdata.2018.20
12. Yongbo L, Elgendi M, Chen Z, Ward R. An optimal filter for short photoplethysmogram signals. *Sci Data*. (2018) 5:180076. doi: 10.1038/sdata.2018.76
13. Allen J, Overbeck K, Nath AF, Murray A, Stansby G. A prospective comparison of bilateral photoplethysmography versus the ankle-brachial pressure index for detecting and quantifying lower limb peripheral arterial disease. *J Vasc Surg*. (2008) 47:794–802. doi: 10.1016/j.jvs.2007.11.057
14. Allen J, Oates CP, Lees TA, Murray A. Photoplethysmography detection of lower limbperipheral arterial occlusive disease: a comparison of pulse timing, amplitude and shape characteristics. *Physiol Meas*. (2005) 26:811–21. doi: 10.1088/0967-3334/26/5/018
15. van Gastel M, Stuijk S, de Haan G. Robust respiration detection from remote photoplethysmography. *Biomed Opt Expr*. (2016) 7:4941–57. doi: 10.1364/BOE.7.004941
16. Pollreisz D, Taheri. Nejad N. Detection and removal of motion artifacts in PPG signals. *Mobile Networks Appl*. (2019). doi: 10.1007/s11036-019-01323-6
17. Waugh W, Allen J, Wightman J, Sims AJ, Beale TAW. Novel signal noise reduction method through cluster analysis, appliedto photoplethysmography. *Comput Math Methods Med*. (2018) 2018:6812404. doi: 10.1155/2018/6812404
18. Allen J, Murray A. Effects of filtering on multi-site photoplethysmography pulse waveform characteristics. *IEEE Comput Cardiol*. (2004) 31:485–8. doi: 10.1109/CIC.2004.1442980
19. Weisberg, HF. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago, IL: University of Chicago Press (2005).
20. Assael H, Keon J. Nonsampling vs. sampling errors in survey research. *J Mark*. (1982) 46:114–23. doi: 10.1177/002224298204600212
21. Biemer P, Lyberg L. *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons, Inc. (2003).
22. Groves R, Fowler F, Couper M, Lepkowski J, Singer E, Tourangeau R. *Survey Methodology*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc. (2009).
23. Hogarth M. A note on aggregating opinions. *Organ Behav Hum Perf*. (1978) 21:40–6. doi: 10.1016/0030-5073(78)90037-5
24. Vogel C, Zwolinsky S, Griffiths C, Hobbs M, Henderson E, Wilkins E. A Delphi study to build consensus on the definition and use of big data in obesity research. *Int J Obesity*. (2019) 43:2573–86. doi: 10.1038/s41366-018-0313-9

Conflict of Interest: Between 2014 and 2018 JA was the Chief Investigator on an NIHR i4i funded grant (II-C1-0412-20003) to develop a miniaturized version of multi-site PPG pulse vascular measurement technology—specifically for peripheral arterial disease detection in a primary care setting.

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

Questionnaire Summary for PPG Morphological Analysis

1. When obtaining a PPG trace, what would you recommend as a minimum duration of recording?

Time: ☐ ≤10 s ☐ 11–30 s ☐ 31–60 s ☐ 2 min ☐ 5 min ☐ 10 min ☐ 20 min ☐ >20 min

2. Given the specified measurement setting over a 2-min period, what percentage of beats would you expect to be of good quality?

Finger site: ☐ ≤10 ☐ 11–25 ☐ 26–50 ☐ 51–75 ☐ 76–85 ☐ 86–95 ☐ >95

Toe site: ☐ ≤10 ☐ 11–25 ☐ 26–50 ☐ 51–75 ☐ 76–85 ☐ 86–95 ☐ >95

3. At what percentage of noise, relative to the number of pulse waves acquired, should we discard a recording and/or repeat it?

Finger site: ☐ ≤10 ☐ 11–25 ☐ 26–50 ☐ 51–75 ☐ 76–85 ☐ 86–95 ☐ >95

Toe site: ☐ ≤10 ☐ 11–25 ☐ 26–50 ☐ 51–75 ☐ 76–85 ☐ 86–95 ☐ >95

4. In the exploration of the photoplethysmographic pulse shape, what would you recommend as a minimum number of diagnostic quality pulse waves, in order that the average obtained is a true representation of the individuals PPG pulse shape?

Finger site: ☐ ≤10 ☐ 11–20 ☐ 21–30 ☐ 31–40 ☐ 41–50 ☐ 51–60 ☐ >60

Toe site: ☐ ≤10 ☐ 11–20 ☐ 21–30 ☐ 31–40 ☐ 41–50 ☐ 51–60 ☐ >60

5. In the exploration of the photoplethysmographic pulse shape, what would you recommend as a minimum number of successive diagnostic quality pulse waves, in order that the average obtained is a true representation of the individuals PPG pulse shape?

Finger site: ☐ ≤10 ☐ 11–20 ☐ 21–30 ☐ 31–40 ☐ 41–50 ☐ 51–60 ☐ >60

Toe site: ☐ ≤10 ☐ 11–20 ☐ 21–30 ☐ 31–40 ☐ 41–50 ☐ 51–60 ☐ >60

6. What minimum number of respiratory cycles would you recommend recording a PPG trace for?

Finger site: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6–10 ☐ >10
Toe site: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6–10 ☐ >10

frontiers in Digital Health

METHODS
published: 23 December 2020
doi: 10.3389/fdgth.2020.610956

Impact of Data Transformation: An ECG Heartbeat Classification Approach

*Yongbo Liang[1], Ahmed Hussain[2], Derek Abbott[3,4], Carlo Menon[5], Rabab Ward[1] and Mohamed Elgendi[1,2,5,6]**

[1] School of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada, [2] Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada, [3] School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA, Australia, [4] Centre for Biomedical Engineering, The University of Adelaide, Adelaide, SA, Australia, [5] Menrva Research Group, School of Mechatronic Systems Engineering and Engineering Science, Simon Fraser University, Surrey, BC, Canada, [6] British Columbia Children's and Women's Hospital, Vancouver, BC, Canada

Cardiovascular diseases continue to be a significant global health threat. The electrocardiogram (ECG) signal is a physiological signal that plays a major role in preventing severe and even fatal heart diseases. The purpose of this research is to explore a simple mathematical feature transformation that could be applied to ECG signal segments in order to improve the detection accuracy of heartbeats, which could facilitate automated heart disease diagnosis. Six different mathematical transformation methods were examined and analyzed using 10s-length ECG segments, which showed that a reciprocal transformation results in consistently better classification performance for normal vs. atrial fibrillation beats and normal vs. atrial premature beats, when compared to untransformed features. The second best data transformation in terms of heartbeat detection accuracy was the cubic transformation. Results showed that applying the logarithmic transformation, which is considered the go-to data transformation, was not optimal among the six data transformations. Using the optimal data transformation, the reciprocal, can lead to a 35.6% accuracy improvement. According to the overall comparison tested by different feature engineering methods, classifiers, and different dataset sizes, performance improvement also reached 4.7%. Therefore, adding a simple data transformation step, such as the reciprocal or cubic, to the extracted features can improve current automated heartbeat classification in a timely manner.

Keywords: feature mapping, feature representation, feature transformation, feature conversion, feature restructuring, data Wrangling

OPEN ACCESS

Edited by:
Anders Nordahl-Hansen,
Østfold University College, Norway

Reviewed by:
Erick Andres Perez Alday,
Emory University, United States
Paolo Crippa,
Marche Polytechnic University, Italy

***Correspondence:**
Mohamed Elgendi
mohamed.elgendi@cw.bc.ca

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

Received: *30 September 2020*
Accepted: *03 December 2020*
Published: *23 December 2020*

Citation:
Liang Y, Hussain A, Abbott D, Menon C, Ward R and Elgendi M (2020) Impact of Data Transformation: An ECG Heartbeat Classification Approach. Front. Digit. Health 2:610956. doi: 10.3389/fdgth.2020.610956

INTRODUCTION

Electrocardiographs (ECGs) have been a staple in medical practice for around a century. A complete heartbeat process is initiated by the sinus node—consisting of the depolarization of atriums and ventricles and the repolarization of the ventricles—in which atrial depolarization forms a P wave, ventricular depolarization forms a QRS complex wave, and the repolarization of the ventricles forms a T wave. Since its inception, ECGs have been used to diagnose physical heart abnormalities (1). When beats conform to the basic structure of a QRS complex, they are called normal beats; otherwise, they may be called arrhythmic. In an arrhythmic heartbeat—such as a beat that occurs too fast, too slow, or is irregularly timed—the morphology of the ECG waves changes accordingly.

The accurate determination of beat types can assist in the diagnosis of ECG signals. However, a more simplistic and accurate way to distinguish the heartbeats is still an unmet need. Past research has explored several ECG morphological features and many complex classifiers for achieving higher classification performance. Note that ECG morphological features (2), including RR-interval features (3) and PT-interval features (4). have been proposed, and some complex classification models, such as artificial neural networks (5), extreme learning machines (6), and deep neural networks (7) have been adopted. Although these methods can achieve slight performance improvements, the heavy computation (requiring off-line processing) limits the application of these methods. Due to the development of mobile medical technology, a greater need for robust lower computational overhead is emerging.

The rapidly increasing accessibility of mobile devices and their use in classifying different types of beats makes the investigation of different potential patterns in ECG data especially important. Any simple mathematical model can be incorporated in a mobile app to provide preliminary diagnoses of heart-related problems; this would offer patients awareness of a problem before receiving a formal diagnosis from a physician. It could also be used to promote timely self-treatment (e.g., electrolytic rebalance and breathing techniques). Recently, in 2019, Oscar et al. noted that the total number of smartphone users worldwide was projected to surpass 2.5 billion. Furthermore, the United States of America found that, as of 2017, ∼64% of its population uses smartphones (8). Hence, the transformation method developed in this study was aimed for use in the growing field of mobile health. Therefore, this study focuses on whether it is possible to devise a simple feature transformation method that can improve the classification of different heartbeat events using multiple feature calculations of ECG signals without the need for complex algorithms that require high computational power to achieve similar results.

MATERIALS AND METHODS

Hypotheses

To our knowledge, no study has investigated feature transformations to classify different heartbeat events. The main research question is: "What is the simplest mathematical transformation that can improve classification performance compared with the original features?"

Feature Engineering

ECG signals contain a wealth of heartbeat process information, and high-quality, clear ECG signals can be used for the diagnosis and evaluation of a variety of heart diseases. However, in some cases, researchers can obtain only a small portion of the total number of ECG features. For example, ECG signals can be obtained from wearable ECG devices, but they are generally of low quality and have high interference rates, thus making it difficult to extract more accurate and effective information. In this study, to reduce the effects of noise, an 8th order 0.5–30 Hz bandpass Butterworth filter was used for the raw ECG signal (9). Meanwhile, the raw ECG signal without any filter to process (unfiltered data) was also explored in this study in order to compare the information with the filtered ECG data. Feature engineering methods in time-series biomedical signals are a common step in; for example, extraction of skewness, kurtosis, entropy, signal-noise ratio, and so on, all of which have been applied to ECG (10), photoplethysmogram (PPG) (11) and electroencephalogram (EEG) (12) signals. In this study, we adopted six feature-engineering methods to calculate ECG features based on unfiltered and filtered ECG signals (13), such as skewness (f_S), kurtosis (f_K), entropy (f_E), the zero-crossing rate (f_Z), the signal-to-noise ratio (f_N), and relative power (f_R). The formulas of these features are as follows:

1. **Skewness (f_S)**

$$f_S = 1/N \sum_{n=1}^{N} [x[n] - \hat{\mu}_x/\sigma]^3 \tag{1}$$

where $\hat{\mu}_x$ and σ are the empirical estimates of the mean and standard deviation of x, respectively, and N is the number of sampling points in unfiltered and filtered ECG signals.

2. **Kurtosis (f_K)**

$$f_S = 1/N \sum_{n=1}^{N} [x[n] - \hat{\mu}_x/\sigma]^4 \tag{2}$$

where $\hat{\mu}_x$ and σ are the empirical estimates of the mean and standard deviation of x, respectively, and N is the number of sampling points in unfiltered and filtered ECG signals.

3. **Entropy (f_E)**

$$f_E = -\sum_{n=1}^{N} x[n]^2 \log_e(x[n]^2) \tag{3}$$

where x is the unfiltered and filtered ECG signal and N is the number of sampling points.

4. **Zero-crossing rate (f_Z)**

$$f_Z = 1/N \sum_{n=1}^{N} \Pi\{y < 0\} \tag{4}$$

where y is the filtered ECG signal of length N and Π–the indicator function $\Pi\{A\}$–is 1 if its argument A is true, and 0 otherwise.

5. **Signal-to-noise ratio (f_N)**

$$f_N = \sigma^2_{signal}/\sigma^2_{noise} \tag{5}$$

where σ_{signal} is the standard deviation of the absolute value of the unfiltered and filtered ECG signal (y) and σ_{noise} is the standard deviation of the y signal.

6. **Relative power (f_R)**

Because most of the energy of the ECG signal is concentrated within the 5–15 Hz frequency band, the ratio of the power spectral density (PSD) in this band to the PSD of the overall 1–40 Hz signal provides a measure of f_R:

$$f_R = \sum_{f=5}^{15} \mathrm{PSD} / \sum_{f=1}^{40} \mathrm{PSD} \tag{6}$$

where PSD is calculated using Welch's method.

Feature Transformation

To improve features' separability, a feature transformation is applied to convert an original feature to a high dimensional space. The original input features, obtained from the previous subsection, are written as f. Feature transformation is a function of the input attributes $\varphi(f)$, defined as follows:

$$\varnothing(f) = \begin{bmatrix} f \\ \ln(f) \\ \frac{1}{f} \\ \sqrt{f} \\ f^2 \\ f^3 \\ asin(f) \end{bmatrix} \tag{7}$$

where f is the tested feature. This study explores whether the classification performance can be improved based only on mathematical transformations, without the use of other external features. We investigated six data transformations based on the recommendations in (14, 15).

Database

The ECG dataset used in this study was obtained from the MIT-BIH Arrhythmia Database (16, 17), which includes ECG signals and corresponding annotated beat types. The database is made up of 30-min ECG recordings from 48 patients. For the consistency of feature calculations, we calculated the features from the 10s-long ECG signal segments; each segment was part of a separate heartbeat category. Some categories only had a small sample size, so only normal beats (Norm), atrial fibrillation (AF), atrial premature beats (APBs), and premature ventricular contractions (PVCs) were included in this study, as they had the largest sample sizes. **Table 1** shows the statistics of the heartbeat segments in the MIT-BIH Arrhythmia Database, in which we can see that the number of samples in the different categories varies greatly. Specifically, the number of Norms (category 1) is four times larger than that of APBs (category 4) and two times larger than that of AFs (category 2) and PVCs (category 3). An unbalanced dataset can incorrectly represent classification performance and, because a balanced dataset was needed, a resampling technique was required. However, under-sampling and oversampling have their own flaws (18), and in order to reduce the impact of a resampling technique, two different techniques were adopted to balance the data, namely the random under-sampling technique (RUS) and the synthetic minority oversampling technique (SMOTE) (19). For the RUS process, the samples of category 1 (Norm) were resampled randomly according to the numbers in category 2 (AF), category 3 (PVC), and category 4 (APB) to classify each other. For the SMOTE process, categories 2, 3, and 4 were resampled according to the number in category 1. The RUS and SMOTE random samples were generated using MATLAB version R2019a (The MathWorks, Inc., MA, USA). Finally, RUS-balanced and SMOTE-balanced datasets were generated and used to classify the different categories, as shown in **Table 2**.

TABLE 1 | The statistics of heartbeat segments in the MIT-BIH arrhythmia database (10).

Index	Heartbeat type	Description	Number of segments
1	Norm	Normal beat	283
2	AF	Atrial fibrillation	135
3	PVC	Premature ventricular contraction (PVC)	133
4	APB	Atrial premature beat	66
Total	(All)	-	617

Note that the segment in this table means the 10 s length ECG segment.

Feature Evaluation

The original features were calculated first using six feature-engineering methods and were named f_S, f_K, f_E, f_Z, f_N, and f_R. New features were then developed from different dimensionalities based on these original features, including logarithmic [$\ln(f)$], reciprocal ($1/f$), square-root ($\sqrt{f}$), square (f^2), cube (f^3), and arcsine [$\mathrm{asin}(f)$] calculations. To classify the different ECG categories, these ECG features were extracted and constructed based on the different feature-engineering methods and mathematical transformations. The features were evaluated by classifying the different heartbeat categories using multiple linear and non-linear classifiers.

The dataset—and each category within the dataset—was divided into a training set (70%) and a testing set (30%). In the training set, 10-fold cross-validation was adopted to validate the generalization ability of the trained classifier. In the testing phase, the performance evaluation was based on the testing set by the trained model. The F1 score was calculated as an evaluation measure as follows:

$$F1 = 2 \times Recall \times Precision/(Recall + Precision) \tag{8}$$

where precision = TP/(TP + FP) and recall = TP/(TP + FN). Here, TP stands for true positives, FP stands for false positives, and FN stands for false negatives.

In order to compare the change in the classification performance, the difference measure (D-value) was adopted and calculated as follows:

$$D\ value = F1_{transformed\ feature} - F1_{original\ feature} \tag{9}$$

If D is positive (D > 0) then the transformed feature improved the F1 accuracy, if D is zero, then transformed feature scored the

TABLE 2 | The statistics of different categories in this study.

Trial number	Heartbeat type	Number of segments (unbalanced dataset)	Number of segments (RUS dataset)	Number of segments (SMOTE dataset)
1	Norm vs. AF	283 vs. 135	135 vs. 135	283 vs. 283
2	Norm vs. PVC	283 vs. 133	133 vs. 133	283 vs. 283
3	Norm vs. APB	283 vs. 66	66 vs. 66	283 vs. 283

Norm, AF, PVC, and APB each represent normal beat, atrial fibrillation beat, Premature ventricular contraction, and Atrial premature beat. The unbalanced column shows that the quantity of different categories varies greatly; the balanced column shows an approximately equal number of different categories after random sampling. RUS represented random under-sampling and SMOTE represented Synthetic Minority Oversampling Technique.

same F1 score as the original feature, and if D is negative ($D < 0$), then the transformed feature scored less F1 score than the original feature. Certainly, the main goal of this study is to find the transformation method that is consistently achieve a D value > 0, regardless of the feature extraction method, classifier, signal quality, and sampling technique.

Classification

We used five linear and non-linear classifiers to evaluate the different feature-engineering methods and mathematical transformations. The classifiers were the k-nearest neighbor (KNN), neural net (NN), support vector machine (SVM), decision tree (TREE), and Naïve Bayes (NB). For the KNN classifier, the number of neighbor points was set to 10. The NN classifier was a feedforward neural network with an input layer, a hidden layer with 10 neurons, and an output layer. The SVM classifier with a quadratic kernel was used. For the TREE, the split criterion of TREE was Gini's diversity index (gdi); the maximal number of decision splits was 100. For the Naïve Bayes classifer, the kernel smoother type was the Gaussian kernel, and the kernel smoothing density support was unbounded. In training the model, 10-fold cross-validation was used, which protected against overfitting by partitioning the dataset into multiple parts and estimating the accuracy of each fold. Code written in MATLAB was used to perform the feature evaluation and model training. **Figure 1** shows a work flowchart of this study.

Data Availability

The MIT-BIH Arrhythmia Database is publicly available and can be downloaded from https://www.physionet.org/physiobank/database/mitdb/.

RESULTS AND DISCUSSION

Each feature was used to classify different heartbeat categories, and the F1 score for each of the feature classifications was recorded and summarized in Appendix **Table A**. In Appendix **Tables A, A.1–A.3** showed the performance of three classification trials (Norm vs. APB, Norm vs. AF, and Norm vs. PVC) that were achieved based on the unbalanced dataset. **Tables A.4–A.6** showed the performance of three classification trials that were achieved based on the RUS balanced dataset. **Tables A.7–A.9** showed the performance of three classification trials that were achieved based on the SMOTE balanced dataset. Each table in Appendix **Table A** contained the F1 score of five classifiers that were achieved by unfiltered and filtered ECG signals. The D-value in the table was calculated based on the maximum F1 score of classifiers between the original feature and the transformed features. For the original feature, the D-value was always zero as the F1 score is subtracted from itself, while the transformed feature had either a positive, negative, or zero D value. A positive value meant that the transformed feature improved the classification performance, and a negative value meant that the transformed feature was not helpful in classification. The first and last columns of Appendix **Table A** are the D-values. The results show that not all the mathematical transformations were consistently effective. We set the original feature as the baseline of classification with a D-value of zero. The D-value after the mathematical transformations was either positive or negative. For example, the D-value of f_N had the most negative value, which showed that the mathematical transformations for the signal-to-noise ratio feature were not helpful.

When we analyzed Appendix **Table A**, we found some interesting changes for filter processing, feature engineering, feature transformations, and different classifiers. In **Table A.2**, the F1 score of f_S is only 40% by the KNN classifier. However, the reciprocal of f_S ($\frac{1}{f_S}$) improves the F1 score to 75.6%. In **Table A.8**, the F1 score of the same solution improved from 71.6 to 91.9%. From other tables, we also found a 10% improvement of KNN after the reciprocal transformation of skewness. For filter processing, the F1 score of KNN also had a big improvement. In **Table A.1**, the F1 score of KNN for f_Z, calculated by an unfiltered ECG signal, was only 18.2%, which means that the solution can't work well at all. However, after the filtering process, the F1 score of KNN reached 70%. In **Table A.4**, the F1 score of KNN for f_Z also improved from 46.5 to 73.7% after filtering. In addition, when we see the F1 score difference in classifiers, we found that KNN was not the best classifier, as it only shows a poor result compared to other classifiers. However, when we adopted some proper processing, such as filtering, transformation, and resampling methods, the performance of KNN increased from 18.2 to 95.2%, which was higher than other classifiers found in **Tables A.1, A.4, A.7**.

Actually, other feature engineering methods, except f_Z, don't improve the performance of KNN. This tells us that the use of a zero-crossing rate (f_Z) and a KNN classifier should be concerned with the noise of physiological signals, and a suitable filter should be implemented first. It is an imperative that a solution regarding the combining of optimal filter, feature engineering, feature

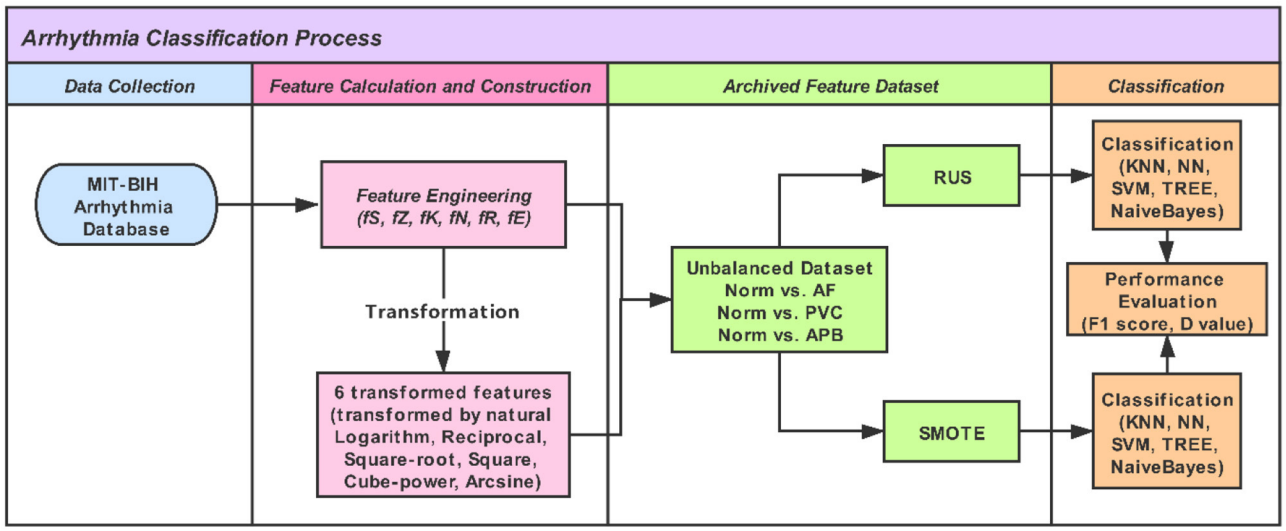


FIGURE 1 | A work flowchart of this study.

transformation, and classifiers achieve better performance. In addition, different feature engineering methods show different characteristics. The reciprocal transformation of skewness ($\frac{1}{f_S}$) significantly improves the F1 score. However, the reciprocal transformation of other feature-engineering methods does not show improvement at all, or only shows minor improvement. For example, in **Table A.2**, the reciprocal transformation of skewness improved the F1 score by 35.6% for the KNN classifier, and improved the F1 score by more than 10% for other classifiers. **Tables A.1, A.4, A.5, A.7–A.9** also show similar changes.

We also analyzed cases of performance reduction. In **Table A.4**, the D-value of $(f_K)^2$ and the original f_K was −15.1%, which demonstrates that the square transformation does not provide any help for the classification. From **Table A.2**, $\ln(f_Z)$ reduces the F1 score by 4.3%, and other transformed features of f_Z also had a negative D-value. As mentioned above regarding f_Z, the logarithm transformation of f_Z wasn't helpful in improving the classification, which was not the best choice of mathematical transformation.

For different classification issues, reciprocal transformation did not improve the classification accuracy of all trials. In **Table A.9**, the cube transformation made the highest improvement for Norm vs. PVC classification by reaching about 9.3%. Although the reciprocal transformation also made an improvement, the improvement was only about 3.8%. In **Tables A.3, A.6**, cube transformation was also superior to the reciprocal transformation. According to this analysis, cube transformation had an advantage in classifying Norm and PVC compared with AF and APB.

As observed above, there was no one transformation to solve all problems. Based on Appendix **Table A**, we summarize the D-value as follows below:

A. **D-value = 0 (No classification accuracy improvement)**

The D-value being equal to zero means that applying feature transformation can lead to no improvement. In fact, transformed features can only provide the same result as the feature itself. As can be seen in **Table A.1**, when the square root was applied to the zero-crossing feature, the F1 score of f_Z and $\sqrt{f_Z}$ were the same result (89.8%); this was achieved by the Naïve Bayes classifier. Feature transformation did not improve the classification accuracy.

For the filtered process, similar results are shown in **Tables A.2, A.3, A.5, A.7**; when f_S was transformed to $\sqrt{f_S}$, the F1 score did not improve and the D-value was equal to zero. For the unfiltered process, there were also similar results, as shown in Tables A.1–A.3, A.6, A.7. The square root did not improve the performance when compared to the original feature.

B. **D-value > 0 (Classification Improvement)**

The D-value > 0 meant that the transformation improved the classification performance when compared to the original feature. We easily found the phenomenon in **Table A.2** when the f_Z was transformed to $\frac{1}{f_S}$, the F1 score improved from 80.8 to 89.9%. The D-value reached up to 9.1%. In **Table A.4**, the D-value of $(f_N)^3$ and f_N reached up to 12.7% and $\frac{1}{f_S}$ obtained a 16% D-value in **Table A.5**. In **Table A.8**, the same transformation of $\frac{1}{f_S}$ improved the F1 score from 71.6 to 91.9%, which obtained a 20.3% D-value. More positive D-values can be found in **Tables A.2–9**.

C. **D-value < 0 (Negatively impact classification accuracy)**

In Appendix **Table A**, it's easy to find some negative D-values. Negative D-values mean that the transformation is terrible and is not helpful in classifying. It demonstrates that it is better not to transform for an inappropriate transformation. In **Table A.4**,

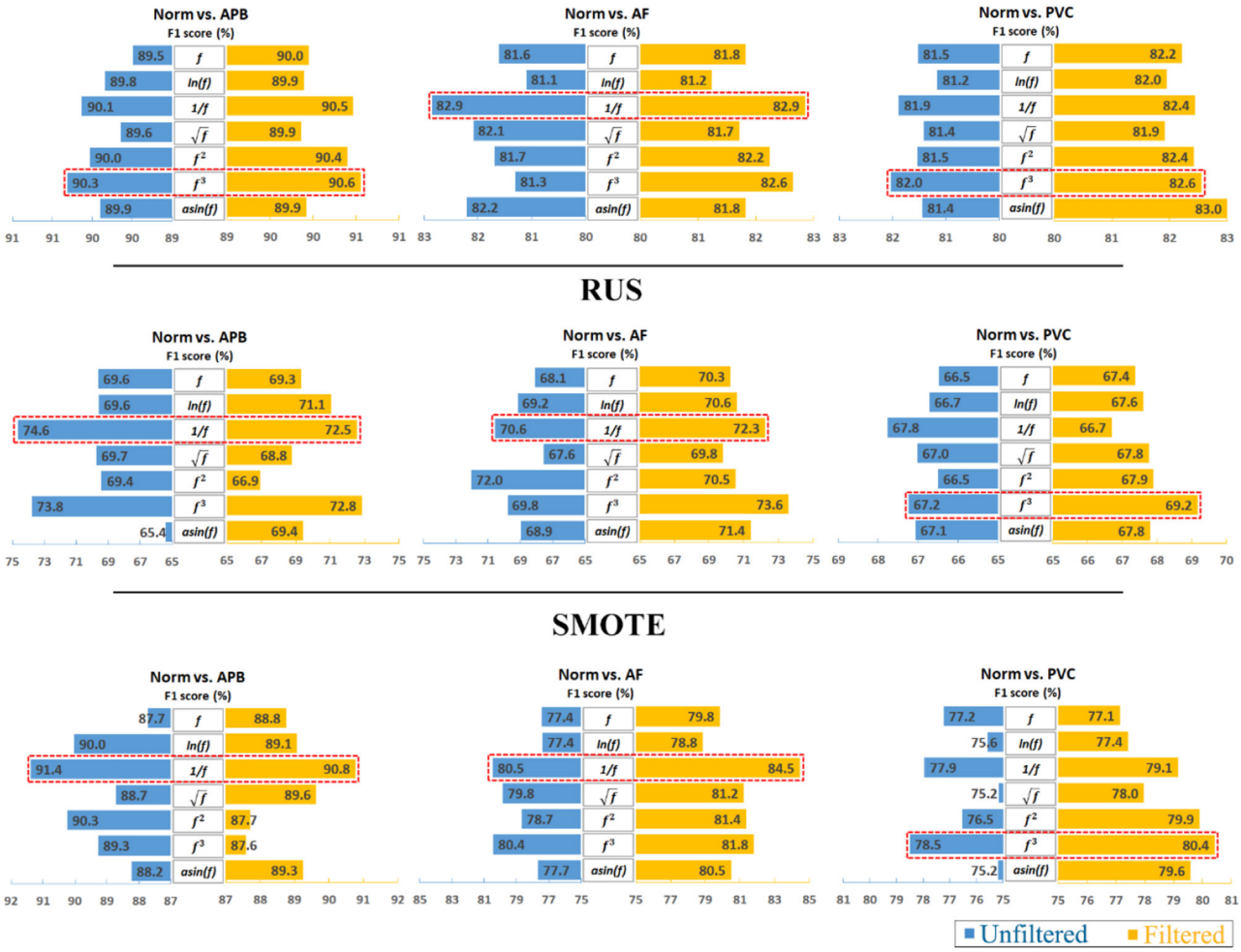


FIGURE 2 | Comparison of classification performance between unfiltered data and filtered data. Note that the reciprocal is the most consistently effective data transformation.

$(f_K)^2$ obtained a lower F1 score, from 68.1 to 52.9%, and the D-value was −15.2%. Similarly, $(f_N)^3$ decreased the F1 score from 74.2 to 65.9% in **Table A.5**, and its D-value was −8.3%.

The D-value clearly shows the change of different solutions. This performance difference occurs due to the intercorrelation between the classification, feature engineering methods, and transformations. The optimal combination is valuable to explore when the available methods are limited, such as in the application with low computation and low battery.

Furthermore, to analyze the results clearly, an average calculation for an overall analysis was conducted based on the six feature engineering methods, and the averaged results were plotted in **Figure 2**. Meanwhile, the corresponding D-values between original and transformed features were shown in **Figure 3**. From **Figures 2**, **3**, we see that the filter improved the performance for the unbalanced, RUS, and SMOTE. Overall, the SMOTE achieved greater improvement than the unbalanced and RUS balanced datasets. In addition, most of the transformed features achieved a positive improvement compared to the original feature. In **Figure 2**, the red box pointed out the best transformation for each figure. According to the statistics of F1 score, $\frac{1}{f}$ and f^3 made greater improvements, which were 4.7% (Norm vs. AF) and 3.3% (Norm vs. PVC). Furthermore, $\frac{1}{f}$ achieved the five best results, and f^3 achieved the four best results. From these two sides, $\frac{1}{f}$ was the most stable mathematical transformation. **Figure 4** shows the average F1 score of the unbalanced, RUS, and SMOTE datasets, which were a further overall calculation based on **Figure 2**. And **Figure 5** shows the corresponding D-values between original and transformed features. The performance of Norm vs. APB, Norm vs. AF, and Norm vs. PVC were improved by 1.9, 2.6, and 1.8% by $\frac{1}{f}$, $\frac{1}{f}$, and f^3, respectively.

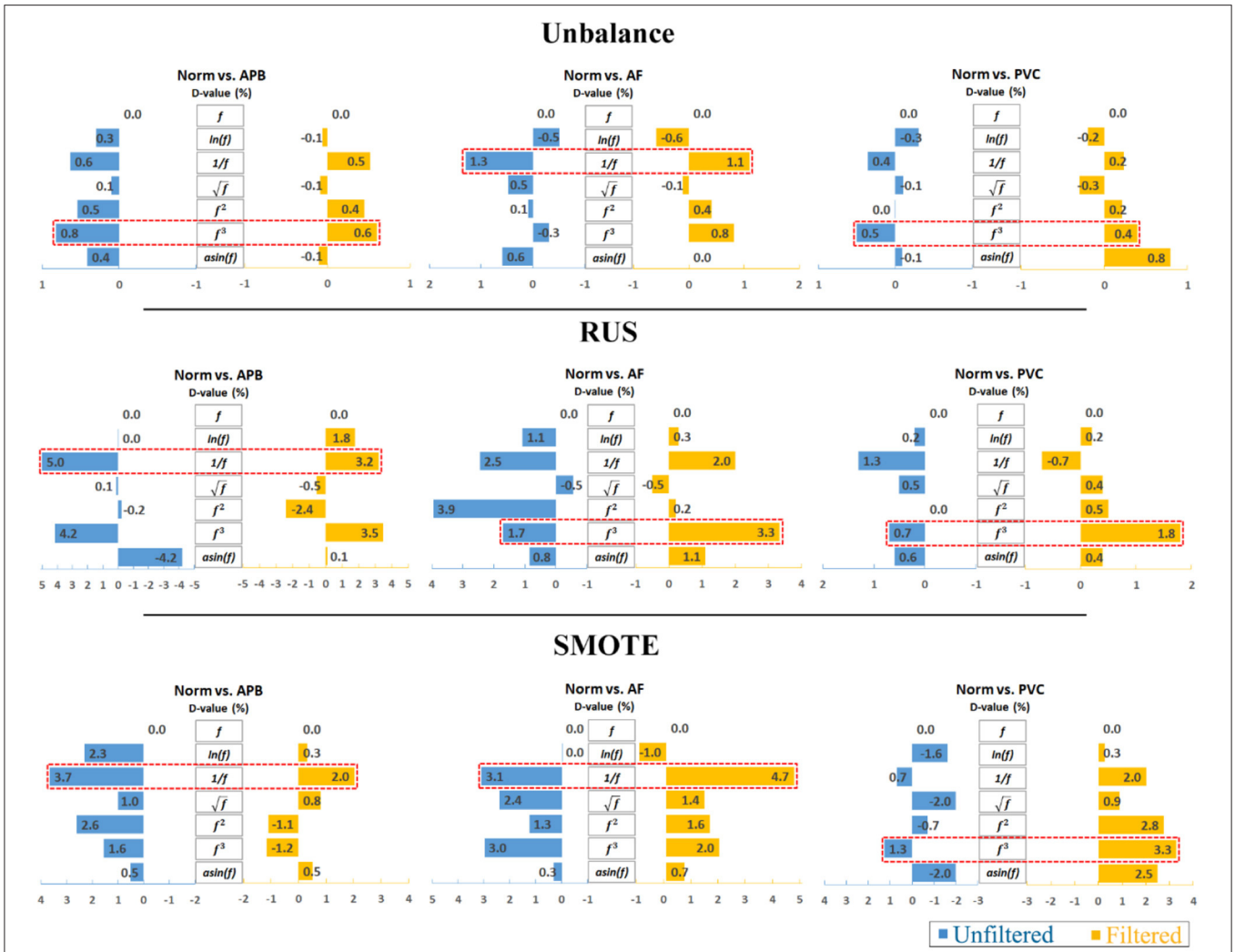


FIGURE 3 | D-value comparison between the original and transformed features. Note that the reciprocal is the most consistently effective data transformation.

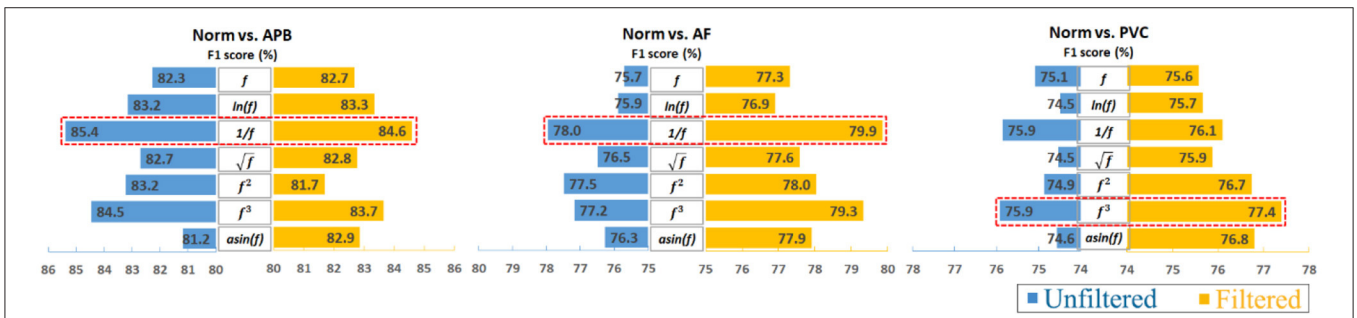


FIGURE 4 | Overall impact of data transformations on classifying ECG heartbeats by averaging the unbalanced, RUS, and SMOTE results. Note that the reciprocal is the most consistently effective data transformation.

Most of the previous research on this topic focused on the RR interval and various other morphological features. Because of the broad spectrum of varying heart diseases—as well as signal noise—ECG waveform morphologies can vary greatly. One of the most difficult issues in this field of study is the correct extraction of ECG morphological features. Although some researchers have achieved higher classification performances using methods based on hundreds of ECG morphological features (20–23), ECG

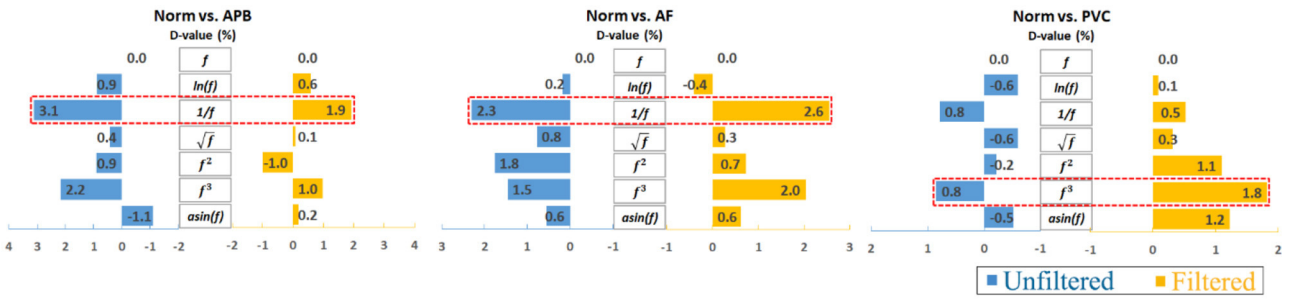


FIGURE 5 | Overall D-value changes after averaging the unbalanced, RUS, and SMOTE results. Note that the reciprocal is the most consistently effective data transformation.

signals with high levels of noise and many morphological features are usually incorrectly extracted (7, 24, 25), and this affects the robustness and effectiveness of the morphological method used. A simple feature extraction followed by optimal mathematical transformation, as this paper proposes, could be a new way to improve the detection of heart rate abnormalities.

A simple mathematical transformation, as discussed in this study, would likely not use significant processing energy or battery life and would be especially helpful in situations in which time and mobile battery life are critical. In contrast, complex calculations and classifiers, which take longer to run on even high-performance computers, would likely be less useful, as mobile devices are usually used for their simplicity and quickness, both of which the algorithm in this study promotes. However, mobile processing is evolving rapidly, so it may be possible to incorporate more complex methods into mobile devices in the near future. It is therefore important to continue improving the accuracy and reproducibility of the transformed features so they can compete with more complex methods while having the advantage of speed and simplicity.

Our findings are significant as they show the impact of mathematical transformations on extracted features and the overall accuracy in detecting abnormal heartbeats. Calculating the optimal feature (such as skewness in our study) and the optimal mathematical feature transformation is easily programmable into simple-to-use heart activity detection devices, unlike their more complex counterparts. As mentioned earlier, considering battery life, processing power, and the urgent need for an algorithm that will quickly classify heartbeats, whether for personal or medical purposes, the reciprocal feature-engineering method will perform as desired, especially considering the stable improvement in classification performance of the transformed feature.

It is also important to note that AF, PVC, and APB are some of the most prevalent arrhythmic conditions, and this algorithm will definitely be useful in the medical and self-care industry. With the development of digital health technology, wearable ECG devices are proliferating rapidly. The reciprocal transformation method used in this study is useful in this context for promoting the real-time detection of heart activity by reducing complexity and improving accuracy.

The findings of this study do not mean that we always have to use reciprocal transformation to improve classification accuracy. However, it shows the importance of feature transformation. Meanwhile, it also has some disadvantages and limitations. First, the sample size is small in this study, and more clinical data is needed to explore these new findings. Our study focused on ECG segments that are 10s in length, which only contain three types of abnormal heartbeats. Future research will explore other types of abnormalities. Second, the dataset is not balanced, which required the use of methods to upsample and downsample the dataset. Third, reciprocal transformation is applicable in solving the AF and APB detection, but it is not helpful in recognizing the PVC category. In other ways, the cube transformation is more favorable in solving the PVC category recognition. It is important to note that this is an interrelated optimization problem. In addition, given the data we have, we found that the reciprocal improves feature sensitivity and separability, which is beneficial to the classification. However, only AF, APB, and PVC were studied. More abnormal heartbeat events should be explored and validated in future work. Meanwhile, for upcoming studies, regardless of the application, we recommend testing the reciprocal as a feature transformation step and examining different classifiers for better classification results.

CONCLUSION

We tested the hypothesis that a simple mathematical transformation can lead to better heartbeat classification, which could improve studies that rely on transformed features as biomarkers. Six mathematical transformations were evaluated for heart activity classification performance, and we found that some processing steps for original f_Z made the F1 score improve from 18.2 to 95.2%. Meanwhile, we also found that reciprocal $\frac{1}{f}$ improved the overall (tested over different feature engineering methods and different dataset sizes) classification accuracy to 4.7%. The main finding was that the application of a reciprocal transformation to features extracted from the ECG signals improved heartbeat classification consistently. The proposed extra mathematical step is therefore useful for big data analytics and can be easily incorporated into mobile and portable health applications.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YL, AH, and ME processed the signals and drafted the manuscript. DA, CM, and RW advised and supervised the project. ME designed the experiment and led this investigation. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the NSERC grant RGPIN-2014-04462 and Canada Research Chairs (CRC) program. RW was supported by NPRP grant # NPRP12S-0305-190231 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2020.610956/full#supplementary-material

REFERENCES

1. Klabunde RE. *Cardiovascular Physiology Concepts*, *2nd edn*. Lippincott Williams & Wilkins (2011).
2. Akhoondi MPaF. Providing an efficient algorithm for finding R peaks in ECG signals and detecting ventricular abnormalities with morphological features. *J Med Signals Sensors.* (2016) 6:218–23. doi: 10.4103/2228-7477.195090
3. Tateno K, Glass L. A method for detection of atrial fibrillation using RR intervals. *Comp Cardiol.* (2000) 27:391–4. doi: 10.1016/j.amjcard.2011.01.028
4. Elgendi M, Jonkman M, De Boer F. In: *2008 7th IEEE International Conference on Cognitive Informatics.* (2008). p. 83–8.
5. Krishna Prasad GJSS. Classification of ECG arrhythmias using multi-resolution analysis and neural networks. In *TENCON 2003*. Bangalore (2003). p. 227–31.
6. Kim J, Shin HS, Shin K, Lee M. Robust algorithm for arrhythmia classification in ECG using extreme learning machine. *Biomed Eng Online.* (2009) 8:31. doi: 10.1186/1475-925X-8-31
7. Rahhal MMA, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR, et al. Deep learning approach for active classification of electrocardiogram signals. *Inform Sci.* (2016) 345:340–54. doi: 10.1016/j.ins.2016.01.082
8. Trespalacios O, Nandavar S, Albert Newton JD, Demant D, Phillips JG. Problematic use of mobile phones in Australia…is it getting worse? *Front Psychiatry.* (2019) 10:1–15. doi: 10.3389/fpsyt.2019.00105
9. Elgendi M, Meo M, Abbott D. A proof-of-concept study: simple and effective detection of P and T waves in arrhythmic ECG signals. *Bioengineering.* (2016) 3:26. doi: 10.3390/bioengineering3040026
10. Zhao Z, Zhang Y. SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. *Front Physiol.* (2018) 9:727. doi: 10.3389/fphys.2018.00727
11. Elgendi M. Optimal signal quality index for photoplethysmogram signals. *Bioengineering.* (2016) 3:1–15. doi: 10.3390/bioengineering3040021
12. Grosselin F, Navarro-Sune X, Vozzi A, Pandremmenou K, De Vico Fallani F, Attal Y, et al. Quality assessment of single-channel EEG for wearable devices. *Sensors.* (2019) 19:601. doi: 10.3390/s19030601
13. Lyon A, Mincholé A, Martínez JP, Laguna P, Rodriguez B. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *J R Soc Interface.* (2018) 15:20170821. doi: 10.1098/rsif.2017.0821
14. Manikandan S. Data transformation. *J Pharmacol Pharmacotherapeut.* (2010) 1:126. doi: 10.4103/0976-500X.72373
15. McDonald JH. *Handbook of Biological Statistics, Vol. 2.* Baltimore, MD: Sparky House Publishing (2009).
16. Goldberger AL, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* (2000) 101:e215–20. doi: 10.1161/01.CIR.101.23.e215
17. Moody GBMR. The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol.* (2001) 20:45–50. doi: 10.1109/51.932724
18. Rajesh KNVPS, Dhuli R. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomed Signal Process Control.* 41:242–54. doi: 10.1016/j.bspc.2017.12.004
19. Alberto Fern'andez SG, Nitesh FH, Chawla V. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-years anniversary. *J Artificial Intelligence Res.* (2018) 61:863–905. doi: 10.1613/jair.1.11192
20. Tsipouras MGDF, Sideris D. Arrhythmia classification using the RR-interval duration signal. *Comput Cardiol.* (2002) 29:485–8. doi: 10.1109/CIC.2002.1166815
21. Cuesta P, Lado MJ, Vila XA, Alonso R. Detection of premature ventricular contractions using the RR-interval signal: a simple algorithm for mobile devices. *Technol Health Care.* (2014) 22:651–6. doi: 10.3233/THC-140818
22. deChazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans Biomed Eng.* (2004) 51:1196–206. doi: 10.1109/TBME.2004.827359
23. Asl BM, Setarehdan SK, Mohebbi, M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence Med.* (2008) 44:51–64. doi: 10.1016/j.artmed.2008.04.007
24. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Inform Sci.* (2017) 405:81–90. doi: 10.1016/j.ins.2017.04.012
25. Luz EJdS, Nunes TM, de Albuquerque VHC, Papa JP, Menotti D. ECG arrhythmia classification based on optimum-path forest. *Expert Syst Appl.* (2012) 40:3561–73. doi: 10.1016/j.eswa.2012.12.063

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: 31 March 2021
doi: 10.3389/fdgth.2021.610006

Remote Digital Measurement of Facial and Vocal Markers of Major Depressive Disorder Severity and Treatment Response: A Pilot Study

Anzar Abbas[1], Colin Sauder[2,3], Vijay Yadav[1], Vidya Koesmahargyo[1], Allison Aghjayan[2], Serena Marecki[2], Miriam Evans[2] and Isaac R. Galatzer-Levy[1,4]*

[1] AiCure, New York, NY, United States, [2] Adams Clinical, Watertown, MA, United States, [3] Karuna Therapeutics, Boston, MA, United States, [4] Psychiatry, New York University School of Medicine, New York, NY, United States

Objectives: Multiple machine learning-based visual and auditory digital markers have demonstrated associations between major depressive disorder (MDD) status and severity. The current study examines if such measurements can quantify response to antidepressant treatment (ADT) with selective serotonin reuptake inhibitors (SSRIs) and serotonin–norepinephrine uptake inhibitors (SNRIs).

Methods: Visual and auditory markers were acquired through an automated smartphone task that measures facial, vocal, and head movement characteristics across 4 weeks of treatment (with time points at baseline, 2 weeks, and 4 weeks) on ADT ($n = 18$). MDD diagnosis was confirmed using the Mini-International Neuropsychiatric Interview (MINI), and the Montgomery–Åsberg Depression Rating Scale (MADRS) was collected concordantly to assess changes in MDD severity.

Results: Patient responses to ADT demonstrated clinically and statistically significant changes in the MADRS [$F_{(2, 34)} = 51.62$, $p < 0.0001$]. Additionally, patients demonstrated significant increases in multiple digital markers including facial expressivity, head movement, and amount of speech. Finally, patients demonstrated significantly decreased frequency of fear and anger facial expressions.

Conclusion: Digital markers associated with MDD demonstrate validity as measures of treatment response.

Keywords: major depressive disorder, Montgomery-Åsberg Depression Rating Scale, machine learning, computer vision, digital biomarker, antidepressant treatment, digital phenotyping

OPEN ACCESS

Edited by:
Mohamed Elgendi,
University of British Columbia, Canada

Reviewed by:
Ella Daly,
Janssen Pharmaceuticals, Inc., United States
Domenico De Berardis,
Azienda Usl Teramo, Italy
Neven Henigsberg,
University of Zagreb, Croatia

***Correspondence:**
Vidya Koesmahargyo
vidya.koesmahargyo@aicure.com

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

Received: *10 November 2020*
Accepted: *19 February 2021*
Published: *31 March 2021*

Citation:
Abbas A, Sauder C, Yadav V, Koesmahargyo V, Aghjayan A, Marecki S, Evans M and Galatzer-Levy IR (2021) Remote Digital Measurement of Facial and Vocal Markers of Major Depressive Disorder Severity and Treatment Response: A Pilot Study. Front. Digit. Health 3:610006. doi: 10.3389/fdgth.2021.610006

INTRODUCTION

Patients with major depressive disorder (MDD) are heterogeneous in both their clinical presentation and their response to antidepressant treatment (ADT) (1, 2). It is theorized that treatment effects may be obfuscated because MDD measurements combine heterogeneous symptoms that reflect distinct neurobiological and social processes while pharmacological treatments target specific neurobiological processes such as serotonergic tone. For example, patients with different subtypes of MDD, such as cognitive and neurovegetative phenotypes, have demonstrated differential treatment response to distinct classes of ADTs (3, 4). As such,

there are significant efforts to refocus treatment research on measures that match the underlying neurobiological treatment target (5). Disentangling the heterogeneity in MDD can lead to better risk and treatment response assessment by shifting the focus of investigation to narrow phenotypes that reflect the underlying neurological deficit and target of treatment (5, 6).

The use of digital measurements that relate to underlying biological phenotypes, termed digital phenotyping (7), has been proposed as a methodology to improve measurement of underlying illness by capturing digital proxy measures of clinical functioning. An example of digital phenotyping is the measurement of activity as a proxy measure of mood or anxiety states using actigraphy or geolocation captured from an individual's smartphone (8, 9). While novel measurements are promising, validation is required before such metrics can be interpreted clinically. The key steps to validation include comparison with traditional clinical measures, both cross-sectionally and as they change with the disease or treatment course (10). Such measures should strive for ease of collection and increased sensitivity to facilitate frequent, accurate assessment and should be validated in relation to narrower biological phenotypes and treatment targets than those that traditional endpoints assess. This will ultimately lead to improved, dynamic treatment research and clinical decision making (9) based on modulation of underlying neurobiological deficits (11).

Based on prior knowledge, visual, and auditory data sources represent a compelling direction for objective measurement of patient functioning in MDD. Beginning with observations by Emil Kraepelin, patients with depression have been shown to produce slowed and spaced out speech, where they appear to "become mute in the middle of a sentence" and demonstrate altered facial behavior, regarding which he states, "the facial expression and the general attitude are sleepy and languid" (12). These clinical observations by Kraepelin have been corroborated and extended with standardized methods to assess facial expressions, vocal characteristics, and movement patterns using audio and video data sources. The same paucity of speech has been observed in acutely suicidal patients (13). Indeed, both speech and facial/bodily movement represent sensitive biological outputs that change with physiological and cognitive variability (13–15).

A number of visual and auditory characteristics that correspond to known MDD symptoms can now be directly quantified. This includes reduced gross motor activity (16), slumped posture (17), reduced head movement variability (17–19), reduced facial expressivity (20), reduced speech production (21), and increased negative affect (22, 23). The automated measurement of these clinical features introduces the possibility of objective automated assessment. Given that audio and video data sources can be captured remotely, this further introduces the possibility of greatly scaling the reach and frequency of assessment. Increased scale and objectivity can facilitate increased accuracy and accessibility of clinical risk and treatment response assessments.

Serotonin signaling deficits represent a primary biological target for treatment in MDD. Serotonergic tone mechanistically impacts motor functioning directly through interactions with dopamine and norepinephrine signaling (24–26). Postmortem comparison of suicides compared with controls demonstrates significant reductions of brain serotonin (27, 28). More specific mapping of mRNA expression patterns demonstrates reduced expression of serotonin mRNA subtypes that are relatively widespread and other subtypes that are specific to the frontopolar cortex amygdala circuitry (29). This circuitry governs the expression and regulation of threat and anxiety (30).

In this exploratory pilot study, we tested the ability of digitally measured facial, vocal, and movement behaviors to measure depression severity and treatment response across 4 weeks of ADT in individuals with MDD. We hypothesized that overall facial expressivity, amount spoken, and head movement measured from video and audio captured during smartphone-based tasks would increase in response to ADT. We also hypothesized that negative facial affect (i.e., fear and anger) would decrease in response to treatment. In doing so, we aimed to evaluate the ability of remote, automated, digital assessments to measure depressive symptomatology with reliability and accuracy. We also hoped that findings from this pilot study would inform future studies with larger sample sizes that can delve further into how such measurements are affected in different MDD subpopulations and varying treatment regimens.

METHODS

Study Participants

Participants were identified through advertisements posted on social media. Individuals who self-identified as experiencing depression were screened over the telephone to assess depression symptoms. Potentially eligible subjects were then scheduled for an in-person pre-screening visit with a clinician to assess primary eligibility criteria. Individuals who met the criteria and provided informed consent participated in a screening assessment with a psychological rater, which included the Mini-International Neuropsychiatric Interview (MINI), Structured Interview Guide for the Montgomery–Åsberg Depression Rating Scale (SIGMA-MADRS), Columbia Suicide Severity Rating Scale (C-SSRS), and the Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR16). All study activities were approved by an institutional review board.

To be included in the study, subjects had to meet *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition (DSM-5) criteria for single or recurrent MDD based on the MINI with a current major depressive episode of $\geq$8 weeks and a MADRS total score of $\geq$20. Participants must have also been, in the opinion of the study psychiatrist, medically stable and a good candidate for treatment with a monoamine ADT. Key exclusion criteria included significant medical complications (e.g., uncontrolled cardiac or endocrine disorders, and diagnosis or treatment for cancer within the past 2 years), significant psychiatric complications (e.g., other primary psychiatric diagnoses and substance use disorders), intellectual disability (though no participants had to be excluded based on this criteria), or the use of certain prohibited concomitant medications (e.g., prescription painkillers/opioids; though use of benzodiazepines was not an

exclusion criterion, none of the study participants reported in this manuscript were on benzodiazepines). Subjects who endorsed active suicidal ideation with intent or recent suicidal behavior (within the past 6 months), or who, in the opinion of the investigator, were at significant risk for suicidal behavior were excluded.

Participants who met screening eligibility criteria subsequently completed a visit with a study psychiatrist and were prescribed an ADT consistent with standard of care. Participants who demonstrated significant decreases in depression severity, indicated by a 30% reduction in MADRS total score over 4 weeks of ADT, were included in the sample ($n = 18$). The sample included seven men and 11 women (mean age $= 30.2 \pm 8.6$). The mean body mass index (BMI) was 28.7 ± 5.6. Baseline total MADRS scores ranged from 25 to 45 (mean $= 34.1 \pm 4.9$). Five study participants (28%) were on ADT at the time of screening, and most (89%) had recurrent MDD. The mean major depressive episode duration was 11 months, ranging from 2 to 43 months.

Treatment and Assessment Conditions

All patients were prescribed either a selective serotonin reuptake inhibitor (SSRI) or serotonin–norepinephrine uptake inhibitor (SNRI) at label-specified doses based on the clinician's discretion. Time elapsed between the first participant in and last participant out was 6 months. Treatment response was measured at biweekly intervals using two independent assessments described below.

Assessments

Clinical Assessment

Montgomery–Åsberg Depression Rating Scale: The MADRS is a 10-item clinician administered scale for the measurement of MDD with validated clinical cut points for severe (>34), moderate (20–34), mild (7–19), and asymptomatic (<7) depression. The MADRS has demonstrated validity as a sensitive measure of ADT response (31). The MADRS was administered by trained psychological raters with prior scale experience at week 0 (baseline) and at ∼2 and 4 weeks posttreatment initiation.

Remote Smartphone-Based Video Assessments

All participants were asked to download the AiCure app (AiCure, LLC, New York, NY www.aicure.com) on their personal smartphone for measurement of digital markers of MDD. They were then trained by the study team on how to use the app to participate in remote assessments. This software platform has historically been used in clinical research for reporting of patient behavior to clinicians, including medication adherence, electronic patient-reported outcomes, and ecological momentary assessments, with considerable work done on patient acceptance and usability (32, 33). An additional functionality of capturing video and audio in response to prompts (as described below) was utilized for the purposes of this study (34, 35).

Participants completed weekly remote assessments for the length of the study. The assessment consisted of a smartphone-based adaptation of a paradigm to examine emotional valence in response to varied emotional imagery (27, 28, 36). At each assessment time point, they were prompted to view images taken from the Open Affective Standardized Image Set (OASIS) (37). The image set has emotional valence scores for each image based on responses recorded from a large, heterogeneous population, with lower scores referring to negatively valenced images and higher scores referring to positively valenced images. The valence scores were z-scored, and images with resulting scores of −0.5 to 0.5 standard deviation from the mean were considered neutrally valenced, images with resulting scores <1.5 standard deviation from the mean were considered negatively valenced, and images with resulting scores >1.5 standard deviation from the mean were considered positively valenced. The space in standard deviations between the classifications was added to ensure adequate separation between the image valences while also ensuring that enough images were left in each class to allow for there to be no repetition of images presented to the patients over the course of the study.

As part of the weekly remote assessments, patients were shown three positive images and three negative images padded with seven neutral images in between. The images were shown in series, starting with a neutral image, followed by a positive image, and then another neutral image before showing a negative image. This pattern was repeated until three positive and three negative images were shown and ended with a neutral image. This order was selected to avoid drastic shifts in image valences, i.e., switching directly between negative and positive images; by padding with neutral images, we hoped to alleviate any priming effects that may be present. For each image, the participant was asked to speak to the image by describing what they see in the picture and how it makes them feel (see **Figure 1**) and were required to speak for at least 10 s per image. Special care was also taken to ensure that participants were not shown the same image twice over the course of the study in order to limit any habituation effects of participating in the assessments.

Digital Marker Calculation

Video and audio were captured continuously during the smartphone assessment using the smartphone front-facing camera and microphone. Data were uploaded and processed through Health Insurance Portability and Accountability Act (HIPAA)-compliant backend services for transfer and storage of protected health information (PHI). Video was extracted for analysis for the portion of the task where the participant is observing the image and responding to it. Both video and audio were extracted and analyzed for the portion of the task when the participant was describing the image.

All analyses were conducted in python with use of open-source tools. All digital biomarker variables analyzed were acquired through the use of OpenDBM, an open-source software package that combines tools for measurement of facial, vocal, and movement behaviors, developed partially for the research presented in this manuscript (https://github.com/AiCure/open_dbm). Code for all subsequent statistical analyses presented in this manuscript has also been made available online: https://github.com/AiCure/ms_dbm_adamsclinicalstudy. A total of 17 digital measurements in addition to the MADRS scores were used to measure response

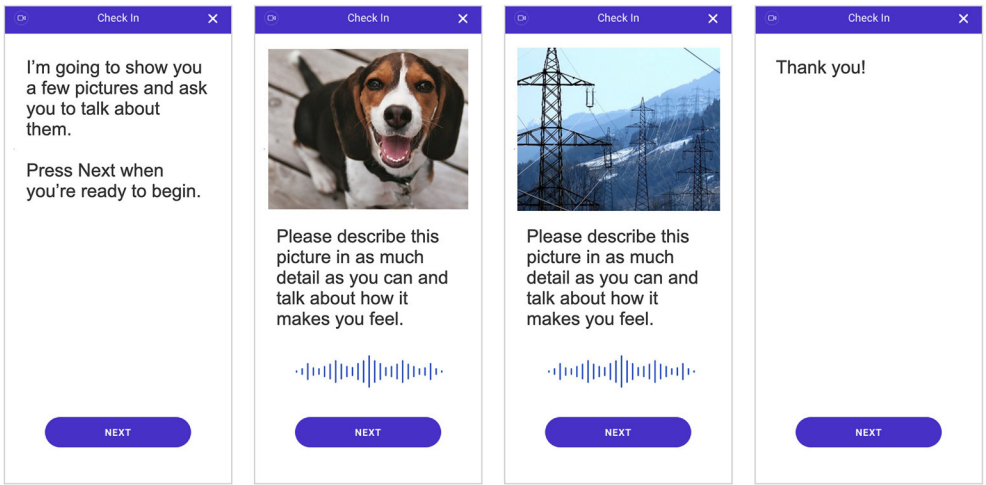


FIGURE 1 | Depiction of the smartphone-based assessment that all individuals completed. Video and audio of participant responses were recorded during the assessment and used to quantify behavioral characteristics and subsequently measure digital markers of major depressive disorder (MDD) severity.

TABLE 1 | Repeated-measures ANOVA results for all visual and voice markers measured in response to positive, negative, and neutral visual stimuli from baseline to 4 weeks of ADT.

	Neutral stimuli				Positive stimuli				Negative stimuli			
	Sph.	**F**	***p***	**+/–**	**Sph.**	**F/W**	***p***	**+/–**	**Sph.**	**F**	***p***	**+/–**
Voice percentage	**T**	**5.60**	**0.0095**	**+**	**T**	**3.59**	**0.0042**	**+**	**T**	**4.66**	**0.0187**	**+**
Anger intensity	T	13.28	<0.0001	+	T	21.19	<0.0001	+	T	19.96	<0.0001	+
Anger count	F	2.54	0.1214	n/a	F	0.92	0.3787	n/a	F	2.40	0.1307	n/a
Disgust intensity	T	12.00	0.0002	+	T	9.00	0.0009	+	T	9.43	0.0007	+
Disgust count	T	0.51	0.6033	n/a	T	1.33	0.2796	n/a	T	0.31	0.7358	n/a
Fear intensity	T	41.23	<0.0001	+	T	32.50	<0.0001	+	T	60.38	<0.0001	+
Fear count	F	4.84	0.0413	–	F	0.67	0.5182	n/a	F	0.77	0.4287	n/a
Happiness intensity	F	6.03	0.0232	+	F	5.21	0.0306	+	F	4.38	0.0445	+
Happiness count	T	0.03	0.9666	n/a	T	0.46	0.6362	n/a	F	1.72	0.2089	n/a
Sadness intensity	T	13.53	<0.0001	+	T	8.67	0.0012	+	T	10.54	0.0004	+
Sadness count	F	0.59	0.4690	n/a	T	2.05	0.1473	n/a	T	1.90	0.16815	n/a
Surprise intensity	T	22.29	<0.0001	+	T	17.31	<0.0001	+	T	26.10	<0.0001	+
Surprise count	T	0.14	0.8665	n/a	T	0.20	0.8194	n/a	T	0.16	0.8497	n/a
Overall expressivity	**T**	**32.60**	**<0.0001**	**+**	**T**	**40.67**	**<0.0001**	**+**	**T**	**36.95**	**<0.0001**	**+**
Head movement mean	**F**	**8.90**	**0.0069**	**+**	**F**	**3.58**	**0.0413**	**+**	**F**	**2.49**	**0.1335**	**n/a**
Head movement standard deviation	**F**	**3.68**	**0.0378**	**+**	**T**	**1.53**	**0.2333**	**n/a**	**F**	**1.59**	**0.2274**	**n/a**
Head pose change mean	**F**	**5.01**	**0.0325**	**+**	**T**	**3.18**	**0.0570**	**+**	**F**	**1.41**	**0.2595**	**n/a**

In particular, overall facial expressivity, voice percentage, and head movement markers showed an overall increase in response to treatment, consistent with the decrease in MDD symptom severity. Sph. indicates sphericity assumption being true or false, F indicates F-statistic, p indicates p-value of F-statistic, and +/– indicates increased or decreased values compared with baseline (n/a indicates no significant change was observed).
ADT, antidepressant treatment; MDD, major depressive disorder. Bold indicate the main findings.

to treatment. A subset of the results from those comparisons is presented in the main text (**Table 1**). A full list of comparisons is provided **Supplementary Table 1**. There was no primary endpoint that was being analyzed as part of this study; rather, the ability of a set of digital markers (facial, vocal, and movement) was being analyzed individually, with the collective comparisons indicating the usefulness of digital measurement tools in general.

Facial Marker Calculation

First, all videos were segmented into individual video frames at 30 frames per second. Next, each frame was segmented into three matrices consisting of red, blue, and green spectrum pixels for use in computer vision (CV) modeling using OpenCV, an open-source CV software package (38). Subsequently, each frame was analyzed using OpenFace (39), an open-source software package that has demonstrated validity next to expert human ratings

of Facial Action Coding System (FACS) (23), a standardized methodology to measure facial movements that reflect the activity in the underlying human facial musculature used in the production of basic emotions (i.e., *happiness*, *fear*, *anger*, *surprise*, *sadness*, and *disgust*).

Specifically, for each frame OpenFace outputs, (1) binary activation of each facial action unit (AU) was utilized to calculate the presence of facial emotions, and (2) the degree of expressivity for that AU was utilized to calculate intensity of facial emotions. From AU measurements, emotion behavior was calculated including (1) the presence or absence of each emotion for each frame selected as the most probable based on the observed AU activation, termed "count," and (2) the level of activation for each emotion and across all emotions, termed "intensity." Following the calculation of these variables for each frame, a set of variables was calculated that represented the count of emotions expressed across all frames divided by number of frames (*fear count*, *anger count*, *surprise count*, *sadness count*, *happy count*, and *disgust count*) and the intensity of emotion averaged over all frames (fear intensity, anger intensity, surprise intensity, sadness intensity, and disgust intensity). Additionally, a composite score of overall facial intensity summed across all emotions was calculated (*overall facial expressivity*).

Voice Marker Calculation

Recordings were segmented into speech and non-speech parts using parselmouth, an open-source software package that utilizes Praat software library (40) functions for vocal analysis (41). The ratio of speech to white space between words was calculated to represent the amount of time participants spoke compared with non-speech (*voice percentage*).

Movement Marker Calculation

For each frame of video, head position and angle were acquired using OpenFace. The average framewise displacement of the head between frames (*head movement mean*) and its standard deviation (*head movement standard deviation*) were calculated as measures of head movement. The mean change in angle of the head (*head pose change mean*) was calculated as an additional measure of head movement.

Data Analysis

Change over time in MADRS and facial, voice, and movement variables (termed digital markers) was calculated using repeated-measures analysis of variance (ANOVA). To avoid capitalizing on change when doing multiple comparisons or testing for multiple hypotheses, *p*-values were corrected using false discovery rate (FDR) correction (42). The sphericity assumption, which is the condition where the variances of the differences between all combinations of related groups are equal, was formally tested for each ANOVA. When this assumption proved to hold, the F-statistic and corresponding *p*-value were used. When the sphericity assumption was violated, Mauchly's W statistic and corresponding *p*-value were used (43). Additionally, pairwise comparisons were calculated between each time point to determine where change across time points occurs (i.e., baseline to 2 weeks, baseline to 4 weeks, and 2–4 weeks) controlling for FDR using Tukey's test.

RESULTS

Depression Response

Participants demonstrated a main effect for change in MADRS scores from baseline to week 4 [$F_{(2, 34)} = 51.62$, $p < 0.0001$]. Descriptive statistics demonstrate clinically relevant change with patients moving from the clinical to non-clinical range (**Supplementary Table 1**; **Figure 2**).

Participants demonstrated change in MDD severity as measured by digital markers. To align time points between digital markers and the MADRS scores, measurements from days 7 to 21 were averaged as the week 2 time point, and measurements from days 22to 35 were averaged as the week 4 time point. Due to missed remote assessments, a subset of the total sample of 18 had complete data across time points, with $n = 12$ for facial markers and $n = 11$ for voice markers. All statistical results for digital markers are presented in **Table 1**. Examples of marker profiles across treatment are presented in **Figure 2** alongside the participants' MADRS profile across treatment. All scores, including MADRS, were normalized to a range of 0–1 to allow visual comparison of the magnitude of change on digital markers in comparison with change in MADRS clinical scores (**Figure 2**).

Facial Markers

All facial activity measures across all emotions (fear intensity, anger intensity, surprise intensity, sadness intensity, disgust intensity, and overall expressivity) along with the overall expressivity score demonstrated significant positive change from baseline to week 4 in response to all image prompts (positive, neutral, and negative; see **Table 1**). This result indicates that ADT produces a main effect on facial activity overall, which is not bound to one particular facial musculature group or type of external stimulus (**Figure 2**).

Across conditions, the frequency of expressions of anger (anger count) decreases. The frequency of expressions of fear also decreases, but only in response to neutral and negative stimuli (fear count). Additionally, the frequency of expressions of happiness (happy count) decreases in response to negative stimuli only. Together, results indicate a general decrease in expressions of anger and context-specific decreases in fear and happiness expressions.

Voice Markers

The single variable representing the ratio of speech to silence across sentences uttered (voice percentage) additionally demonstrated significant positive change in response to ADT across all conditions, indicating an increase in speech relative to silence. This result is consistent with increased motor/muscle activity observed in facial activity (**Figure 2**).

Movement Markers

Additionally, movement parameters demonstrated consistent effects across conditions. The rate of head movement (head movement mean) and the degree of variability in the rate

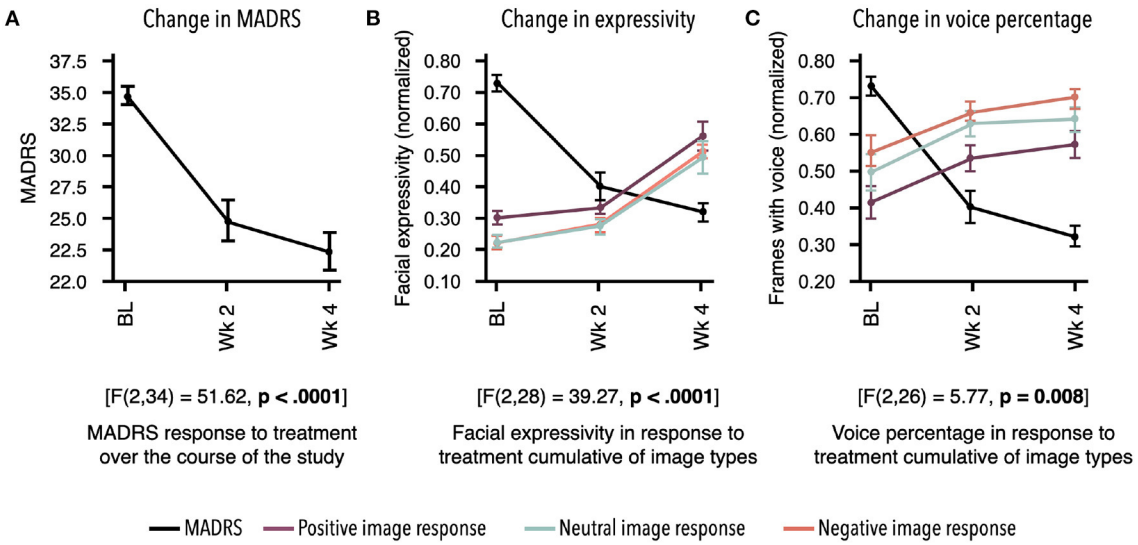


FIGURE 2 | Response to treatment as measured by two independent assessments. Mean for each time point and standard error bars are shown. Results demonstrate treatment's significant effect on digital markers, which are highly concordant with change in depression symptom severity. **(A)** Montgomery–Åsberg Depression Rating Scale (MADRS) scores acquired at baseline (BL), week 2, and week 4 showed significant decrease in response to antidepressant therapy (ADT) [$F_{(2,\ 34)} = 51.62$, $p < 0.0001$]. **(B)** Overall facial expressivity measured in response to positive [$F_{(2,\ 28)} = 40.66$; $p < 0.0001$], neutral [$F_{(2,\ 28)} = 32.6$; $p < 0.0001$], and negative [$F_{(2,\ 28)} = 36.95$; $p < 0.0001$] images demonstrated a significant increase in response to ADT as MADRS scores decreased. **(C)** Percentage of frames with voice measured in response to positive [$F_{(2,\ 26)} = 3.59$; $p = 0.04$], neutral [$F_{(2,\ 26)} = 5.59$; p 0.009], and negative [$F_{(2,\ 28)} = 4.65$; $p = 0.02$] images also demonstrated a significant increase in response to ADT as MADRS scores decreased. All values in **(B,C)** were normalized between 0 and 1.

of head movement (head movement standard deviation) both demonstrated significant increases in response to ADT. Head pose change mean also demonstrated significant increase during neutral and positive stimuli (see **Table 1**).

DISCUSSION

Results demonstrate a consistent effect of monoamine ADTs (SSRIs/SNRIs) on digital markers of motor functioning, which are highly concordant with change in MDD symptom severity. Specifically, facial and vocal activities demonstrated robust increases across 4 weeks following the initiation of treatment, which mirrored decreases in symptom severity as assessed by the clinician administered MADRS. The current findings suggest that SSRI/SNRI treatment, which produces graded increases in serotonin, reduces depression severity in part by rescuing motor functioning (e.g., increased facial expressivity and increased speech production).

Additionally, a decrease was observed across conditions in the expression of anger. Patients with depression have long demonstrated increased rates of anger than healthy counterparts (44, 45). Furthermore, polymorphisms of the serotonin 1B receptor that are associated with increased depression and suicide risk are also associated with increased anger and fear (46, 47). These results further indicate that the observed change in digital markers in response to serotonin reuptake inhibitors reflects a more specific phenotypic change in measurement of serotonergic profile in the central nervous system.

Serotonin levels in the central nervous system are known to have both direct and indirect effects (*via* dopamine) on motor activity (48, 49). Both suicide (as measured in postmortem brain tissue) and suicidal attempts, a key symptom class of MDD, are associated with depleted serotonin (50). As such, digital measurements that reflect motor behavior may represent a sensitive measure of serotonergic tone and potentially other neurotransmitter activities that affect motor functioning and ultimately the overall clinical presentation.

The current work presents a number of limitations that should be overcome through research that confirms and extends the findings reported. First, while treatment success was confirmed with clinical measures of MDD, dosage and treatment type were not controlled in a manner to make direct inferences about dose–response relationships. Future studies with larger sample sizes that consider different treatment types will have to be conducted to make comparisons on how they might affect digital measurements in varying ways. In addition, the current study was not adequately powered to assess the intra-subject variability in treatment response. Future research should provide more extensive experimental control of medication and dosage to assess the relationship between magnitude of clinical response and digital markers of motor activity.

Second, while facial movement results were robust, we do not know if findings related to specific emotions would rise to significance given a larger sample size or more sampling occasions of the stimuli. One of the goals of the data collection was to implement a very simple remote assessment of objective visual and auditory markers to facilitate ease of frequent assessment. However, the minimum sample to

accurately measure each marker needs to be assessed through the use of larger samples. For example, we observed decreases in happiness in response to negative images. This result is difficult to directly interpret. However, given a larger sample, we may be powered to identify increases in happiness in response to positive images, consistent with observations that depressed patients display context-inappropriate affect (51, 52). It is also possible that priming to the stimuli, i.e., the images shown during the remote assessments, was a factor in the behavior recorded and subsequently the data analyzed; both priming to the stimuli and habituation to the assessment need to be evaluated in future work.

Third, subtypes of depression and the range of depression severity observed at the start of treatment were not evaluated as variables in the analysis presented in this study due to the small sample size used. However, the findings observed in this study will greatly inform future work to determine sample sizes needed to measure how digital measurements of facial, vocal, and movement behaviors may differ across subpopulations of MDD as well as different treatment types.

Ultimately, the current work holds promise as an example of the potential to observe treatment effects that reflect underlying neurobiological target engagement by shifting the focus to monotonic neurobiologically based domains rather than heterogeneous diagnoses (6). Further work should determine if these same markers are relevant in other disorders and treatments that are mechanistically affected by serotonergic tone, as well as their relevance to other disorders with motor and movement profiles including Parkinson's disease and schizophrenia (53). Second, the current work demonstrates the success of non-invasive objective digital assessment as a tool to assess treatment effects in MDD, which was the core focus of the study. Importantly, no markers were scientifically novel; rather, they were based on validated methods that are open and public and have been previously reported in scientific literature.

The current work demonstrates, in the context of MDD, that these data sources can be captured remotely through ubiquitously available digital tools to provide measurements that are at least as robust as traditional rating scales. It will be important to determine if such models reliably track with other disease states and treatment responses, as such models and applications have significant potential to increase the rate and accuracy of treatment decision making.

Together, the current study demonstrates that scalability, through digital measurement, of monotonic characteristics that reflects the underlying central nervous system activity. This observation holds promise that frequent remote digital assessment can be used to monitor, titrate, and even personalize treatment for MDD and other psychiatric or neurological conditions by grounding the measurements in narrow phenotypes that match the underlying mechanistic target of the treatment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://github.com/AiCure/ms_dbm_adamsclinicalstudy.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Adams Clinical Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AA, VY, VK, and IG-L contributed toward the conception of study, development of technology used, data analysis conducted, reporting of results, and writing the manuscript. AA, SM, ME, and CS carried out all participant recruitment, technology onboarding, clinical data collection, and manuscript revisions. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.610006/full#supplementary-material

Supplementary Table 1 | Means and standard errors across time points of each digital marker in response to neutral, positive and negative stimuli.

Supplementary Table 2 | Pairwise comparisons between time points for each digital marker in response to neutral, positive and negative stimuli.

Supplementary Figure 1 | Clockwise from top-left, weekly change in digital measurement of overall expressivity, voice percentage, head pose change, and head movement. Each of the variables have been split up by the kind of image that the participants were speaking to: negatively, neutrally, or positively valenced images. In the comparison presented in the main text, values for week 2 and 3 and values for weeks 4 and 5 were averaged into single time points to align the digital measurement time points with the MADRS time points for side-by-side comparison. It also increased the sample size, as not all participants provided consistent weekly data and aggregation across weeks increased the n that could be included in the repeated measures ANOVA. These figures demonstrate the weekly change, further emphasizing the point made in the main text that digital measurements can be conducted with greater frequency than traditional assessments such as the MADRS. However, in these figures, the same patients do not represent each time point. This explains the dip in values is observed at time point 5, which is biased toward the subset of patients that had week 5 data rather than being indicative of a consistent trend.

REFERENCES

1. Lux V, Kendler KS. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. *Psychol Med.* (2010) 40:1679–90. doi: 10.1017/S0033291709992157
2. Ostergaard SD, Jensen SOW, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatr Scand.* (2011) 124:495–6. doi: 10.1111/j.1600-0447.2011.01744.x
3. Uher R, Maier W, Hauser J, Marusic A, Schmael C, Mors O, et al. Differential efficacy of escitalopram and nortriptyline on dimensional

measures of depression. *Br J Psychiatry J Ment Sci.* (2009) 194:252–9. doi: 10.1192/bjp.bp.108.057554
4. Jaracz J, Gattner K, Moczko J, Hauser J. Comparison of the effects of escitalopram and nortriptyline on painful symptoms in patients with major depression. *Gen Hosp Psychiatry.* (2015) 37:36–9. doi: 10.1016/j.genhosppsych.2014.10.005
5. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* (2013) 11:126. doi: 10.1186/1741-7015-11-126
6. Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry.* (2014) 171:395–7. doi: 10.1176/appi.ajp.2014.14020138
7. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA.* (2017) 318:1215. doi: 10.1001/jama.2017.11295
8. Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology.* (2016) 41:1691–6. doi: 10.1038/npp.2016.7
9. Torous J, Onnela J-P, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping *via* smartphones and connected devices. *Transl Psychiatry.* (2017) 7:e1053. doi: 10.1038/tp.2017.25
10. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *Npj Digit Med.* (2019) 2:1–5. doi: 10.1038/s41746-019-0090-4
11. Lenze EJ, Rodebaugh TL, Nicol GE. A framework for advancing precision medicine in clinical trials for mental disorders. *JAMA Psychiatry.* (2020) 77:663–4. doi: 10.1001/jamapsychiatry.2020.0114
12. Kraepelin E. *Clinical Psychiatry*. New York, NY: Macmillan (1907).
13. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* (2015) 71:10–49. doi: 10.1016/j.specom.2015.03.004
14. Jabbi M, Keysers C. Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emot Wash DC.* (2008) 8:775–80. doi: 10.1037/a0014194
15. Dagum P. Digital biomarkers of cognitive function. *Npj Digit Med.* (2018) 1:1–3. doi: 10.1038/s41746-018-0018-4
16. Buyukdura JS, McClintock SM, Croarkin PE. Psychomotor retardation in depression: biological underpinnings, measurement, and treatment. *Prog Neuropsychopharmacol Biol Psychiatry.* (2011) 35:395–409. doi: 10.1016/j.pnpbp.2010.10.019
17. Kim Y, Cheon S-M, Youm C, Son M, Kim JW. Depression and posture in patients with Parkinson's disease. *Gait Posture.* (2018) 61:81–5. doi: 10.1016/j.gaitpost.2017.12.026
18. Alghowinem S, Goecke R, Wagner M, Parkerx G, Breakspear M. Head pose and movement analysis as an indicator of depression. In: *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* ACII '13 (Geneva: USA: IEEE Computer Society) (2013). p. 283–8. doi: 10.1109/ACII.2013.53
19. Dibeklioglu H, Hammal Z, Cohn JF. Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE J Biomed Health Inform.* (2018) 22:525–36. doi: 10.1109/JBHI.2017.2676878
20. Girard JM, Cohn JF, Mahoor MH, Mavadati S, Rosenwald DP. Social risk and depression: evidence from manual and automatic facial expression analysis. In: *Proc Int Conf Autom Face Gesture Recognit IEEE Int Conf Autom Face Gesture Recognit.* (2013). p. 1–8. doi: 10.1109/FG.2013.6553748
21. Garcia-Toro M, Talavera JA, Saiz-Ruiz J, Gonzalez A. Prosody impairment in depression measured through acoustic analysis. *J Nerv Ment Dis.* (2000) 188:824–9. doi: 10.1097/00005053-200012000-00006
22. Berenbaum H, Oltmanns TF. Emotional experience and expression in schizophrenia and depression. *J Abnorm Psychol.* (1992) 101:37–44. doi: 10.1037//0021-843x.101.1.37
23. Ekman P, Rosenberg EL. *What the Face RevealsBasic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York, NY: Oxford University Press (2005). doi: 10.1093/acprof:oso/9780195179644.001.0001
24. Morrissette DA, Stahl SM. Modulating the serotonin system in the treatment of major depressive disorder. *CNS Spectr.* (2014) 19(Suppl.1):57–67. doi: 10.1017/S1092852914000613
25. Herrera-Guzmán I, Gudayol-Ferré E, Herrera-Guzmán D, Guàrdia-Olmos J, Hinojosa-Calvo E, Herrera-Abarca JE. Effects of selective serotonin reuptake and dual serotonergic–noradrenergic reuptake treatments on memory and mental processing speed in patients with major depressive disorder. *J Psychiatr Res.* (2009) 43:855–63. doi: 10.1016/j.jpsychires.2008.10.015
26. El Mansari M, Guiard BP, Chernoloz O, Ghanbari R, Katz N, Blier P. Relevance of norepinephrine–dopamine interactions in the treatment of major depressive disorder. *CNS Neurosci Ther.* (2010) 16:e1–17. doi: 10.1111/j.1755-5949.2010.00146.x
27. Stockmeier CA. Involvement of serotonin in depression: evidence from postmortem and imaging studies of serotonin receptors and the serotonin transporter. *J Psychiatr Res.* (2003) 37:357–73. doi: 10.1016/s0022-3956(03)00050-5
28. Stockmeier CA, Shapiro LA, Dilley GE, Kolli TN, Friedman L, Rajkowska G. Increase in serotonin-1A autoreceptors in the midbrain of suicide victims with major depression—postmortem evidence for decreased serotonin activity. *J Neurosci.* (1998) 18:7394–401. doi: 10.1523/JNEUROSCI.18-18-07394.1998
29. Anisman H, Du L, Palkovits M, Faludi G, Kovacs GG, Szontagh-Kishazi P, et al. Serotonin receptor subtype and p11 mRNA expression in stress-relevant brain regions of suicide and control subjects. *J Psychiatry Neurosci JPN.* (2008) 33:131–41.
30. Gold AL, Morey RA, McCarthy G. Amygdala-prefrontal cortex functional connectivity during threat-induced anxiety and goal distraction. *Biol Psychiatry.* (2015) 77:394–403. doi: 10.1016/j.biopsych.2014.03.030
31. Montgomery SA, Smeyatsky N, de Ruiter M, Montgomery DB. Profiles of antidepressant activity with the Montgomery-Asberg Depression Rating Scale. *Acta Psychiatr Scand Suppl.* (1985) 320:38–42. doi: 10.1111/j.1600-0447.1985.tb08073.x
32. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. *Stroke.* (2017) 48:1416–9. doi: 10.1161/STROKEAHA.116.016281
33. Hanina A, Shafner L. Traveling through the storm: leveraging virtual patient monitoring and artificial intelligence to observe, predict, and affect patient behavior in CNS drug development. *Handb Behav Neurosci.* (2019) 29:427–34. doi: 10.1016/B978-0-12-803161-2.00031-X
34. Abbas A, Yadav V, Smith E, Ramjas E, Rutter SB, Benavidez C, et al. Computer vision-based assessment of motor functioning in schizophrenia: use of smartphones for remote measurement of schizophrenia symptomatology. *Digit Biomark.* (2021) 5:29–36. doi: 10.1159/000512383
35. Galatzer-Levy IR, Abbas A, Koesmahargyo V, Yadav V, Perez-Rodriguez MM, Rosenfield P, et al. Facial and vocal markers of schizophrenia measured using remote smartphone assessments. *medRxiv*. (2020) 2020:9741. doi: 10.1101/2020.12.02.20219741
36. Carretié L, Tapia M, López-Martín S, Albert J. EmoMadrid: an emotional pictures database for affect research. *Motiv Emot.* (2019) 43:929–39. doi: 10.1007/s11031-019-09780-y
37. Kurdi B, Lozano S, Banaji MR. Introducing the open affective standardized image set (OASIS). *Behav Res Methods.* (2017) 49:457–70. doi: 10.3758/s13428-016-0715-3
38. Brahmbhatt S. Introduction to computer vision and OpenCV. In: *Practical OpenCV* (Berkeley, CA: Apress). p. 3–5. doi: 10.1007/978-1-4302-6080-6_1
39. Baltrušaitis T, Robinson P, Morency L-P. OpenFace: an open source facial behavior analysis toolkit. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV).* (2016). p. 1–10. doi: 10.1109/WACV.2016.7477553
40. Weenink D, Boersma P. *Praat: Doing Phonetics by Computer*. (2018). Available online at: http://www.praat.org/ (accessed June 04, 2020).

41. Jadoul Y, Thompson B, de Boer B. Introducing parselmouth: a python interface to Praat. *J Phon.* (2018) 71:1–15. doi: 10.1016/j.wocn.2018.07.001
42. Li A, Barber RF. Multiple testing with the structure-adaptive Benjamini-Hochberg algorithm. *J R Stat Soc Ser B Stat Methodol.* (2019) 81:45–74. doi: 10.1111/rssb.12298
43. Mendoza JL. A significance test for multisample sphericity. *Psychometrika.* (1980) 45:495–8. doi: 10.1007/BF02293611
44. Koh KB, Kim DK, Kim SY, Park JK. The relation between anger expression, depression, and somatic symptoms in depressive disorders and somatoform disorders. *J Clin Psychiatry.* (2005) 66:485–91. doi: 10.4088/jcp.v66n0411
45. Riley WT, Treiber FA, Woods MG. Anger and hostility in depression. *J Nerv Ment Dis.* (1989) 177:668–74. doi: 10.1097/00005053-198911000-00002
46. Hakulinen C, Jokela M, Hintsanen M, Merjonen P, Pulkki-Råback L, Seppälä I, et al. Serotonin receptor 1B genotype and hostility, anger and aggressive behavior through the lifespan: the Young Finns study. *J Behav Med.* (2013) 36:583–90. doi: 10.1007/s10865-012-9452-y
47. McDevitt RA, Hiroi R, Mackenzie SM, Robin NC, Cohn A, Kim JJ, et al. Serotonin 1B autoreceptors originating in the caudal dorsal raphe nucleus reduce expression of fear and depression-like behavior. *Biol Psychiatry.* (2011) 69:780–7. doi: 10.1016/j.biopsych.2010.12.029
48. Gerson SC, Baldessarini RJ. Motor effects of serotonin in the central nervous system. *Life Sci.* (1980) 27:1435–51. doi: 10.1016/0024-3205(80)90368-9
49. Jacobs BL. Serotonin and behavior: emphasis on motor control. *J Clin Psychiatry.* (1991) 52(Suppl.):17–23.
50. Mann JJ. The serotonergic system in mood disorders and suicidal behaviour. *Philos Trans R Soc Lond B Biol Sci.* (2013) 368:20120537. doi: 10.1098/rstb.2012.0537
51. Coifman KG, Bonanno GA, Ray RD, Gross JJ. Does repressive coping promote resilience? Affective-autonomic response discrepancy during bereavement. *J Pers Soc Psychol.* (2007) 92:745–58. doi: 10.1037/0022-3514.92.4.745
52. Coifman KG, Bonanno GA. When distress does not become depression: emotion context sensitivity and adjustment to bereavement. *J Abnorm Psychol.* (2010) 119:479–90. doi: 10.1037/a0020113
53. Stahl SM. Parkinson's disease psychosis as a serotonin-dopamine imbalance syndrome. *CNS Spectr.* (2016) 21:355–9. doi: 10.1017/S1092852916000602

Conflict of Interest: At the time of the study, AAb, VY, VK and IG-L were employees of AiCure and held stock options in AiCure. CS, AAg, SM, and ME were employees of Adams Clinical, and CS was also an employee of Karuna Therapeutics.

The authors declare that the study was jointly funded by AiCure, LLC and Adams Clinical, both of which may benefit from research reported in the manuscript and were involved in study design and execution. Adams Clinical conducted patient recruitment, enrollment, and clinical assessments. AiCure developed methods for digital phenotyping and provided software tools for remote collection of video data used. Both AiCure and Adams Clinical were involved in subsequent data analysis, interpretation, and presentation of findings.

frontiers in Digital Health

REVIEW
published: 07 April 2021
doi: 10.3389/fdgth.2021.662811

Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring

*Mahsa Sheikh, M. Qassem and Panicos A. Kyriacou**

Research Centre for Biomedical Engineering, School of Mathematics, Computer Science & Engineering, City, University of London, London, United Kingdom

Collecting and analyzing data from sensors embedded in the context of daily life has been widely employed for the monitoring of mental health. Variations in parameters such as movement, sleep duration, heart rate, electrocardiogram, skin temperature, etc., are often associated with psychiatric disorders. Namely, accelerometer data, microphone, and call logs can be utilized to identify voice features and social activities indicative of depressive symptoms, and physiological factors such as heart rate and skin conductance can be used to detect stress and anxiety disorders. Therefore, a wide range of devices comprising a variety of sensors have been developed to capture these physiological and behavioral data and translate them into phenotypes and states related to mental health. Such systems aim to identify behaviors that are the consequence of an underlying physiological alteration, and hence, the raw sensor data are captured and converted into features that are used to define behavioral markers, often through machine learning. However, due to the complexity of passive data, these relationships are not simple and need to be well-established. Furthermore, parameters such as intrapersonal and interpersonal differences need to be considered when interpreting the data. Altogether, combining practical mobile and wearable systems with the right data analysis algorithms can provide a useful tool for the monitoring and management of mental disorders. The current review aims to comprehensively present and critically discuss all available smartphone-based, wearable, and environmental sensors for detecting such parameters in relation to the treatment and/or management of the most common mental health conditions.

Keywords: mental health monitoring, wearables, personal sensing, physiological and behavioral monitoring, digital phenotyping

OPEN ACCESS

Edited by:
Pradeep Nair,
Central University of Himachal Pradesh, India

Reviewed by:
Carlo Massaroni,
Campus Bio-Medico University, Italy
Ian Cleland,
Ulster University, United Kingdom

***Correspondence:**
Panicos A. Kyriacou
p.kyriacou@city.ac.uk

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

Received: *01 February 2021*
Accepted: *02 March 2021*
Published: *07 April 2021*

Citation:
Sheikh M, Qassem M and Kyriacou PA (2021) Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring. Front. Digit. Health 3:662811. doi: 10.3389/fdgth.2021.662811

INTRODUCTION

Mental illness is a health condition that alters a person's thoughts, feelings, and/or behaviors and causes the person distress and difficulty in functioning. The aggregate lifetime prevalence of common mental disorders across 39 countries has been estimated at around 30% (1). Mental health disorders are a major contributor to global disease burden due to their high prevalence, impairment of critical brain functions, and clinical course that is either chronic or remitting and relapsing. Some of the general symptoms and warning signs for mental disorders include marked personality change, inability to cope with problems and daily activities, strange and grandiose ideas, extreme

mood swings, excessive anxieties, violent behavior, and thinking or talking about suicide or harming oneself (2).

Conventionally, mental disorders are diagnosed by self-report screening questionnaires or based on the *Diagnostic and Statistical Manual of Mental Disorders V*. Therefore, clinical diagnosis of mental disorders is achieved through the patient's subjective description of the symptoms, interviews and psychological questionnaires, and the physician's own expertise. However, the conventional assessment of psychiatric disorders based on a patient's or informant's recall is subject to inherent biases, and unreliability as the key feature in mental disorders is the variation of mood over time. In addition, mental health disorders are often chronic and relapsing in nature, and thus, long-term treatment management and assessment are essential for patient symptom reduction and recovery, but this is difficult to achieve with traditional methods, which rely on retrospective reports that are subject to recall bias. The lack of accurate methods for characterizing behavior and the need for regular monitoring of mental health conditions poses several physical and economical challenges, which, in recent years, has prompted great interest in digital phenotyping of mental health disorders through monitoring of physiological and behavioral parameters that can be translated into biomarkers of conditions, such as depression, anxiety disorders, bipolar disorder, and schizophrenia.

Physiological measures, such as heart rate variability (HRV), skin temperature (ST), electromyography (EMG), blood volume pulse (BVP), blood pressure, and cortisol levels as well as behavioral data, including sleep duration, social activities, and voice features, are among the many factors influenced by many mental health disorders. Consequently, capturing physiological and behavioral data has been widely implemented for the monitoring and general management of mental health.

In general, context sensing, personal sensing, mobile sensing, and digital phenotyping are all terms referring to identifying behaviors that relate to physical and mental health (2). Personal sensing involves collecting and analyzing data from sensors embedded in the context of daily life to detect and measure physical properties. To translate the raw sensor data into markers of behaviors and states related to mental health, the captured data is converted into features that are used to define behavioral markers through machine learning (3). Smartphones and connected devices, such as smartwatches, can be used to monitor behavior through passive detection of location, acceleration, social activities, and voice features.

Passive collection of physiological and behavioral data has been implemented for the diagnosis of various mental health disorders. This includes a wide range of studies that have investigated smartphone-based, wearable, and environmental sensors for digital phenotyping and monitoring of mental health. Until recent years, direct measurement of physiological data (e.g., HR, sleep quality, skin conductance) has been impractical as such devices have traditionally been bulky, large, and expensive (4). However, the development of various wearable devices, such as wristbands, smartwatches, and fitness tracking devices, and the common use of smartphones have facilitated the collection of biologically significant data. As a result, new technologies are rapidly being developed in fostering mental health and will allow for collection and analysis of data derived from self-reports, monitoring of behavioral patterns, and physiological sensing. More importantly, patients with serious mental illnesses have mostly shown good adherence and interest in using such devices to collect physiological and behavioral data (4), which would provide a vital complement to clinical visits.

The enthusiastic consumer and patient use of mobile devices, the availability of high-speed wireless internet, and a high prevalence of connected mobile devices with operating systems have culminated in patient and provider interest in monitoring mental symptoms and capturing behavioral and physiological data. This paper provides a review on recent developments in digital phenotyping based on physiological and behavioral monitoring of symptoms related to mental health and discusses the impact of such technologies in mental health care as, despite the expanded interest in mobile and connected technology in the arena of mental health, very few e-tools have been well-studied and validated for their impact when implemented on a broad scale (4).

METHODS

Searches of the literature were conducted in Web of Science, PubMed, and Ovid [including Journals from Ovid, CityLibrary Journals@Ovid, AMED (Allied and Complementary Medicine), Embase, Global health, and Ovid MEDLINE]. Keywords used in this search included "sensors," "mental health monitoring," "personal sensing," "mental disorders," "physiological and behavioral monitoring," and "digital phenotyping." Database searches yielded 851 results of which 21 were review papers. The references of relevant review papers were scanned to identify applicable studies. From the combinations of the keywords and 36 relevant articles found in review references, 866 articles were identified. Studies investigating physiological and behavioral monitoring in any condition other than mental disorders were excluded. Studies in which no sensing device was employed for monitoring physiological and behavioral parameters were also excluded. From careful analysis of titles and abstracts, 139 articles were identified, 73 of which met the inclusion criteria and were included for analysis (**Figure 1**).

The literature search yielded 73 studies, 55 of which investigated wearable, mobile phone, and ambient sensors, summarized in **Tables 1**–**3**. The remaining 18 papers focused on identifying behavioral and physiological changes linked to mental health and used existing monitoring systems without considering and investigating a specific sensing mode. These studies focused on identifying mental health–related physiological and behavioral changes, some of which are specifically discussed in section Mental Health–Related Physiological and Behavioral Changes. The years of publications analyzing different sensing devices for mental health monitoring ranged from 2008 to 2020 with the years 2015 to 2020 producing the majority of publications (**Figure 2**). Wearables and sensor-enabled smartphone applications constituted the majority of studies with 24

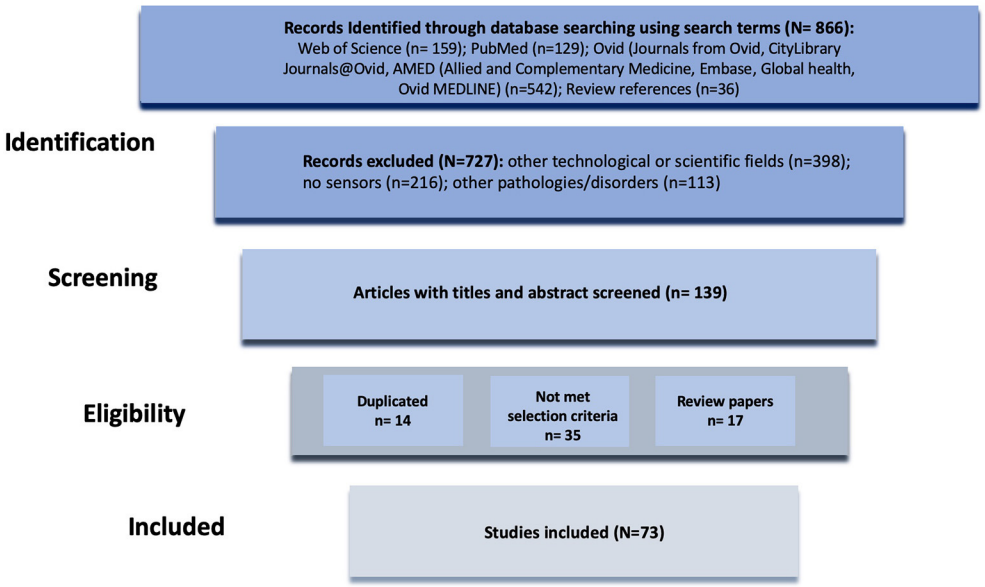


FIGURE 1 | Diagram of the methodology used for the literature review process.

and 22 publications, respectively. A total number of nine studies were found to investigate ambient sensors in mental health monitoring.

In sorting the included studies by application types, it is demonstrated that the majority of publications considered in this review investigated the use of monitoring systems for stress along with anxiety disorders and depression with just 2% difference between the two. Bipolar disorder monitoring, accounting for 15% of the studies, is the next well-studied condition for employment of behavioral and physiological monitoring systems. Schizophrenia and posttraumatic stress disorder (PTSD) monitoring comprise the smallest number of publications (**Figure 3A**). The application of different devices, including wearable, mobile phone, and ambient sensors, was also analyzed for each mental health condition. In the management of stress and anxiety disorders, wearable sensors were the most employed monitoring systems, mostly sensing physiological parameters, such as HR, electrodermal activity, and ST to evaluate anxiety levels. Whereas, for depression and bipolar disorder most of the studies used smartphone sensors as the monitoring system. In the monitoring of depression, the majority of studies focused on behavioral parameters, including physical activity, sleep duration, and conversation frequency to identify depressed mood. Furthermore, to assess the pathological mood states involved in bipolar disorder, physiological signals, such as HR and respiration activity, as well as behavioral parameters, such as mobility, social activity, sleep patterns, and voice features, were monitored. Although ambient sensors are generally the least investigated devices for mental health monitoring, they have been widely considered for detecting cognitive impairment (**Figure 3B**).

MENTAL HEALTH–RELATED PHYSIOLOGICAL AND BEHAVIORAL CHANGES

A wide range of studies have demonstrated the association between physiological and behavioral patterns and different mental disorders. For example, it has been suggested that demonstrations of more phone activity from a combination of increased GPS position changes, erratic accelerometer movements, and increased social activity may be indicative of the manic phase in bipolar patients (55). Moreover, analysis of voice data through machine learning and natural language processing and detection of the pitch, tempo, and loudness of voice might serve as a potential marker of many psychiatric illnesses, such as depression, anxiety, and even suicidality (56). Portable noise sensors and GPS trackers have been used to identify links between mental health and personal noise exposure with high exposure to noise at the entire day level associated with worse mental health (57). Moreover, HRV, respiratory sinus arrythmia (RSA), electrodermal activity (EDA), ST, EMG, BVP, blood pressure, and cortisol levels are among the physiological markers that react to emotional experiences and can be used to detect stress and other emotional states (58). Wearables can also measure variables such as skin conductance and HR with greater asymmetries in skin conductance amplitude on the two sides of the body identified as an indicator of emotional arousal (3). Therefore, sensor data can be used to monitor human emotion states. However, the signals identified as physiological responses during various emotional states should be stable and reliable (59). Data from accelerometer-based wearable devices have also been used to detect the association between physical activity and depression (60). Although decreased levels of mobility and social

TABLE 1 | Wearable physiological and emotional monitoring research prototypes.

Reference/product	Purpose	Device form factor	Sensors/parameters	Clinical application
Wearables				
Jiang et al. (5), Newcastle University, UK, 2019	Audio sensing, motion detection, behavior monitoring	Watch	Light and temperature sensors, microphone, accelerometers	Anxiety, autism
Guo et al. (6), Purdue University, USA, 2016	Emotion recognition	Scarf	Heart rate sensor, Electrodermal Activity (EDA) sensor	Emotion regulation
Hui et al. (7), The University of Reading, 2018	Emotion recognition	Glove	PPG, electrodermal activity, skin temperature, and electromyography (EMG) sensors	Emotion regulation
MyFeel (8), Sentio Solutions Inc	Anxiety detection	wristband	Heart rate, electrodermal activity, and skin temperature	Emotion recognition and stress management
Reveal (9), Awake labs	Anxiety detection	Smartwatch	Heart rate, electrodermal activity, and skin temperature	Emotion recognition and stress management
Thync (10), Thync Global Inc	Increase energy and lower stress	Patch	Electrical Nerve Stimulation (TENS) and Transcranial Direct Current Stimulation (tDCS)	Neurostimulation
AutoSense (11), Washington, USA, 2011	Stress detection	Sensor Suite	Electrocardiogram (ECG) measurement, respiratory inductive plethysmograph (RIP), Galvanic skin response (GSR), skin and ambient temperature sensors, three-axis accelerometer	Stress management
Mobile Sensing Platform (MSP) (12), Choudhury et al., 2008	Automatic activity recognition	Wearable device	Microphone, visible light, pho-transistor, three-axis digital accelerometer, digital barometer, temperature, and digital compass	Cognitive assistance
Fletcher et al. (13), Boston, USA, 2011	ECG heart monitoring	Neoprene band	Electrodermal activity (EDA), 3-axis motion, temperature and electrocardiogram (ECG)	Posttraumatic stress disorder (PTSD), drug-addiction
Vidal et al. (14), Lancaster University, UK, 2011	Monitoring the link between eye movements and cognition	Wearable eye tracking equipment	Electrooculography (EOG)	Mental health monitoring
Psyche (15), Lanata et al., University of Pisa, Italy, 2015	Monitoring of pathological mood states in bipolar disorder	T-shirt	Electrocardiogram-Heart Rate Variability (HRV), piezoresistive sensors, tri-axial accelerometers	Bipolar disorder monitoring
Prociow et al. (16), University of Nottingham, UK, 2012	Monitoring of bipolar symptoms	Wearable and environmental sensors	Belt-worn accelerometer, wearable light sensor, and Bluetooth encounters, PIR motion sensors, door switches, remote control usage monitor	Bipolar disorder monitoring
Jin et al. (17), Newcastle, UK, 2020	Analyzing behavior signals and speech under different emotions	Wristband	6D acceleration and angular sensor, temperature and humidity sensor, MEMS microphones, audio code unit	Anxiety and depression monitoring
Can et al. (18), Istanbul,Turkey, 2019	Continuous stress detection	Samsung Smartwatches, Empatica E4 wristbands	Heart rate activity, skin conductance, and accelerometer signals	Stress detection
Dagdanpurev et al. (19), Tokyo, Japan, 2018	Measuring heart rate (HR), high frequency (HF) component of heart rate variability (HRV), and the low frequency (LF)/HF ratio of HRV before, during, and after the mental task	Fingertip sensor	Photoplethysmograph (PPG) sensor	Major depressive disorder (MDD) screening system
Horiuchi et al. (20), Kanagawa, Japan, 2018	Measuring frequency, duration, and velocity of eye blinks as fatigue indices	Smart glass	Optical sensor: dye-sensitized photovoltaic cells	Fatigue assessment
Minguillon et al. (21), Granada, Spain, 2018	Classifying stress levels based on EEG, ECG, EMG, and GSR	Portable system	RABio w8 (real-time acquisition of biosignals, wireless, eight channels) system, electroencephalography, electrocardiography, electromyography, and galvanic skin response	Stress detection

(Continued)

TABLE 1 | Continued

Reference/product	Purpose	Device form factor	Sensors/parameters	Clinical application
Tsanas et al. (22), Edinburgh, United Kingdom, 2020	Monitoring activity, sleep, and circadian rhythm patterns	Wrist-worn sensor	Actigraphy, light, and temperature data	Posttraumatic stress disorder (PTSD) monitoring
Razjouyan et al. (23), Houston, USA, 2020	Quantifying physical activity patterns and nocturnal sleep using accelerometer data	Chest-worn sensor	Accelerometer data	Identification of cognitive impairment
McGinnis et al. (24), Burlington, USA, 2018	Detection of internalizing disorders in children	Belt-worn measurement unit	Motion data	Early detection of depression and anxiety disorders
Correia et al. (25), Braga, Portugal, 2020	Monitoring changes in Heart rate variability (HRV) elicited by a mentally stressful task	Earlobe PPG sensor	PPG sensor	Stress detection
Mohino-Herranz et al. (26), Madrid, Spain, 2015	Detection of stress, mental overload and emotional status in real time	Vest	Electrocardiogram (ECG) and thoracic electrical bioimpedance (TEB) signals	Stress detection

communications were found to be indicative of higher depressive symptoms, factors such as light intensity and smartphone screen usage were unlikely to be predictive of depressed mood (61). Also, physiological sensing, such as measurement of HRV data as a marker of decompensation, has shown promising results in monitoring schizophrenia. Furthermore, Depp et al. suggests that, in patients with schizophrenia, less GPS mobility is associated with greater negative symptom severity and motivational deficit (62). However, it is important to note that these relationships are not simple, and due to their complexities, passive data may still have limited clinical utility (63). Moreover, the validity of the data collected with consumer technologies and how the information collected on these platforms correlates with the disorders should be further investigated (4).

GENERAL WORKFLOW OF MENTAL HEALTH MONITORING SYSTEMS

The internet of things–based layers involved in the structure of mental health monitoring systems is demonstrated in **Figure 4**. The general system collects health information from different sensing devices, including wearables, mobile phone sensors, and ambient sensors. These devices are capable of recording various physiological and behavioral data, including HR, lung function, sleep duration, etc. The sensing layer is typically followed by network and analysis layers, which involve employing the appropriate data acquisition and analysis techniques to eventually translate the raw sensor to achieve digital phenotyping in the application layer. The network layer comprising wired or wireless networks (e.g., Bluetooth, Wi-Fi) performs the collection and storage of data. Thereafter, the raw data is processed to extract features in the analysis layer, which is discussed in detail below. The activity classification step eventually categorizes the extracted features into different conditions in the application layer.

Sensing Layer

An important aspect in the development of mental health monitoring systems is the sensing layer, whose role is to collect and transmit physiological, behavioral, and/or environmental parameters, thereby detecting changes in these in a continuous manner. To achieve the measurement of attributes related to individuals, various sensors, such as inertial sensors (e.g., accelerometer, gyroscope, or barometric pressure sensors) or physiological sensors (e.g., spirometer, ST, and blood pressure sensors) can be used. Moreover, a number of on-object sensors, such as environmental sensors, can be used for measuring indoor environmental conditions, including humidity and temperature.

The studies investigating wearable and mobile monitoring systems require constant interactions between fields such as medicine and computer science. Therefore, prerequisites are an important phase in these studies, some of which include the need for user consent on data collection; ethical approvals; and IT infrastructure, software, and hardware requirements (64). Furthermore, the type and number of sensors are often determined on the basis of application, all while considering their impact on factors such as battery consumption, obtrusiveness, and privacy (65). Sensor sampling rates, defined as the frequency at which the data is collected, should also be determined. Although high sampling rates provide more fine-grained pattern details, its impact on the device battery life is an important consideration (65). The studies are often performed as longitudinal and quantitative to provide the data required for building automatic predictive models and to measure a certain state over a period of time. Clinicians, health researchers, and patients should also couple mental health monitoring systems with frequent standard self-report scales, such as the Hamilton Depression Scale and the Young Mania Rating Scale, to

TABLE 2 | Smartphone-based physiological and emotional monitoring research prototypes.

Reference/product	Purpose	Platform	Sensors/parameters	Clinical application
Mobile phone sensors				
BeWell+ (27), Lin et al., USA, 2012	Evaluation of well-being based on physical activity, social interaction and sleep patterns	Sensor enabled smartphone application	Human voice, sleep, physical activity, and social interaction	Well-being monitoring
MoodRhythm, Matthews et al. (28), USA, 2016	Monitoring of bipolar symptoms	Sensor enabled smartphone application	Phone's light sensor, accelerometers, and microphone, as well as phone use events such as screen unlocks and battery charging state, communication patterns including SMS and call logs	Bipolar disorder monitoring
MONARCA (29), Faurholt-Jepsen et al., Copenhagen, Denmark, 2014	Monitoring of pathological mood states in bipolar disorder	Sensor enabled smartphone application	Phone usage, social activity measured as the number and length of incoming and outgoing phone calls and text messages, physical activity measured through step counter, mobility based on location estimation and speech activity collected by extraction of different voice features	Bipolar disorder monitoring
Evidence-Baes Behavior (eB2) app (30), Berrouiguet et al., Brest, France, 2018	Monitoring mobility patterns of patients with depression	Sensor enabled smartphone application	Physical activity, phone calls and message logs, app usage, nearby Bluetooth and Wi-Fi connections, and location	Depression monitoring
CrossCheck (31), Wang et al., New Hampshire, USA, 2016	Monitoring passive smartphone sensor data and self-reported indicators of schizophrenia	Sensor enabled smartphone application	Sleep, mobility, conversations, smartphone usage	Schizophrenia
SIMBA (32), Beiwinkel et al., Lüneburg, Germany, 2016	Tracking daily mood, physical activity, and social communication in bipolar patients	Sensor enabled smartphone application	Geolocation	Bipolar disorder monitoring
Ben-Zeev et al. (33), New Hampshire, USA, 2015	Analyzing stress, depression, and subjective loneliness over time	Smartphone sensor data	Speech, sleep duration, geospatial activity, and kinesthetic activity	Stress and depression monitoring
Jacobson et al. (34), Lebanon, USA, 2020	Predicting depressed mood within a day	Smartphone sensor data collected from "Mood Triggers" app	location, local weather information: temperature, humidity, precipitation, light level; heart rate information: average heart rate and heart rate variability; and outgoing phone calls	Depression monitoring
Pastor et al. (35), Barcelona, Spain, 2020	Digital phenotyping of patients with alcohol use disorder and anxiety symptoms	Smartphone sensor data collected from "HumanITcare" app	Sleep cycle, heart rate, movement patterns, and sociability	Monitoring anxiety symptoms and alcohol use disorder
Purple robot (36), Schueller et al., Chicago, USA, 2014	Depression detection	Sensor enabled web-based and smartphone application	physical activity, social activity	Depression monitoring
FINE, Dang et al., Hannover, Germany, 2016	Depression detection	Sensor enabled smartphone application	Smartphone use, social activity, movement	Depression monitoring
Mobilyze! (37), Burns et al., Chicago, USA, 2011	Prediction and intervention of depression	Sensor enabled smartphone application	Physical activity, social activity	Depression monitoring
PRIORI (38), Gideon et al., Michigan, USA, 2016	Analysis of mood in individuals with bipolar disorder	Sensor enabled smartphone application	Analysis of voice patterns collected from phone calls	Bipolar disorder monitoring
SIMPle 1.0 (39), Hidalgo-Mazzei et al., Barcelona, Spain, 2015	Bipolar disorder symptom management and psycho-educational Intervention	Sensor enabled smartphone application	Smartphone usage, calls, and physical activity	Bipolar disorder monitoring
Me app (40), Asare et al., Oulu, Finland, 2019	Predictive measures of major depressive disorder	Sensor enabled smartphone application	GPS location, physical activity, light, noise, screen interaction, battery, application notifications, calls, messages	Depression monitoring
SOLVD app (41), Cao et al., Houston, USA, 2020	Monitoring depressive symptoms of adolescents	Sensor enabled smartphone application	Steps, GPS, SMS, call, light, and screen time	Depression monitoring

(Continued)

TABLE 2 | Continued

Reference/product	Purpose	Platform	Sensors/parameters	Clinical application
Toi Meme app (42), Daregel et al., Paris, France, 2020	Monitoring of pathological mood states in bipolar disorder	Sensor enabled smartphone application	Motor activities (eg, number of steps, distance) measured using the smartphone's motion sensors	Bipolar disorder monitoring
DeMasi et al. (43), California, USA, 2017	Mood monitoring	Sensor enabled smartphone application	Activity and sleep tracking using accelerometer data	Depression and bipolar disorder monitoring
Mobile Sensing and Support (MOSS) app (44), Wahle et al., Zurich, Switzerland, 2016	Detecting depression and providing personalized CBT intervention	Sensor enabled smartphone application	WiFi, accelerometer, GPS, and phone use	Depression monitoring
Daus et al. (45), Stuttgart, Germany, 2020	Mood recognition in bipolar patients	Sensor enabled smartphone application	Location, Accelerometer, Smartphone usage, social interaction	Bipolar disorder monitoring

TABLE 3 | Ambient mental health monitoring research prototypes.

Reference	Sensors/parameters	Application
Ramón-Fernández et al. (46), Alicante, Spain, 2018	Level of noise in the room, flow of people moving through it, temperature, luminosity, and air quality	Stress detection
Kim et al. (47), Michigan, USA, 2017	Passive infrared motion sensors	Depression monitoring
Dawadi et al. (48), Washington, USA, 2013	Motion sensors on the ceiling, door sensors on cabinets and doors, and item sensors on selected	Cognitive assessment
Alam et al. (49), Yongin, Korea, 2016	Web of objects system comprising a smart home environment and lightweight biosensors	Psychiatric emergency
Hayes et al. (50), Oregon, USA, 2009	Measures of walking speed and amount of activity in the home	Cognitive assessment
Mielke et al. (51), Braunschweig, Germany, 2020	Detection of Psychomotor agitation by collecting motion data	Mental health monitoring
Kasteren et al. (52), Amsterdam, Netherlands, 2010	Activity monitoring using wireless sensor systems	Cognitive assessment
Bradford et al. (53), Canberra, Australia, 2013	Sensor-based in-home monitoring system for the elderly	Cognitive assessment
Ribonia et al. (54), Cagliari, Italy, 2016	SmartFABER sensor network collecting behavioral anomalies in the elderly	Cognitive assessment

evaluate their correlations. Therefore, the wearables and sensor-enabled smartphone applications are combined with self-reports to compare the sensor-derived measures with the reported conditions and determine the associations between the sensor-based auto-report and the self-report.

Network Layer

The network layer is responsible for connecting all the devices in the sensory layer and allows the data to be collected, stored, transmitted, shared, and aggregated. To protect the participants' privacy, the data is anonymized, encrypted, transmitted via secure connections, and finally stored in secure servers. The latter can be performed in various ways; e.g., data acquired directly from smartphones can be stored in the internal device memory. Additional methods exist for data collected using wearables, such as watches and wristbands, with which data may be sent to a smartphone via Bluetooth or stored in secure storage platforms managing sensitive data (66) or to a remote server or a computer using wired, wireless, or internet connection. Various networks, including Blutooth, Wi-Fi, Zigbee, and cellular, can be adopted in the network layer.

Analysis Layer

Data Labeling

Machine learning algorithms rely on training data to find patterns and generate predictive models. Therefore, the data labeling phase, including tagging the sensor data with their corresponding ground truth state, is important in training the final predictive models. The data labeling phase can happen on-site, at the hospital/clinic, or off-site while the participants perform their daily routines at home/work/etc. Data tagging can be achieved through periodic specialist/doctor assessments on-site or by phone (66, 67), using self-reports often presented by a mobile application in periodic intervals in which the participant is in charge of the labeling process (68, 69) or according to some event, such as labeling stress during an exam situation (70).

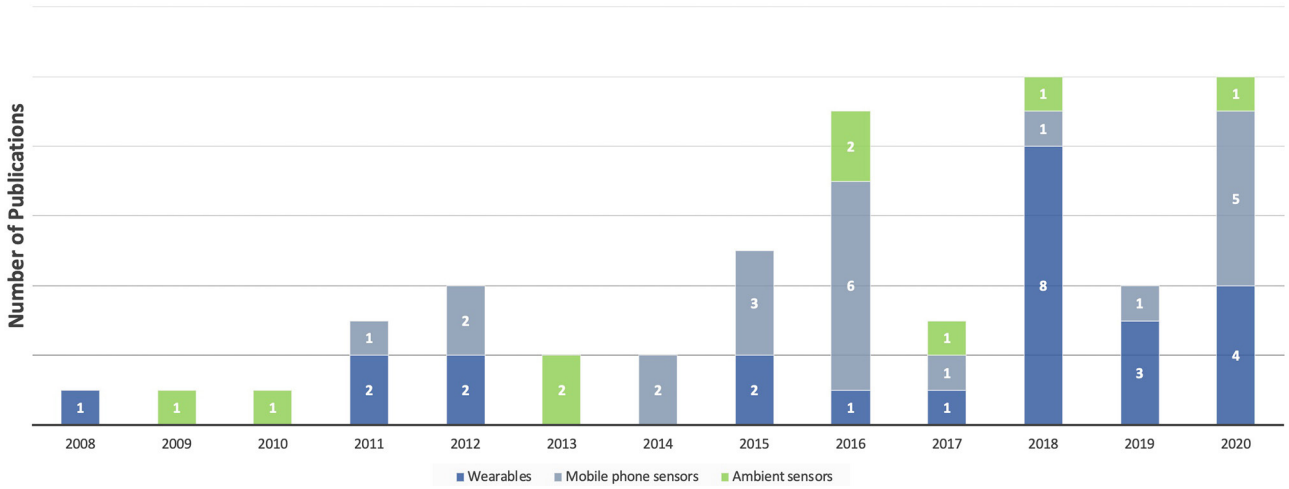


FIGURE 2 | Number of publications investigating monitoring systems in mental health from 2008 to 2020.

Data Analysis and Preprocessing

After collecting the data, exploratory data analysis and preprocessing are used to better understand and visualize the data and detect outliers. Preprocessing includes applying filters and transformations, such as scaling, quantization, binarization, and so on, to the raw data to reduce noise and remove outliers. Dimensionality reduction techniques, including principle component analysis (71) and multidimensional scaling (72), are among the most common preprocessing techniques. Feature extraction is then used to build feature vectors from the raw sensor data. Feature vectors are required for machine learning algorithms as they are representations of the original data. Arithmetic mean, standard deviation, min, max, skewness, kurtosis, root mean square, power spectrum density, energy, and correlation coefficient are some of the common extracted features for mental states detection (68, 73, 74).

Machine Learning Model Training

Different types of machine learning algorithms, including supervised, unsupervised, semi-supervised, transfer, and reinforcement learning, are used to find particular patterns and relationships from the vast amount of collected data and finally yield the appropriate predictive model(s). It should be noted that algorithms are often combined to get the final predictive models. However, there is a distinction between certain training schemes, such as user-dependent and user-independent (general) models. User-dependent models, trained with data from the specific user under consideration and capturing the specific behavior of each user, yield better results but require a lot of data training. User-independent models, trained with data from all other users excluding the target user, do not require any data for the target user but might not perform well for atypical users. Although Lu et al. suggest that user-dependent models perform much better for stress detection (75), some studies propose that hybrid models have the best of both user-dependent and -independent models (68, 76).

As mentioned previously, physiological and/or behavioral monitoring systems are often coupled with data analysis platforms whose role is to convert raw sensor data into measures of mental disorders. Accordingly, different machine learning software tools/libraries are used for evaluating and training personalized or general models. These include the Weka software and the scikit-learn library used for bipolar detection (77, 78) and anxiety recognition (79), respectively. Furthermore, InSTIL (Intelligent Sensing to Inform and Learn) is a software platform designed for digital phenotyping. The platform provides acquisition of sensor data from consumer smartphones and is being used by studies that seek to collect passive and active sensor signals (80). Another developed platform is the Remote Assessment of Disease and Relapse (RADAR)-base platform, which can be integrated with remote monitoring initiatives and caters to large-scale data collection (81). These platforms provide general remote data collection at scale, management of studies, and real-time visualizations and ensure user privacy and security. However, other tools, such as Matlab machine learning toolboxes, can be also used based on different application types.

Application Layer

Remote mental health monitoring is of great significance to individuals with mental disorders as well as their caregivers and physicians. Physiological and behavioral patterns can reflect mental states of the patients, and thus, recording such data provides physicians and caregivers with a useful method for accurate intervention and diagnosis. For example, data from accelerometers, GPS, and mobile phone sensors can be used to detect physiological patterns associated with depression or depressed mood states involved in bipolar disorder. Furthermore, HR and EDA sensors can be employed to monitor physiological markers that react to emotional experiences and

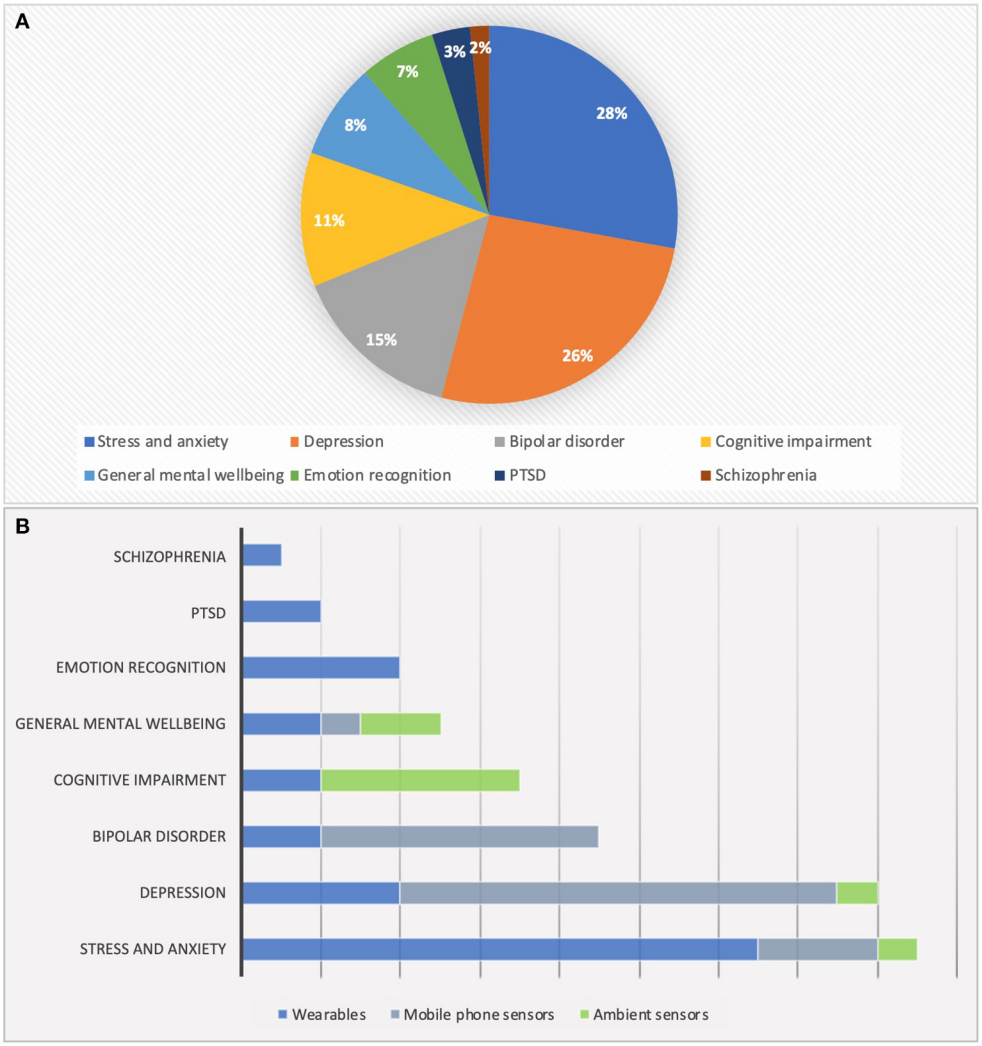


FIGURE 3 | (A) Domain of the reviewed papers. **(B)** Investigated sensing devices for different mental disorders.

allow emotion recognition and the detection of stress levels. Sleep, mobility patterns, and conversation frequency and patterns are among the parameters used for monitoring schizophrenia. Last, microphone, visible light, accelerometer, temperature, and digital compass sensors are often utilized to provide cognitive assistance.

Mental health monitoring systems can be also used to facilitate monitoring of symptoms in home and hospital environments. Using these systems instead of the conventional questionnaires or manual tests delivers particular information for physicians and caregivers and, thus, potentially assists self-management of well-being, reduces health care costs, and avoids undesirable consequences in a personalized manner. Furthermore, employing these systems to provide ambient assisted living, particularly for the elderly, provides an emergency system that is essential for monitoring and detecting abnormal physiological and behavioral states.

APPLICATIONS OF PHYSIOLOGICAL AND BEHAVIORAL MONITORING SYSTEMS

A wide range of devices aimed at physiological monitoring are presented, including wristbands and watches, chest straps, vests, garments and shirts, patches and sleeves (82–87). A list of wearable physiological and emotional monitoring systems can be found in **Table 1**. Various mental health studies on personal sensing have also utilized mobile phone sensors as they are widely owned and contain a large number of embedded sensors able to detect behavioral markers, such as sleep, social context, mood, and stress (**Table 2**). The use of smartphone sensors has been investigated for detecting the presence and severity of several mental disorders, including depression, bipolar disorder, and schizophrenia (29, 31, 88–90). Additionally, environmental or ambient sensors are generally installed in a single room, a house, or an entire building to understand the context of

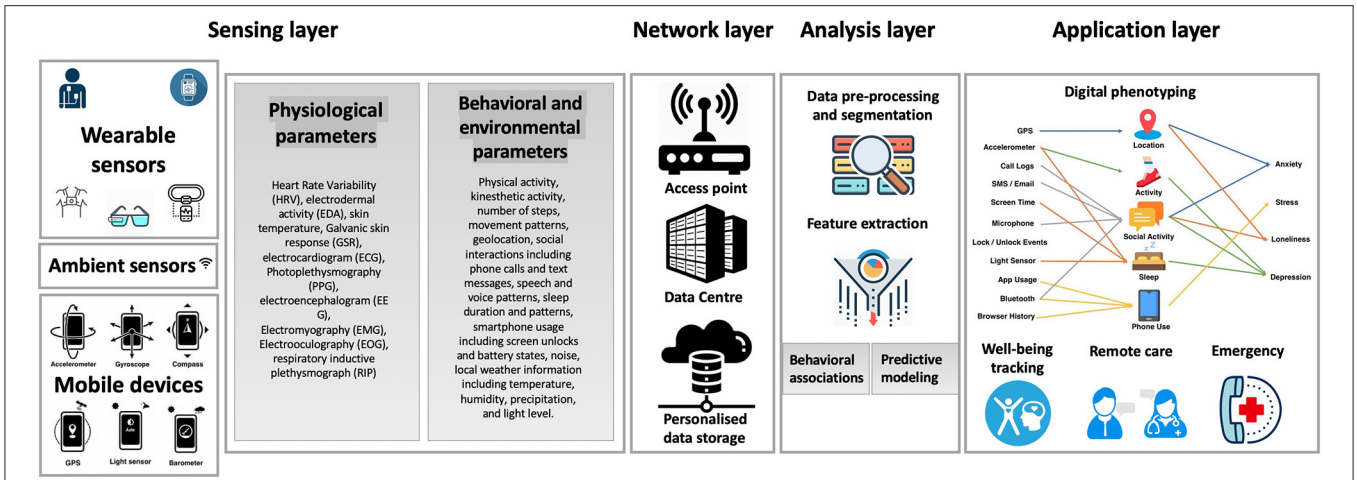


FIGURE 4 | Overview of the workflow of mental health monitoring systems.

that environment by gathering information from sensors and providing assistance to the inhabitants. Ambient sensors have been investigated in some mental health monitoring studies, especially for detecting cognitive impairment (**Table 3**). The sections below discuss in detail the various wearable, ambient, and smartphone-based systems that have been investigated in recent years for treatment and management of common mental health disorders.

Emotion and Autonomic Activity Recognition

Several wearable devices have been developed to monitor emotional information. The scarf-based monitor developed by Guo et al. uses an HR sensor and an EDA sensor to detect emotional states. The scarf also emits an odor and changes color to enhance the mood in response to negative emotions (**Figure 5**) (6). Similarly, a smart glove developed by Hui et al. utilizes a photoplethysmography (PPG) sensor for HR, EDA, ST, and EMG sensors to recognize different emotions, such as happiness, anger, fear, and disgust (7). To achieve emotion recognition using sensor data, a study by Jang et al. investigates the consistency on changes of bio-signals as physiological responses induced by six basic emotions—happiness, sadness, anger, fear, disgust, and surprise—using 60 different emotional stimuli. HR, skin conductance level (SCL), mean of ST (meanSKT), and mean of PPG (meanPPG) were measured before and during the presentation of the stimuli. The results suggest that biosensors are useful tools for emotion recognition as the physiological responses by six emotions were consistent over time; in particular, physiological features, such as SCL, HR, and PPG, are found to be very reliable (59). Moreover, a smart glass containing dye-sensitized photovoltaic cells, in which the optical sensors are positioned at the lateral side of the eye, has been designed for fatigue assessment based on eye blinks. The glass measures frequency, duration, and velocity of eye blinks as fatigue indices, which can potentially benefit the maintenance of physical and mental health (20). Also, a novel measurement technique using wearable eye tracking for mental health monitoring has been proposed based on the link between eye movements and cognition (14). A number of wearable monitoring systems have been studied to assess cognitive health in older adults. Using the data from a chest-worn sensor, it has been suggested that physical activity, including sedentary and light activities, percentage of walk and step count, and total sleep time and time in bed, are significant metrics for identifying cognitive frailty in older adults (23). Furthermore, a link between greater gait symmetry and better mental health has been identified as well by employing a high-speed 3-D motion sensing system to record gait mechanics in older adults (92). Another system is the mobile sensing platform, which is a small, wearable device designed for automatic activity recognition and cognitive assistance using on-body sensors, including microphone, visible light, pho-transistor, three-axis digital accelerometer, digital barometer, temperature, and digital compass (12).

Among studies employing ambient sensors, Alam et al. objectified different home appliances and sensors as a web of objects to create a smart home environment and combined it with lightweight biosensors and web-based psychiatric screening scales to assess patients' psychiatric symptoms (49). Additionally, with agitation being a symptom of many mental illnesses, including dementia, Mielke et al. attempted to detect psychomotor agitation patterns via installation of sensors in the apartment and building a smart home environment (51). The study proposes that the movement sequences identified as conspicuous or critical can be indicative of psychomotor agitation or possible mood and behavioral changes (51). Particularly, ambient sensors have been used in a couple of studies to achieve monitoring of cognitive health in older adults. In a study by Kasteren et al. state change sensors were located in doors, cupboards, refrigerator, etc., to provide an activity monitoring system and assist in caregiving for the elderly in their homes (52, 93). The Smarter Safer Homes project (53) and SmartFABER (54) sought to detect cognitive decline and

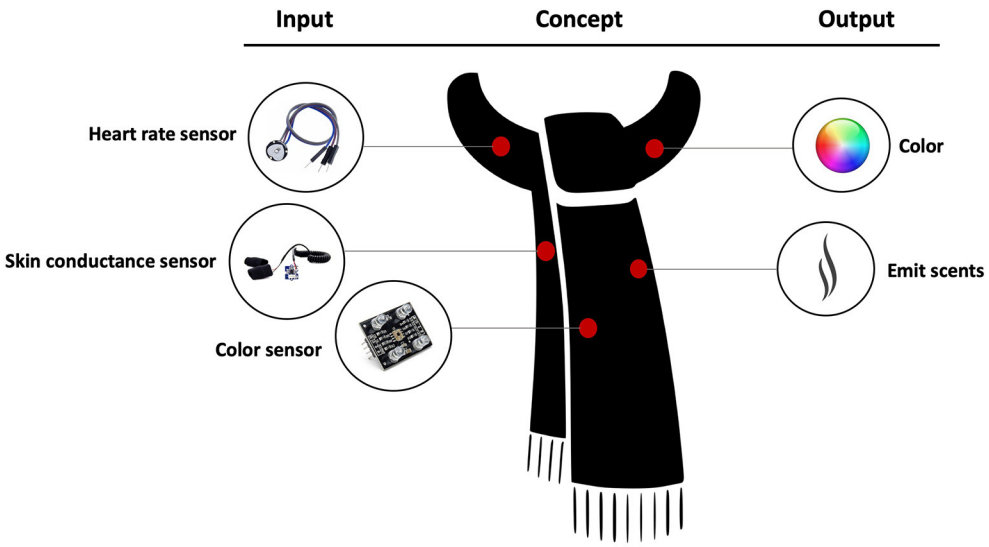


FIGURE 5 | Concept of a smart scarf designed for emotion recognition (91).

abnormal behaviors in the elderly by creating sensor-based in-home monitoring systems that acquire data about the interaction of the senior with the home environment. Early detection of cognitive impairment has also been investigated in (50), by installing an unobtrusive activity assessment system containing motion sensors and contact sensors to monitor the activity of older adults in their homes. Moreover, passive infrared motion sensors have been employed to monitor the daily activities of elderly persons and achieve long-term depression monitoring by creating smart homes (**Figure 6**) (47). Last, a smart home test bed was designed at Washington State University represented by an apartment that was instrumented with motion sensors on the ceiling and sensors on cabinets and doors as well as sensors on selected kitchen items. The task quality of smart home activities were quantified to assess the cognitive health of the individuals (48).

FIGURE 6 | Sample configuration in an elderly house for long-term monitoring of depression. [Modified from Kim et al., *IEEE Sensors*, 2017 (47)].

Stress and Anxiety Disorders

Anxiety disorders are characterized by unrealistic and excessive worry accompanied by symptoms such as extreme vigilance, motor tension, restlessness, muscle tension, and disturbed sleep (94). Various devices have been developed for emotion recognition and stress management applications. As such, the MyFeel wristband (8) developed by Sentio Solutions, Inc., and Reveal (9), collect HR, EDA, and ST to evaluate anxiety levels by applying data analytics and cognitive-behavioral techniques. Thync (10) developed by Thync Global, Inc., employs electrical nerve stimulation (TENS) and transcranial direct current stimulation (tDCS) as neuro-stimulation technologies to increase energy and lower stress. TENS delivers small electrical impulses that reach the brain through the spinal cord and may help alleviate stress levels by relaxing the muscles (95). tDCS has been found to prevent chronic stress in a preliminary animal study by modulating the neural activity through the application of a small current (96). A wireless sensor suite called AutoSense (11) also provides information relating to general stress with 90% accuracy by collecting and processing cardiovascular, respiratory, and thermoregularity measurements. AutoSense integrates six sensors in a small device and includes wireless transmission to a mobile phone in real time, which provides monitoring of physiological responses to real-life stressors and various behaviors that may be related to stress, such as drinking, smoking, physical activity, movement patterns, conversations,

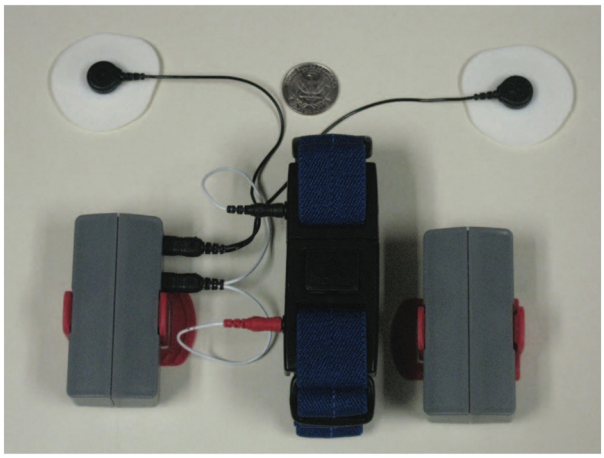


FIGURE 7 | AutoSense sensor suite including ECG leads and RIP connectors plug in to the unit on the side. The red clip in the wearable packaging is used to attach the unit to the (blue) RIP band that goes around the chest [Modified from Ertin et al., *SenSys*, 2011(11)].

etc., (**Figure 7**) (11). Furthermore, in a study by Can et al., HR activity, skin conductance, and accelerometer signals have been captured by smart wearable devices to detect and discriminate stress during situations such as contests, higher cognitive load lectures, and relaxed time activities using machine learning methods. The study used Samsung Smartwatches and Empatica E4 wristbands for data acquisition with higher accuracy and data quality obtained with the later class of devices (18). The combination of different physiological parameters seems to provide the highest accuracy for emotion classification. For example, Can et al. suggest that a higher classification accuracy was obtained when heart activity is combined with EDA than when these modalities were used separately (18).

In relation to stress detection, a portable stress detection system based on the RABio w8 (real-time acquisition of biosignals, wireless, eight channels) system has been designed, which is combined with software that comprises an application programming interface and a graphical user interface. By inducing different levels of stress, the system is demonstrated to be efficient in classifying three levels of stress (stress, relaxed, and neutral). The team is also working on a more wearable version of this system as a cap embedded with all the electronics (21). Furthermore, Jiang et al. develop a long-term wearable well-being sensing watch that continuously and unobtrusively detects details of behaviors that might be related to the onset of anxiety and autism (5). The portable wearable device consists of audio feature calculation without preserving the raw data and a variety of digital sensors for collecting multimodal data from the environment as well as physiological signals and behavioral activity. It has been demonstrated that the social audio features captured by the device are highly correlated with questionnaire scores for monitoring individuals with high to low autism traits (5). An undergoing study is also aiming to develop digital phenotyping of patients with alcohol use disorder and anxiety symptoms using data collected from a smartphone and a wearable sensor. In this study, factors including sleep cycle, HR, movement patterns, and sociability are sent and saved to the HumanITcare app (35).

Utilizing the sensors embedded in smart mobile phones, a new generation of well-being applications have been designed to automatically monitor multiple aspects of physical and mental health, for example, changes in the speech production process, which happens during stress. Therefore, microphones, embedded in mobile phones and carried ubiquitously by people, provide the opportunity for the non-invasive and continuous detection of voice-based stress (75). StressSense (75) and BeWell+ (27) are sensor-enabled smartphone apps designed to monitor user behavior. StressSense is designed to identify stress from the human voice using microphones embedded in mobile phones. BeWell+ continuously monitors user behavior, namely sleep, physical activity, and social interaction, and promotes improved behavioral patterns via feedback rendered as notifications, and this has been shown to successfully convey information and increase awareness (31). Ultimately, in the study by Boukhechba et al. it is suggested that mobility features, such as location entropy, are negatively associated with social anxiety with socially anxious students avoiding public areas and engagement in leisure activities. These findings were demonstrated by passively generating GPS data from an app installed on participants' personal mobile phones. Last, Ramon-Fernandez et al. propose an environmental stressor monitoring system to identify stressful environments based on parameters such as level of noise in the room, flow of people moving through it, temperature, luminosity, and air quality (46).

Depression

Depression is one of the most common mood disorders, and it is characterized by the absence of a positive affect; loss of interest in activities and experiences; low mood; and a range of associated emotional, cognitive, physical, and behavioral symptoms (94). Interpretation of data from smartphone-based geolocation sensors has identified the association between digital markers and mental illness concepts, especially mood (97). These geolocation-derived digital markers include number of locations visited, distance traveled, and time spent at a specific location. GPS embedded in participants' smartphone, GSM cellular network, and Wi-Fi are the common tools for measuring geolocation (97). Conversation frequency and duration, sleep disruption, social withdrawal and avoidance, mobility and GPS features are among the factors related to depression that can be detected using smartphone sensors (3). The evidence-based behavior (eB2) app has been designed to identify changes in the mobility patterns of patients with depression based on the smartphone's native sensors and advanced machine learning and signal processing techniques. The app captures inertial sensors, physical activity, phone calls and message logs, app usage, nearby Bluetooth and Wi-Fi connections, and location. The study has proposed that some specific mobility pattern changes can be indicators of relapses or clinical changes; however, that might not always be the case (30). Similarly, in a study by Ben-Zeev et al. it is suggested that increased geospatial activity

is associated with better depression scores. Nevertheless, the same study identified that geospatial activity and sleep duration are inversely associated with daily stress, and no correlations were identified between sleep duration, geospatial activity, or speech duration and loneliness (33). Furthermore, a momentary assessment study aims to predict future depressed mood within hour-to-hour time windows. Utilizing ideographically weighted machine learning models and passive mobile phone sensor data, the study shows the significant correlation between observed and predicted hourly mood (34).

A study by Jin et al. proposes an attention-based deep learning architecture combined with wearable sensing devices, which can effectively classify mental states by analyzing behavior signals and speech under different emotions (17). The sensing system of the wearable device is used for analyzing the degree of anxiety and depression and consists of 6-D acceleration and angular sensors, a temperature and humidity sensor, MEMS microphones, and an audio code unit (17). Also, a study by McGinnis et al. sought to identify anxiety and depression in children under the age of eight by developing objective measures. To recognize children with internalizing disorders, the study utilizes an instrumented fear induction task and captures the six degree–of-freedom movement of a child using data from a belt-worn inertial measurement unit. The results suggest that the collected motion data are sensitive to behaviors that are representative of child psychopathology (24). Last, a fingertip PPG sensor has been developed by Dagdanpurev et al. as a major depressive disorder (MDD) screening system (19). The system employs autonomic nerve transient responses induced by a mental task and logistic regression analysis to identify MDD patients from healthy subjects. The self-monitoring system achieved 83% sensitivity and 93% specificity in MDD screening determined by ECG-derived HRV (20).

Bipolar Disorder

Bipolar disorder (BD), characterized by recurrent episodes of depressed and manic mood states, is a leading cause of disability worldwide and is associated with significant functional impairment (98). In the management of BD, PSYCHE, a personalized wearable monitoring system was trialed on 10 bipolar patients exhibiting depression, hypomania, mixed state, and euthymia symptoms to assess the pathological mood states by recording physiological signals (**Figure 8**). The pathological mood states were assessed using a data mining strategy by recording the physiological signals through information collected from the autonomic nervous system. PSYCHE employs a commonly used measure of entropy to analyze more than 400 h of cardiovascular dynamics and characterize the symptoms. This helped to improve the diagnosis and management of psychiatric disorders (15). The system consists of a comfortable t-shirt with embedded sensors, such as textile electrodes, to monitor electrocardiogram (ECG) for HRV series, piezoresistive sensors for respiration activity, and tri-axial accelerometers for activity recognition, and provides a smartphone-based interactive platform and data visualization to the patient and physician, respectively. The t-shirt with the above-described sensors is coupled with embedded electronics to acquire and store the data and an internal tri-axial accelerometer to monitor movement activity (15). Once the electronic device is detected by a mobile application through Bluetooth, the physiological data streaming is initiated. Several algorithms are then used to process the physiological data collected and correlate it with the mental health status of the patient. The data is then uploaded automatically by patients who are connected at home to produce predictive results that allow the physician to optimize the patient's treatment. Therefore, the textile-integrated platform together with the smartphone framework offered long-term acquisition of HRV data and helped discriminate different pathological mood states and investigated the response to treatment on BD patients (15). It has been demonstrated that data coming from the PSYCHE platform (long-term HRV series) could be considered as a viable biomarker for discriminating of bipolar patients and their response to treatment (15).

In a study by Prociow et al. wearable sensing techniques were coupled with environmental sensors to observe the areas of life influenced by BD and match the collected information with bipolar symptoms (16). The system included a wearable light sensor and a belt-worn accelerometer, which was used to detect restless behaviors, such as psychomotor agitation and increased or decreased physical activity to identify mood states. Moreover, the environmental system included bed sensors and light detectors installed in the patients' home to monitor altered sleep patterns, such as insomnia, hypersomnia, and self-deprivation of sleep, which are important diagnostic indictors that a manic or depressive episode is occurring. Basic processing performed on the behavioral data yielded from these sensors provided information about early effects of a bipolar episode on activity patterns. The study was performed on four healthy subjects and one participant with BD. The recruited bipolar patient remained euthymic (asymptomatic) throughout the monitoring period (16).

Combining self-reports and passive sensor data, several studies have sought to monitor mood fluctuations involved in bipolar disorder. MoodRhythm is a smartphone application that combines self-report and passive sensing via the use of smartphone sensors for the long-term monitoring of bipolar disorder. The application incorporates existing self-report strategies from interpersonal and social rhythm therapy and combines them with inputs from smartphone sensors (28). Additionally, the MONARCA system was developed to investigate the monitoring, treatment and prediction of bipolar disorder episodes (**Figure 9**) (67, 78, 100–102). The MONARCA project is a smartphone-based behavior monitoring technology that leverages a variety of phone sensors to detect changes in mental states. The system also includes a feedback loop between patients and clinicians through a web portal that provides detailed historical overviews of a patient's data to generate measures of illness activity (29, 55). The system investigates self-monitored (subjective) or sensor-based automatically generated (objective) behavioral data to monitor mood states in patients with bipolar disorder. The automatically generated behavioral data is collected by smartphones supporting different types of sensors. Some of the sensor-based data on measures of illness activity include phone usage, social activity measured

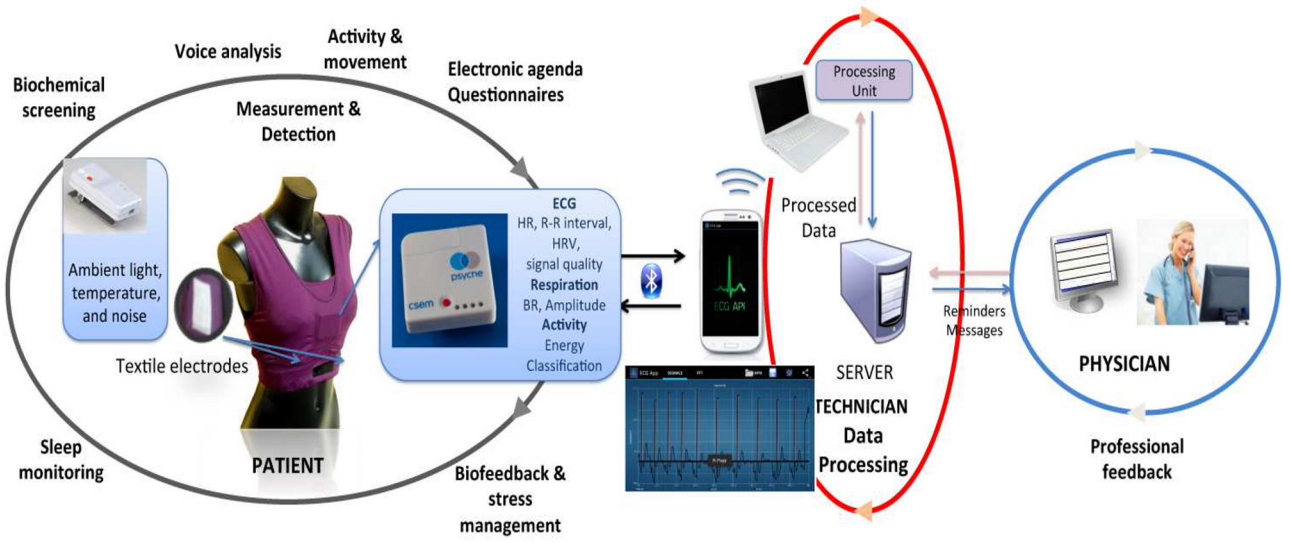


FIGURE 8 | Overview of the PSYCHE wearable monitoring system. The system involves a T-shirt with embedded textile sensors, mobile application for physiological signal acquisition from the wearable platforms, and feedback to patient and clinician [Modified from Lanata et al., *IEEE Journal of Biomedical and Health Informatics*, 2015 (15)].

as the number and length of incoming and outgoing phone calls and text messages, physical activity measured through a step counter, mobility based on location estimation, and speech activity collected by extraction of different voice features without recording the actual conversation. The study was performed for a 9-month trial period to investigate the correlations between subjective and objective behavioral data and the severity of depressive and manic symptoms. Altogether, the MONARCA system has been shown to recognize the clinical states of bipolar patients (depression/mania) with 72–81% accuracy based on location features. The system also shows high acceptance within the patients and health care providers and is proved to be effective in reducing the level of bipolar symptoms (29). Among other studies targeting bipolar patients, Social Information Monitoring for Patients with Bipolar Affective Disorder (SIMBA) is a smartphone monitoring app tracking daily mood, physical activity, and social communication in bipolar patients (32). Using the data collected from the SIMBA app, it has been suggested that the distance traveled by the bipolar patient is negatively associated with manic states (32). However, this is contrary to the findings of a study by Faurholt-Jepsen et al. (67) suggesting that the number of changes in communication are positively associated with manic states.

PTSD

PTSD follows significant trauma and is characterized by emotional numbness; intrusive reliving of the traumatic episode; disturbed sleep, including nightmares; and hyperarousal, such as exaggerated startle responses (94). A couple of wearable monitoring systems have been developed for individuals with PTSD. Namely, a wrist-worn actigraphy sensor has been investigated for the monitoring of activity, sleep, and circadian patterns in PTSD, which involves long-lasting symptoms, such as avoidance behaviors and sleep disturbance. The proposed sleep detection algorithm has been demonstrated to match the participants' sleep diaries. Also, determined from the greater nocturnal activity and awakenings, participants with PTSD were found to exhibit more fragmented sleep patterns compared with non-traumatized control groups (22). Furthermore, a wearable sensor system has been designed and implemented consisting of a neoprene band worn on the ankle as well as an optional custom-designed ECG heart monitor worn on the chest, both containing a Bluetooth radio that enables communication with the patient's mobile phone. The sensor bands contain circuitry for measuring EDA, three-axis motion, temperature, and ECG. The system has been studied for PTSD and drug addiction and works by presenting therapeutic and empathetic messages to the patient in the tradition of cognitive behavioral therapy when a specific arousal event is detected (13).

Schizophrenia

Schizophrenia is a neuropsychiatric syndrome exhibiting psychotic symptoms, such as hallucinations and delusions; negative symptoms, including loss of motivation and blunted affect; and last, cognitive symptoms, including impairments in attention, working memory, verbal fluency, and various aspects of social cognition (94). Continuous remote monitoring and identification of subjective and objective indicators of psychotic relapse is known to improve the management of schizophrenia-spectrum disorders. Accordingly, the CrossCheck project has identified statistically significant associations between passive smartphone sensor data and self-reported indicators of mental health in schizophrenia (31, 90). The study suggests that lower rates of physical activity are linked to

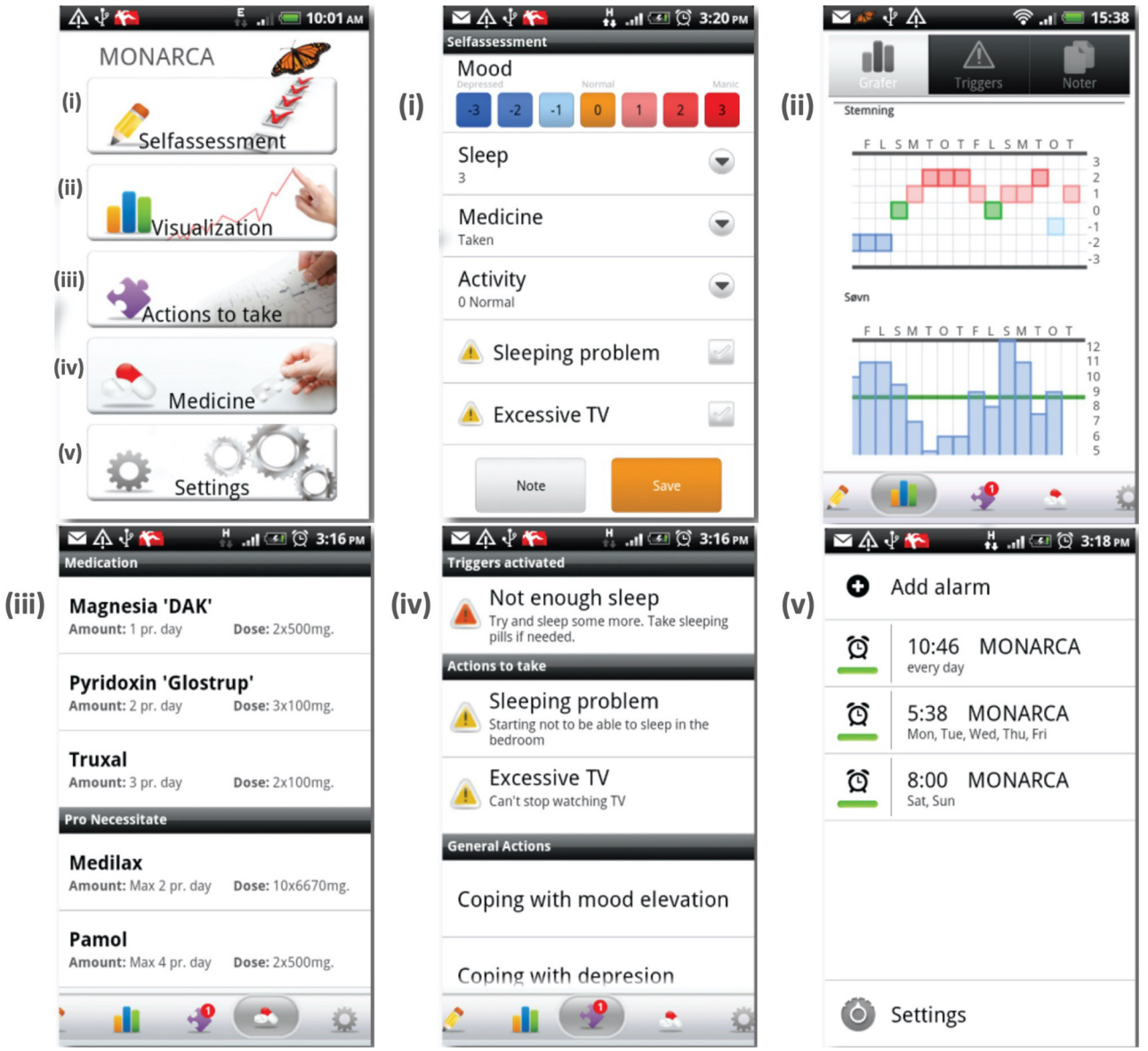


FIGURE 9 | The MONARCA application user interface [Modified from Bardram et al., 2013 (99)].

negative mental health. Also, it is demonstrated that patients showing lower sociability indicated particularly with fewer conversations during the morning and afternoon periods are likely to exhibit negative feelings. Nevertheless, it has been indicated that an increased number of phone calls and SMS messages can also be associated with negative dimensions of mental health as some individuals prefer to use the phone instead of face-to-face communication during negative mental states. Among other findings from monitoring of individuals with schizophrenia-spectrum disorders is that visiting fewer new places and going to bed later is associated with negative feelings, and getting up earlier is suggested to be linked with positive mood (103, 104).

DISCUSSION

The current review presents a comprehensive analysis of developed sensor-based strategies in the management of mental disorders that seek to monitor physiological and behavioral data as indirect measures of health states. The literature review identifies studies investigating monitoring systems for the management of different mental disorders, including stress and anxiety disorders, depression, BD, cognitive impairment, PTSD, and schizophrenia. Principally, mental health monitoring systems collect and interpret physiological and behavioral data to relate them to the symptoms of mental conditions and achieve digital phenotyping. Thereby, a number of environmental,

wearable, and mobile phone sensors have been employed to monitor parameters such as HRV, ST, EMG, BVP, blood pressure, etc. In each study, the type and variety of sensors were determined based on the application of the monitoring system as well as the behavioral and physiological parameters that were expected to be altered as a consequence of a specific disorder. The predicted mood states were often compared with the self-reports coupled with the monitoring systems. The majority of the studies have also suggested specific correlations between the collected behavioral and physiological parameters and the studied mental disorder. However, it should be noted that most of the developed technologies have not been thoroughly validated for their usefulness in clinical applications. Furthermore, the majority of monitoring systems target the general public and healthy individuals for health and fitness applications and are not certified as medical devices, and their impact on emotional states needs to be assessed (58). Also, these studies are subject to small sample sizes and short follow-up times. Therefore, there remain many technical and practical issues to overcome in digital approaches to mental disorders that include maintaining compliance over time, constraints on battery, manufacturing, durability, integration, constraints with analysis and interpretation of information, data accuracy, ethics/privacy issues, acceptability, and engagement issues (98). The mental health monitoring systems employing user-independent (general) models might also need to consider diversities such as age and ethnicity when introducing a predictive model. Moreover, despite the rapid development of sensing technology for monitoring physiological and behavioral data, challenges remain for the effective use of this abundant data. These challenges include processing the raw sensor data and managing the noise and artifacts, considering intrapersonal and interpersonal differences when interpreting the data, cost-effectiveness, and practicality of the device, providing privacy-protecting strategies, designing an accessible data storage and monitoring platform, and providing personalized coaching strategies. The only way to overcome these challenges is by interdisciplinary teams sharing expertise and methods and involvement of end users.

Although body-worn sensors that recognize human activities have the advantage of being at the user's hand, deploying these systems imposes some constraints, including protecting the user's privacy, being lightweight, offering limited computing power, handling the real world's noisy data and complexities, and comprising machine-learning algorithms that are trainable without requiring extensive human supervision. For instance, to avoid privacy issues, speech information should be automatically evaluated without preserving raw audio data (5). Another constraint is that, among wearable systems, the participant acceptability varies between different device forms, which is often related to the appearance and comfort of the wearing sensors. Accordingly, in a study by Huberty et al., it is demonstrated that wrist and upper arm sensors are more preferred than a sensor worn on the non-dominant hip in a sample of middle-aged women (105). Moreover, the results from a patient trial address the issue of non-adhering due to the discomfort of carrying extra devices, forgetfulness, and lack of familiarity with personal technology (16). Therefore, wearing and carrying extra devices on a day-to-day basis has been found to be impractical, and pursuing the environmental sensing via only a mobile phone seems to be ideal. However, it should be noted that wearable sensor features including skin conductance and temperature were found to be more accurate (78.3% accuracy) in identifying stress and poor mental health than mobile phone features (73.5% accuracy) such as phone usage and mobility patterns (106). Another shortcoming of mobile phone apps monitoring behavioral markers is their noticeable effect on the battery life, which can be a problem for users (61). Altogether, incorporating sensors monitoring physiological and behavioral parameters into a single, easy-to-use device has potential for the development of a practical mental health monitoring system. Furthermore, there is a crucial need for transparency and collaborative partnership between providers and patients because trust in the organization that collects the data and the purpose of data collection are the main factors for acceptability of the sensing systems. Additionally, ambient sensors can be expensive and limited to a particular physical space, their proper functioning requires calibration, and they might also raise privacy concerns. Nevertheless, these sensors have the advantage of not requiring direct contact with the user, hence, being free of wearing or carrying devices. Last, there is crucial need for patients, psychiatrists, and psychologists to adopt such means into their treatment and general practice (16).

When considering physiological and behavioral data as measures of mental health, it is important to consider that sensors do not sense the mental state itself, but a behavior that is the consequence of an underlying physiological alteration. Furthermore, relying solely on physiological sensor data might not be optimal as body changes may occur from other factors, such as food or medication intake and physical exercise. Thus, the mental health monitoring system should also consider the individual characteristics of each user. Another important aspect that should be considered is that the nature of the collected data from these sensors differs from the usual information that is available to and discussed with health care professionals. Patients have been found to express more willingness to consent to passive data retrieval from less personal sources, such as mobile phone screen time, whereas respondents were found to be at least willing to consent to more personal data, such as communication and location data, as well as audio recording and analysis (107). It has been noted that participants' comfort with sharing data depends on the data type and the recipient. Although some individuals were shown to be comfortable with sharing their health information, including sleep, mood, and physical activity with their doctors, they were less comfortable sharing personal data, such as communication logs, location, and social activity, especially with the electronic health record systems and their family (108).

Although there have been great advances in the automatic monitoring of mental health, there remain challenges that can pose several research opportunities. One of the key steps in the development of mental health monitoring systems is data labeling, which is used for finding associations between the sensor data and the corresponding true mental state at

that time span. Therefore, machine learning models should be further explored to evaluate their potential for mental health monitoring. Another aspect that needs to be considered when monitoring physiological and behavioral patterns is the inter-user differences. As aforementioned, user-dependent models are suggested to perform better than user-independent (general) models, and hybrid models can be used to combine the strengths of both models (68, 75). Therefore, the potential of this type of hybrid model should be further explored. Additionally, intra-user differences that refer to the variability of physiological and behavioral patterns for the same user should also be addressed. Last, the mental health monitoring systems need to be clinically validated and should be integrated with other systems, including user databases, administrative tools for caregivers and physicians, and support systems to achieve optimum communication platforms. The application of multimodal sensing technologies along with the appropriate machine learning methods is of significance for achieving monitoring of mental health.

CONCLUSION

In conclusion, facile physiological and behavioral sensing systems have the potential to directly enhance the management and monitoring of mental health. Use of these devices can increase patients' self-awareness, which can positively affect the management of their condition. This consciousness about one's mental state can prevent the worsening and further adverse effects of many mental disorders. Therefore, technology needs to be established in mental health as it has in other fields of medicine to provide practical and point-of-care devices for those who are affected. Development of monitoring devices will aid patients in the management of their condition and immensely increase their quality of life.

AUTHOR'S NOTE

The review paper has been written by MS under the supervision of MQ and PK. MQ and PK have reviewed the paper manuscript and made comments and statements. The authors of the review have complementary expertise, ensuring a balanced viewpoint. PK is a leader in the field of medical devices, developing innovative applications for the technology in a range of settings. MQ is a lecturer in biomedical engineering with expertise in optical spectroscopy and developing minimally-invasive monitoring devices. MS research focuses on non-invasive lithium monitoring in the management of bipolar disorder. Together, the authors provide a broad perspective on the topic, ensuring that the review is of interest to a wide audience.

AUTHOR CONTRIBUTIONS

MS and MQ: conceptualization and methodology. MQ and PK: validation, investigation, and supervision. MS: writing-original draft preparation, writing-review, and editing. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication, have read and agreed to the published version of the manuscript.

REFERENCES

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *Int. J. Epidemiol.* (2014) 43:18. doi: 10.1093/ije/dyu038
2. Mohr DC, Shilton K, Hotopf M. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. *Npj Digit Med.* (2020) 3:45. doi: 10.1038/s41746-020-0251-5
3. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu Rev Clin Psychol.* (2017) 13:23–47. doi: 10.1146/annurev-clinpsy-032816-044949
4. Roberts LW, Chan S, Torous J. New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis. *Npj Digit Med.* (2018) 1:20176. doi: 10.1038/s41746-017-0006-0
5. Jiang L, Gao B, Gu J, Chen Y, Gao Z, Ma X. Wearable long-term social sensing for mental wellbeing. *IEEE Sens J.* (2019) 199:8532–42. doi: 10.1109/JSEN.2018.2877427
6. Guo C, Chen YV, Qian ZC, Ma Y, Dinh H, Anasingaraju S. Designing a smart scarf to influence group members' emotions in ambience: design process and user experience. In: Antona M, Stephanidis C, editors. *International Conference on Universal Access in Human-Computer Interaction.* Interaction Techniques and Environments. Vol. 9738. Cham: Springer. doi: 10.1007/978-3-319-40244-4_38
7. Hui TKL, Sherratt RS. Coverage of emotion recognition for common wearable biosensors. *Biosensors*. (2018) 8:30. doi: 10.3390/bios8020030
8. Sentio Solutions Inc. MyFeel'. Available online at: https://www.myfeel.co/science (accessed March 24, 2021).
9. Awake Labs. Reveal'. Available online at: https://awakelabs.com (accessed March 24, 2021).
10. Thync Global Inc. Thync'. Available online at: https://www.thync.com (accessed March 24, 2021).
11. Ertin E, Stohs N, Kumar S, A. Raij, al'Absi M, Shah S. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In: *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems - SenSys '11*. Seattle, WA. (2011). p. 274.
12. Choudhury T, Borriello G, Consolvo S, Haehnel D, Harrison B, Hemingway, B. The mobile sensing platform: an embedded activity recognition system. *IEEE Pervasive Comput*. (2008) 7:32–41. doi: 10.1109/MPRV.2008.39
13. Fletcher RR, Tam S, Omojola O, Redemske R, Kwan J. Wearable sensor platform and mobile application for use in cognitive behavioral therapy for drug addiction and PTSD. *Conf Proc IEEE Eng Med Biol Soc*. (2011) 2011:1802–5. doi: 10.1109/IEMBS.2011.6090513
14. Vidal M, Turner J, Bulling A, Gellersen H.. Wearable eye tracking for mental health monitoring. *Comput Commun.* (2012) 351:1306–11. doi: 10.1016/j.comcom.2011.11.002
15. Lanata, Valenza G, Nardelli M, Gentili C, Scilingo EP. Complexity index from a personalized wearable monitoring system for assessing remission in mental health. *IEEE J Biomed Health Inform*. (2015) 19:132–9. doi: 10.1109/JBHI.2014.2360711
16. Prociow P, Wac K, Crowe J. Mobile psychiatry: towards improving the care for bipolar disorder. *Int J Ment Health Syst*. (2012) 6:5. doi: 10.1186/1752-4458-6-5
17. Jin J, Gao B, Yang S, Zhao B, Luo L, Woo WL. Attention-block deep learning based features fusion in wearable social sensor for mental wellbeing evaluations. *IEEE Access*. (2020) 8:89258–68. doi: 10.1109/ACCESS.2020.2994124

18. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. *Sensors*. (2019) 19:1849. doi: 10.3390/s19081849
19. Dagdanpurev S, Sun G, Shinba T, Kobayashi M, Kariya N, Choimaa L. Development and clinical application of a novel autonomic transient response-based screening system for major depressive disorder using a fingertip photoplethysmographic sensor. *Front Bioeng Biotechnol*. (2018) 6:64. doi: 10.3389/fbioe.2018.00064
20. Horiuchi R, Ogasawara T, Miki, N. Fatigue assessment by blink detected with attachable optical sensors of dye-sensitized photovoltaic cells. *Micromachines*. (2018) 9:310. doi: 10.3390/mi9060310
21. Minguillon J, Perez E, Lopez-Gordo M, Pelayo FM. Sanchez-carrion. Portable system for real-time detection of stress level. *Sensors*. (2018) 18:2504. doi: 10.3390/s18082504
22. Tsanas A, Woodward E, Ehlers A. Objective characterization of activity, sleep, and circadian rhythm patterns using a wrist-worn actigraphy sensor: insights into posttraumatic stress disorder. *JMIR MHealth UHealth*. (2020) 8:e14306. doi: 10.2196/14306
23. Razjouyan J, Najafi B, Horstman M, Sharafkhaneh A, Amirmazaheri M, Zhou, H.. Toward using wearables to remotely monitor cognitive frailty in community-living older adults: an observational study. *Sensors*. (2020) 20:2218. doi: 10.3390/s20082218
24. McGinnis EW, McGinnis RS, Hruschak J, Bilek E, Ip K, Morlen D. Wearable sensors detect childhood internalizing disorders during mood induction task. *PLoS ONE*. (2018) 13:e0195598. doi: 10.1371/journal.pone.0195598
25. Correia B, Dias N, Costa P, Pêgo JM. Validation of a wireless bluetooth photoplethysmography sensor used on the earlobe for monitoring heart rate variability features during a stress-inducing mental task in healthy individuals. *Sensors*. (2020) 204:3905. doi: 10.3390/s20143905
26. Mohino-Herranz, Gil-Pita R, Ferreira J, Rosa-Zurera M, Seoane F. Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones. *Sensors*. (2015) 150:25607–27. doi: 10.3390/s151025607
27. Lin M, Campbell AT, Choudhury T, Lane ND, Mohammod M, Yang, X.. BeWell+: multi-dimensional wellbeing monitoring with community-guided user feedback and energy optimization. In: *Proceedings of the Conference on Wireless Health - WH '12*. San Diego, CA. (2012). p. 1–8.
28. Matthews M, Abdullah S, Murnane E, Voida S, Choudhury T, Gay G. Development and evaluation of a smartphone-based measure of social rhythms for bipolar disorder. *Assessment*. (2016) 23:472–83. doi: 10.1177/1073191116656794
29. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram J, Kessing LV. Daily electronic monitoring of subjective and objective measures of illness activity in bipolar disorder using smartphones– the MONARCA II trial protocol: a randomized controlled single-blind parallel-group trial. *BMC Psychiatry*. (2014) 14:309. doi: 10.1186/s12888-014-0309-5
30. Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Camacho RC, Baca-García E. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the Evidence-Based Behavior (eB2) study. *JMIR Mhealth Uhealth*. 21:e16399. doi: 10.2196/16399
31. Wang R, Aung MSH, Abdullah S, Brian R, Campbell AT, Choudhury, T.. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg. (2016). p. 886–897.
32. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian, G.. Using smartphones to monitor bipolar disorder symptoms: a pilot study. *JMIR Ment Health*. (2016) 3:e2. doi: 10.2196/mental.4560
33. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior mental health. *Psychiatr Rehabil J*. (2015) 38:218–26. doi: 10.1037/prj0000130
34. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors*. (2020) 202:3572. doi: 10.3390/s20123572
35. Pastor N, Khalilian E, Caballeria E, Morrison D, Sanchez Luque U, Matrai S. Remote Monitoring Telemedicine (REMOTE) platform for patients with anxiety symptoms and alcohol use disorder: protocol for a case-control study. *JMIR Res Protoc*. (2020) 9:e16964. doi: 10.2196/16964
36. Schueller SM, Begale M, Penedo FJ, Mohr DC. Purple: a modular system for developing deploying behavioral intervention technologies. *J Med Internet Res*. (2014) 16:e181. doi: 10.2196/jmir.3376
37. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E. Harnessing context sensing to develop a mobile intervention for depression. *J Med Internet Res*. (2011) 13:e55. doi: 10.2196/jmir.1838
38. Gideon, Provost EM, McInnis M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai. (2016) p. 2359–2363.
39. Hidalgo-Mazzei D, Mateu A, Reinares M, Undurraga J, C. del Bonnín M, Sánchez-Moreno J. Self-monitoring and psychoeducation in bipolar patients with a smart-phone application (SIMPLe) project: design, development and studies protocols. *BMC Psychiatry*. (2015) 15:52. doi: 10.1186/s12888-015-0437-6
40. Asare O, Visuri A, Ferriera DST. Towards early detection of depression through smartphone sensing. In: *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers - UbiComp/ISWC '19*. London. (2019). p. 1158–1161.
41. Cao, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. *JMIR Ment Health*. (2020) 7:e14045. doi: 10.2196/14045
42. Dargél, Mosconi E, Masson M, Plaze M, Taieb F, Von Platen C. Toi Même, a mobile health platform for measuring bipolar illness activity: protocol for a feasibility study. *JMIR Res Protoc*. (2020) 9:e18818. doi: 10.2196/18818
43. DeMasi O, Feygin S, Dembo A, Aguilera A, Recht B. Well-being tracking via smartphone-measured activity and sleep: cohort study. *JMIR MHealth UHealth*. (2017) 50:e137. doi: 10.2196/mhealth.7820
44. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR MHealth UHealth*. (2016) 4:e111. doi: 10.2196/mhealth.5960
45. Daus H, Bloecher T, Egeler R, De Klerk R, Stork W, Backenstrass M. Development of an emotion-sensitive mHealth approach for mood-state recognition in bipolar disorder. *JMIR Ment Health*. (2020) 7:e14267. doi: 10.2196/14267
46. de Ramón-Fernández, Ruiz-Fernández D, Marcos-Jorquera D, V. Gilart-Iglesias. A distributed model for stressors monitoring based on environmental smart sensors. *Sensors*. (2018) 18:1935. doi: 10.3390/s18061935
47. Kim JY, Liu N, Tan HX, Chu CH. Unobtrusive monitoring to detect depression for elderly with chronic illnesses. *IEEE Sens J*. (2017) 177:5694–704. doi: 10.1109/JSEN.2017.2729594
48. Dawadi PN, Cook DJ, Schmitter-Edgecombe M, Parsey C. Automated assessment of cognitive health using smart home technologies. *Technol Health Care*. (2013) 21:323–43. doi: 10.3233/THC-130734
49. Alam M, Abedin S, Al Ameen M, Hong C. Web of objects based ambient assisted living framework for emergency psychiatric state prediction. *Sensors*. (2016) 16:1431. doi: 10.3390/s16091431
50. Hayes TL, Abendroth F, Adami A, Pavel M, Zitzelberger TA. Unobtrusive assessment of activity patterns associated with mild cognitive impairment. (2009). doi: 10.4017/gt.2010.09.02.109.00
51. Mielkea C, Antonsa R. Detection of psychomotor agitation pattern from motion sensor data in a living environment of a patient with dementia. *Digit Pers Health Med*. (2020) 270:746–50. doi: 10.3233/SHTI200260
52. van Kasteren TLM, Englebienne G, Kröse BJA. An activity monitoring system for elderly care using generative and discriminative models. *Pers Ubiquitous Comput*. (2010) 14:489–98. doi: 10.1007/s00779-009-0277-9
53. Bradford D. Detecting degeneration: monitoring cognitive health in independent elders. *Nnu Int Conf IEEE Eng Med Biol Soc*. (2013) 2013:029–032. doi: 10.1109/EMBC.2013.6611176

54. Ribonia D, Bettini C. SmartFABER: recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment. *Artif Intell Med*. (2016) 67:57–74. doi: 10.1016/j.artmed.2015.12.001
55. Frost M, Marcu G, Hansen R, Szaántó K, Bardram J. The MONARCA self-assessment system: persuasive personal monitoring for bipolar patients. In: *Presented at the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare,* Dublin: Republic of Ireland. (2011). doi: 10.4108/icst.pervasivehealth.2011.246050
56. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun*. (2015) 71:10–49. doi: 10.1016/j.specom.2015.03.004
57. Ma J, Li C, Kwan MP, Kou L, Chai Y. Assessing personal noise exposure and its relationship with mental health in Beijing based on individuals' space-time behavior. *Environ Int*. (2020) 139:105737. doi: 10.1016/j.envint.2020.105737
58. Taj-Eldin M, Ryan C, O'Flynn B, Galvin P. A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers. *Sensors*. (2018) 182:4271. doi: 10.3390/s18124271
59. Jang EH, Kim AY, Kim SH, Yu HY, Sohn JH. Internal consistency of physiological responses during exposure to emotional stimuli using biosensors'. In: *Proceedings of the 6th International Joint Conference on Pervasive and Embedded Computing and Communication Systems*. Lisbon. (2016). p. 110–115.
60. Vallance JK. Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005–2006). *Prev Med*. (2011) 53:284–8. doi: 10.1016/j.ypmed.2011.07.013
61. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges potential solutions. *J Med Internet Res*. (2018) 20:e10131. doi: 10.2196/10131
62. Depp CA, Bashem J, Moore RC, Holden JL, Mikhael T, Swendsen J. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *Npj Digit Med*. (2019) 2:108. doi: 10.1038/s41746-019-0182-1
63. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB. Automated analysis of free speech predicts psychosis onset in high-risk youths. *Npj Schizophr*. (2015) 1:15030. doi: 10.1038/npjschz.2015.30
64. Riegler M, Gurrin C, Johansen D, Johansen H, Halvorsen P, Lux M. Multimedia and medicine: teammates for better disease detection and survival. In: *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*. Amsterdam. (2016). p. 968–977.
65. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspect Psychol Sci*. (2016) 11:838–54. doi: 10.1177/1745691616650285
66. University of Oslo. Services for sensitive data (TSD). (2018). Available online at: http://www.uio.no/english/services/it/research/sensitive-data/
67. Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Margrethe Christensen E, Bardram JE. Behavioral activities collected through smartphones the association with illness activity in bipolar disorder: smartphone data in bipolar disorder. *Int J Methods Psychiatr Res*. (2016) 25:309–23. doi: 10.1002/mpr.1502
68. Garcia-Ceja E, Osmani V, Mayora O. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE J Biomed Health Inform*. (2016) 20:1053–60. doi: 10.1109/JBHI.2015.2446195
69. Zenonos, Khan A, Kalogridis G, Vatsikas S, Lewis T, Sooriyabandara M. HealthyOffice: mood recognition at work using smartphones and wearable sensors. In: *2016 IEEE Annual Conference on Pervasive Computing and Communications Workshops*. Sydney, NSW: PerCom (2016). p. 1–6. doi: 10.1109/PERCOMW.2016.7457166
70. Bauer G, Lukowicz P. Can smartphones detect stress-related changes in the behaviour of individuals? In: 2012 IEEE Annual Conference on Pervasive Computing and Communications Workshops. Lugano: PerCom (2012). p. 423–426. doi: 10.1109/PerComW.2012.6197525
71. Karl Pearson FRS LIII. On lines planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci*. (2010) 21:559–72. doi: 10.1080/14786440109462720
72. GOWER JC. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*. (1966) 53:325–38. doi: 10.1093/biomet/53.3-4.325
73. Maxhuni, Muñoz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive Mob Comput*. (2016) 31:50–66. doi: 10.1016/j.pmcj.2016.01.008
74. Garcia-Ceja E, Riegler M, Jakobsen P, Tørresen J, Nordgreen T, Oedegaard KJ. Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients. In: *Proceedings of the 9th ACM Multimedia Systems Conference,* Amsterdam. (2018). p. 472–477.
75. Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Andrew T. Campbell. StressSense: detecting stress in unconstrained acoustic environments using smartphones. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. Pittsburgh, PA (2012) 351.
76. Xu Q, Nwe TL, Guan C. Cluster-based analysis for personalized stress evaluation using physiological signals. *IEEE J Biomed Health Inform*. (2015) 18:275–81. doi: 10.1109/JBHI.2014.2311044
77. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G. Smartphone-based recognition of states state changes in bipolar disorder patients. *IEEE J Biomed Health Inform*. (2015) 19:140–8. doi: 10.1109/JBHI.2014.2343154
78. Gruenerbl, Osmani V, Bahle G, Carrasco JC, Oehler S, Mayora, O.. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. In: *Proceedings of the 5th Augmented Human International Conference on - AH '14*. Kobe. (2014). p. 1–8.
79. Miranda D, Favela J, Arnrich B. Detecting anxiety states when caring for people with dementia. *Methods Inf Med*. (2017) 56:55–62. doi: 10.3414/ME15-02-0012
80. Barnett S, Huckvale K, Christensen H, Venkatesh S, Mouzakis K, Vasa, R. Intelligent Sensing to Inform Learn (InSTIL): a scalable governance-aware platform for universal, smartphone-based digital phenotyping for research clinical applications. *J Med Internet Res*. (2019) 211:e16399. doi: 10.2196/preprints.16399
81. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck, D.. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, mobile devices. *JMIR MHealth UHealth*. (2019) 7:e11734. doi: 10.2196/11734
82. 'Wahoo Fitness'. Available online at: https://eu.wahoofitness.com/devices/heart-rate-monitors/wahoo-tickr-x-heart-rate-strap (accessed March 24, 2021).
83. 'Suunto'. Available online at: https://www.suunto.com/Products/sports-watches/suunto-9/suunto-9-baro-black/ (accessed March 24, 2021).
84. 'Healbe, Inc.' Available online at: https://healbe.com/us/gobe2/ (accessed March 24, 2021).
85. 'VialConnect.' Available online at: https://vitalconnect.com/solutions/vitalpatch/ (accessed March 24, 2021).
86. 'Komodo Technologies.' Available online at: http://komodotec.com/product/aio-sleeve/ (accessed March 24, 2021).
87. 'BrainPower,LLC.' Available online at: http://www.brain-power.com/ (accessed March 24, 2021).
88. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR MHealth UHealth*. (2017) 5:e112. doi: 10.2196/mhealth.7297
89. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J Med Internet Res*. (2015) 17:e175. doi: 10.2196/jmir.4273
90. Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer EA, Mistler LA. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. *Psychiatr Serv*. (2016) 67:558–61. doi: 10.1176/appi.ps.201500130
91. Guo C, Chen YV, Qian ZC, Ma Y, Dinh H, Anasingaraju S. Designing a smart scarf to influence group members' emotions in ambience: design process and user experience. In: Antona M, Stephanidis C, editors. *Universal Access in Human-Computer Interaction*. Interaction Techniques and Environments. Vol. 9738. Cham: Springer (2016). p. 392–402.

92. Nagano H, Sarashina E, Sparrow W, Mizukami K, Begg R. General mental health is associated with gait asymmetry. *Sensors*. (2019) 192:4908. doi: 10.3390/s19224908
93. van Kasteren T, Noulas A, Englebienne G, Kröse B. Accurate activity recognition in a home setting. *Assoc Comput*. (2008). 5:e56. doi: 10.1145/1409635.1409637
94. National Collaborating Centre for Mental Health (Great Britain), National Institute for Health and Clinical Excellence (Great Britain), British Psychological Society, and Royal College of Psychiatrists, Common mental health disorders: identification and pathways to care. Liecester; London: British Psychological Society; Royal College of Psychiatrists (2011).
95. Yoshimoto S, Babygirija R. Anti-stress effects of Transcutaneous Electrical Nerve Stimulation (TENS) on colonic motility in rats. *Dig Dis Sci*. (2012) 57:1213–21. doi: 10.1007/s10620-012-2040-8
96. Fregni F. Transcranial direct current stimulation (tDCS) prevents chronic stress-induced hyperalgesia in rats. *Brain Stimulat*. (2018) 11:299–301. doi: 10.1016/j.brs.2017.11.009
97. Fraccaro P, Beukenhorst A, Sperrin M, Harper S, Palmier-Claus J, Lewis S. Digital biomarkers from geolocation data in bipolar disorder schizophrenia: a systematic review. *J Am Med Inform Assoc*. (2019) 261:1412–20. doi: 10.1093/jamia/ocz043
98. Harrison PJ, Geddes JR, Tunbridge EM. The emerging neurobiology of bipolar disorder. *Trends Neurosci*. (2018) 41:18–30. doi: 10.1016/j.tins.2017.10.006
99. Bardram E, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In: *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE) International Conference on Acoustics, Speech, and Signal Processing*. ICASSP, Paris (2013). p. 2627.
100. Frost M, Doryab A, Faurholt-Jepsen M, Kessing LV, Bardram JE. Supporting disease insight through data analysis: refinements of the monarca self-assessment system. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '13*. Zurich. (2013) 133.
101. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. *Psychiatry Res*. (2014) 217:124–7. doi: 10.1016/j.psychres.2014.03.009
102. Fitzgerald J, Fenniri H. Cutting edge methods for non-invasive disease diagnosis using e-tongue and e-nose devices. *Biosensors*. (2017) 7:59. doi: 10.3390/bios7040059
103. Ben-Zeev D, Scherer EA, Brian RM, Mistler LA, Campbell AT, Wang R. Use of multimodal technology to identify digital correlates of violence among inpatients with serious mental illness: a pilot study. *Psychiatr Serv*. (2017) 680:1088–92. doi: 10.1176/appi.ps.201700077
104. Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MSH, et al. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatr Rehabil J*. (2017) 40:266–75. doi: 10.1037/prj0000243
105. Huberty J, Ehlers DK, Kurka J, Ainsworth B, Buman M. Feasibility of three wearable sensors for 24 hour monitoring in middle-aged women. *BMC Womens Health*. (2015) 15:55. doi: 10.1186/s12905-015-0212-3
106. Sano, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E. Identifying objective physiological markers modifiable behaviors for self-reported stress mental health status using wearable sensors mobile phones: observational study. *J Med Internet Res*. (2018) 20:e210. doi: 10.2196/jmir.9410
107. D. Di Matteo, Fine A, Fotinos K, Rose J, Katzman, M.. Patient Willingness to Consent to Mobile Phone Data Collection for Mental Health Apps: Structured Questionnaire. JMIR Ment. Health. 5:e56. (2018). doi: 10.2196/mental.9539
108. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. *JMIR MHealth UHealth*. (2019) 7:e12578. doi: 10.2196/12578

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

BRIEF RESEARCH REPORT
published: 09 April 2021
doi: 10.3389/fdgth.2021.624261

Concurrent Improvement Observed in Patient-Reported Burden and Sensor-Collected Medication Use Among Patients Enrolled in a COPD Digital Health Program

Leanne Kaye[1]*, *Rahul Gondalia*[1], *Meredith A. Barrett*[1], *Melissa Williams*[2] and *David A. Stempel*[2]

[1] ResMed Science Center, San Francisco, CA, United States, [2] Propeller Health, San Francisco, CA, United States

OPEN ACCESS

Edited by:
Richard Ribon Fletcher, Massachusetts Institute of Technology, United States

Reviewed by:
Palash Chandra Banik, Bangladesh University of Health Sciences, Bangladesh
Parisis Gallos, National and Kapodistrian University of Athens, Greece

***Correspondence:**
Leanne Kaye
leanne.kaye@resmed.com

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

Received: 31 October 2020
Accepted: 18 February 2021
Published: 09 April 2021

Citation:
Kaye L, Gondalia R, Barrett MA, Williams M and Stempel DA (2021) Concurrent Improvement Observed in Patient-Reported Burden and Sensor-Collected Medication Use Among Patients Enrolled in a COPD Digital Health Program. Front. Digit. Health 3:624261. doi: 10.3389/fdgth.2021.624261

Background: The COPD assessment test (CAT) is an 8-item questionnaire widely used in clinical practice to assess patient burden of disease. Digital health platforms that leverage electronic medication monitors (EMMs) are used to track the time and date of maintenance and short-acting beta-agonist (SABA) inhaler medication use and record patient-reported outcomes. The study examined changes in CAT and SABA inhaler use in COPD to determine whether passively collected SABA and CAT scores changed in a parallel manner.

Methods: Patients with self-reported COPD enrolled in a digital health program, which provided EMMs to track SABA and maintenance inhaler use, and a companion smartphone application ("app") to provide medication feedback and reminders. Patients completing the CAT questionnaire in the app at enrollment and at 6 months were included in the analysis. Changes in CAT burden category [by the minimally important difference (MID)] and changes in EMM-recorded mean SABA inhaler use per day were quantified at baseline and 6 months.

Results: The analysis included 611 patients. At 6 months, mean CAT improved by −0.9 (95% CI: −1.4, −0.4; $p < 0.001$) points, and mean SABA use decreased by −0.6 (−0.8, −0.4; $p < 0.001$) puffs/day. Among patients with higher burden (CAT ≥ 21) at enrollment, CAT improved by −2.0 (−2.6, −1.4; $p < 0.001$) points, and SABA use decreased by −0.8 (−1.1, −0.6; $p < 0.001$) puffs/day.

Conclusion: Significant and parallel improvement in CAT scores and SABA use at 6 months were noted among patients enrolled in a digital health program, with greater improvement for patients with higher disease burden.

Keywords: chronic obstructive pulmonary disease, COPD assessment test, digital health, medication adherence, short-acting beta-agonist, electronic medication monitoring

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a progressive respiratory illness with substantial impact on the patient's well-being (1). The COPD Assessment Test (CAT) is a validated questionnaire designed to assess the disease burden (2). While the CAT is widely accepted in clinical practice, administration is not performed at any routine cadence, limiting its clinical value to identify patients with declining disease status potentially needing intervention.

Remote patient monitoring (RPM) with digital health tools for COPD may enhance the current standard of care through regular monitoring of symptoms and medication-taking behaviors (3, 4). Today, there exists an abundance of digital tools to support RPM, including wearables, smartphone or mobile phone applications, short-messaging services (SMS), and sensors to track medication use. These tools may help collect data between office visits and provide regular insight into patient health.

In COPD, digital tools may also help enhance care through recording use of as-needed inhaled short-acting beta-agonists (SABA) for symptoms, monitoring adherence to daily maintenance inhaler medication, and measuring lung function. Electronic monitoring of SABA use may help identify patients who have acute worsening of symptoms outside of routine provider visits (5–7). Although SABA use is not captured in CAT, symptoms that are commonly treated with SABA are a major component of the questionnaire.

This study assessed changes in CAT scores and SABA use over 6 months among patients with COPD enrolled in a digital self-management platform, which included electronic medication monitors (EMMs) and a smartphone application ("app").

METHODS

Patients with self-reported COPD enrolled in a digital self-management platform (Propeller Health, Madison WI) between August 2017 and December 2019. Those with an EMM-compatible SABA inhaler and smartphone were eligible for the study. Patients were instructed to attach a small EMM to their SABA inhaler(s), and if available, to their maintenance inhaler(s) (**Figure 1**).

EMMs had an expected battery life of 12–18 months and did not require charging during the study period (8). The EMMs passively monitored the date and time of inhaler actuations when the EMM was depressed (6). These data were transmitted wirelessly via Bluetooth synchronizations ("syncs") to the patient's smartphone app. Patients received evidence-based education, feedback on medication use, and reminders for scheduled medications and were prompted to complete an in-app CAT questionnaire monthly (2). Patients had the option to share their information with their providers but were not required to do so. To participate, patients were required to accept Propeller's Terms of Service which specified use of de-identified, aggregated data for publication (9). This retrospective analysis was reviewed by the Copernicus Institutional Review Board (PRH1-18-132).

Patients completing the CAT within 2 weeks from the date of the first EMM sync (baseline) and again at 6 months (152–212 days) from the date of the first sync were studied. CAT score burden was characterized as low (0–10), medium (11–20), high (21–30), and very high according to standard categories (31–40) (10). The first week of syncing was considered a platform learning period, and thus inhaler use during this period was excluded. Mean daily SABA use (puffs/day) and daily maintenance adherence (percent of puffs taken/prescribed, capped at 100%) were calculated during the 30 days following the baseline CAT and 30 days prior to the 6-month CAT. At baseline, unadjusted comparisons between mean daily SABA use and mean daily maintenance inhaler adherence by CAT burden were evaluated using a Wilcoxon rank-sum test (11). From baseline to 6 months, minimally important differences of CAT (2 points) and changes in CAT burden categories (low: CAT 0–10, medium: CAT 11–20, high: CAT 21–30, and very high: CAT 31–40) (2) were evaluated using the Chi-square test (12). Changes in CAT, mean daily SABA use, and mean daily maintenance adherence were estimated from baseline to 6 months using linear mixed-effect models (13) accounting for within-patient variability and adjusting calendar month to account for potential seasonal variation. Analyses were then stratified by lower (CAT < 21) vs. higher (CAT ≥ 21) burden categories to ensure a sufficient sample size. Because patients served as their own control, adjustment for static individual-level characteristics (e.g., age, gender) was not necessary. However, we did conduct sensitivity analyses to adjust for differences by age, CAT score, mean daily SABA use, and mean daily maintenance adherence (10). All statistical tests were two-tailed with an $\alpha = 0.05$ threshold for statistical significance. All analyses were conducted in R version 3.4 (R Foundation for Statistical Computing).

FIGURE 1 | A small electronic medication monitor (EMM) attaches to a metered dose inhaler (MDI) to collect data on inhaler usage. Data is then transmitted wirelessly via Bluetooth to a paired patient-facing smartphone application.

RESULTS

The analysis included 611 patients {mean [standard deviation (SD)] age: 62 (8) years; 64% were >60 years}. All patients had a SABA inhaler EMM, and 371 (60.7%) also had a maintenance inhaler EMM. At baseline, mean (SD) CAT score was 22.6 (7.8). Mean (SD) SABA use was 2.3 (3.1) puffs/day, and mean (SD) daily adherence was 81 (27.1)%. Among patients with higher CAT burden, 60% were ≥60 years of age compared to patients with lower CAT burden where 71% were ≥60 years of age. Baseline SABA use was greater among patients with higher burden CAT scores compared to patients with lower burden CAT scores [median (IQR): 1.4 (0.4, 3.9) vs. 0.8 (0.1, 2.4) puffs/day; Wilcoxon rank-sum test $p < 0.001$], while baseline maintenance adherence was consistent [median (IQR): 93.3 (72.5, 98.3)% vs. 93.3 (80.7, 100.0)%; Wilcoxon rank-sum test $p = 0.08$].

At 6 months, 277 (45%) patients had CAT scores that improved by the minimally important difference (14). Using the MID as a minimum threshold, 154 (25%) patients moved from a higher burden category to lower burden category, while 351 (57.4%) patients had no category change. A larger percentage of patients in the higher burden group moved to a lower burden category or improved their CAT by the MID compared to those in the lower burden group (34.7 vs. 9.9%, Chi-square $p < 0.001$) (**Table 1**).

Linear mixed-effect models found that, from baseline to 6 months, mean CAT improved by −0.9 (95% CI: −1.4, −0.4; $p < 0.001$) points, mean SABA use decreased by −0.6 (95% CI: −0.8, −0.4; $p < 0.001$) puffs/day, and mean adherence decreased by −4.0 (95% CI: −6.9, −1.2; $p < 0.01$) percent. Among patients with higher CAT burden scores at enrollment, CAT improved by −2.0 (95% CI: −2.6, −1.4; $p < 0.001$) points, and SABA use decreased by −0.8 (95% CI: −1.1, −0.6; $p < 0.001$) puffs/day (**Table 2**). Sensitivity analyses did not change the observed outcomes significantly (**Supplementary Table 1**).

TABLE 1 | Change in CAT burden category[a] and MID change over 6 months using a Chi-square test.

	Overall	Higher burden	Lower burden	*P*-value**
Category improvement and MID decrease, *n* (%)	154 (25.2)	131 (34.7)	23 (9.9)	<0.001
Category worsening and MID increase, n (%)	106 (17.3)	40 (10.6)	66 (28.3)	
No change, *n* (%)	351 (57.4)	207 (54.8)	144 (61.8)	

[a]*Burden categories were defined as low (CAT 0–10), medium (CAT 11–20), high (CAT 21–30), and very high (CAT 31–40) burden.*

***P-value derived from a Chi-square test comparing CAT burden category change in patients with higher vs. lower burden COPD.*

DISCUSSION

This study examines 6-month changes in CAT scores, SABA use, and maintenance adherence among patients using a digital self-management platform for COPD. Patients with higher COPD burden (CAT ≥ 21) had greater improvement with a significant and parallel reduction in both CAT and EMM-measured SABA use at 6 months. Moreover, many patients had a clinically meaningful change in CAT score. The concurrent reductions in CAT scores and SABA use suggest that passively collected SABA data may serve to highlight patients at risk of increasing COPD burden. Adherence to maintenance inhalers at baseline was high (80%) with only a 4% decrease over 6 months and not reflective of changes in CAT.

Digital technologies have demonstrated value in the clinical setting for patients with COPD (6, 7, 15). Alshabani (7) found that integration of an EMM with feedback in a 12-month quality improvement study resulted in earlier identification of patients with worsening condition, and significant reductions in SABA use. Similar outcomes were observed in 190 Medicare-eligible

TABLE 2 | Change in CAT, SABA use and adherence over 6 months using linear mixed effects models.

	Mean (SD) at baseline	Mean (SD) at 6 months	Estimates*	Lower 95% CI*	Upper 95% CI*	*P*-value
Overall, *n* = 611						
CAT	22.6 (7.8)	21.7 (8.1)	−0.9	−1.4	−0.4	$p < 0.001$
SABA use, puffs/day	2.3 (3.1)	1.6 (2.5)	−0.6	−0.8	−0.4	$p < 0.001$
Adherence, %	81.1 (27.1)	76.8 (29.8)	−4.0	−6.9	−1.2	$p = 0.01$
High burden, *n* = 378						
CAT	27.4 (4.7)	25.4 (6.6)	−2.0	−2.6	−1.4	$p < 0.001$
SABA use, puffs/day	2.7 (3.4)	1.8 (2.5)	−0.8	−1.1	−0.6	$p < 0.001$
Adherence, %	79.6 (27.0)	74.5 (30.3)	−4.6	−8.4	−0.9	$p = 0.01$
Low burden, *n* = 233						
CAT	14.7 (4.8)	15.6 (6.5)	0.9	0.1	1.7	$p = 0.03$
SABA use, puffs/day	1.7 (2.4)	1.3 (2.3)	−0.3	−0.6	−0.1	$p = 0.01$
Adherence, %	82.8 (27.2)	80.0 (28.8)	−3.5	−8.2	1.2	$p = 0.14$

**Adjusted for enrollment month.*

patients with COPD (6) where SABA use halved following a 6-month digital health intervention, which included patient outreach from clinical staff. Electronic monitoring of SABA use has also been identified as a potential predictor of exacerbations (5, 16) and higher SABA use has been associated with greater CAT burden (10).

Continued research is needed to better understand the mechanisms of behavior change and identify the most efficacious digital modalities. Programs integrating and/or based on behavior theories typically demonstrate stronger efficacy than those that do not (17). Further, programs relying on digital tracking alone may be less effective due to high attrition rates (18). Hybrid programs integrating both digital and human touch, as well programs including a combination of digital modalities (e.g., EMM plus smartphone app), have demonstrated stronger outcomes.

This study supports the use of a digital self-management platform to complement standard of care, but there were limitations. First, patients self-enrolled and were possibly more motivated to modify their behavior and health. Second, patients self-reported their COPD diagnosis, which was not confirmed given the remote design. Moreover, we did not have access to patient characteristics like gender and disease stage, which may limit the generalizability of the results observed. Lastly, while improved CAT and SABA use were observed over the study period, modest decreases in adherence were also detected. We hypothesize that improvement in COPD burden may have led patients to ease use of their maintenance inhaler; however, further investigation is needed to understand these behavioral changes.

CONCLUSION

Patients enrolled in a digital self-management platform for COPD demonstrated significant and clinically meaningful improvement in CAT and a concurrent reduction in SABA use, especially among those with higher COPD burden. This study highlights the benefit of digital tools like EMMs to identify associations not observed with pre-digital modalities (self-report, prescription fills, etc.). However, EMMs are one tool in a myriad of consumer and medical-grade digital tools today. Future studies in COPD should explore the added value of digital platforms and/or tools that complement patient care by improving patient–provider communication, and better understanding chronic disease management.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the medication use data that supported this study are not publicly available because they are considered Protected Health Information (PHI) under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) in the US, and as such are only accessible under specific authorization of access following HIPAA guidelines. Requests to access the datasets should be directed to leanne.kaye@resmed.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Copernicus Institutional Review Board (PRH1-18-132). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Analyses were performed by RG. All authors contributed to the conceptualization and design of the manuscript, data interpretation, and manuscript preparation and revision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.624261/full#supplementary-material

REFERENCES

1. CDC. *Basics About COPD - Chronic Obstructive Pulmonary Disease (COPD)*. (2019). Available online at: https://www.cdc.gov/copd/basics-about.html (accessed July 6, 2020).
2. Jones PW, Harding G, Berry P, Wiklund I, Chen W-H, Kline Leidy N. Development and first validation of the COPD assessment test. *Eur Respir J*. (2009) 34:648–54. doi: 10.1183/09031936.00102509
3. Ding H, Fatehi F, Maiorana A, Bashi N, Hu W, Edwards I. Digital health for COPD care: the current state of play. *J Thorac Dis*. (2019) 11:S2210–20. doi: 10.21037/jtd.2019.10.17
4. Attaway AH, Alshabani K, Bender B, Hatipoglu US. The utility of electronic inhaler monitoring in COPD management. *Chest*. (2020) 157:1466–77. doi: 10.1016/j.chest.2019.12.034
5. Sumino K, Locke ER, Magzamen S, Gylys-Colwell I, Humblet O, Nguyen HQ, et al. Use of a remote inhaler monitoring device to measure change in inhaler use with chronic obstructive pulmonary disease exacerbations. *J Aerosol Med Pulm Drug Deliv*. (2018) 31:191–8. doi: 10.1089/jamp.2017.1383
6. Chen J, Kaye L, Tuffli M, Barrett MA, Jones-Ford S, Shenouda T, et al. Passive monitoring of short-acting beta-agonist use via digital platform in patients with chronic obstructive pulmonary disease: quality improvement retrospective analysis. *JMIR Form Res*. (2019) 3:e13286. doi: 10.2196/13286
7. Alshabani K, Attaway AA, Smith MJ, Majumdar U, Rice R, Han X, et al. Electronic inhaler monitoring and healthcare utilization in chronic obstructive pulmonary disease. *J Telemed Telecare*. (2020) 26:495–503. doi: 10.1177/1357633X19850404
8. Mosnaim GS, Stempel DA, Gonzalez C, Adams B, BenIsrael-Olive N, Gondalia R, Ket al. The Impact of Patient Self-Monitoring Via Electronic Medication Monitor and Mobile App Plus Remote Clinician Feedback on Adherence to Inhaled Corticosteroids: A Randomized Controlled Trial. *J Allergy Clin Immunol Pract*. (2020). doi: 10.1016/j.jaip.2020.10.064. [Epub ahead of print].
9. *Propeller*. Available online at: https://my.propellerhealth.com/terms-of-service (accessed July 9, 2020).
10. Gondalia R, Bender BG, Theye B, Stempel DA. Higher short-acting beta-agonist use is associated with greater COPD burden. *Respir Med*. (2019) 158:110–3. doi: 10.1016/j.rmed.2019.10.007
11. Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. New York, NY: John Wiley & Sons. (1973). p. 68–75.

12. Agresti A. *An Introduction to Categorical Data Analysis*. 2nd ed. New York, NY: John Wiley & Sons. (2007). p. 38. doi: 10.1002/0470114754
13. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* (2015) 67:1–48. doi: 10.18637/jss.v067.i01
14. Kon SSC, Canavan JL, Jones SE, Nolan CM, Clark AL, Dickson MJ, et al. Minimum clinically important difference for the COPD assessment test: a prospective analysis. *Lancet Respir Med.* (2014) 2:195–203. doi: 10.1016/S2213-2600(14)70001-3
15. North M, Bourne S, Green B, Chauhan AJ, Brown T, Winter J, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: the RESCUE trial. *Npj Digit Med.* (2020) 3:1–8. doi: 10.1038/s41746-020-00347-7
16. Jenkins CR, Postma DS, Anzueto AR, Make BJ, Peterson S, Eriksson G, et al. Reliever salbutamol use as a measure of exacerbation risk in chronic obstructive pulmonary disease. *BMC Pulm Med.* (2015) 15:97. doi: 10.1186/s12890-015-0077-0
17. Lycett HJ, Raebel EM, Wildman EK, Guitart J, Kenny T, Sherlock JP, et al. Theory-based digital interventions to improve asthma self-management outcomes: systematic review. *J Med Intern Res.* (2018) 20:e293. doi: 10.2196/jmir.9666
18. Chan YF, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. the asthma mobile health study, a large-scale clinical observational study using research kit. *Nat Biotechnol.* (2017) 35:354–62. doi: 10.1038/nbt.3826

Conflict of Interest: LK, RG, and MB are employees of ResMed. DS and MW are employees of Propeller Health, an affiliate of ResMed. All authors receive compensation and stock as part of their employment.

The authors declare that this study received funding from Propeller Health, a subsidiary of ResMed. The funder had the following involvement in the study: study design, collection, analysis, interpretation of data, the writing of this article and the decision to submit it for publication.

frontiers in Digital Health

STUDY PROTOCOL
published: 26 July 2021
doi: 10.3389/fdgth.2021.708159

A Protocol Study to Establish Psychological Outcomes From the Use of Wearables for Health and Fitness Monitoring

Frans Folkvord[1,2]*, *Amy van Breugel*[1], *Sanneke de Haan*[3], *Marcella de Wolf*[3], *Marjolein de Boer*[3] *and Mariek Vanden Abeele*[1,4]

[1] Tilburg School of Humanities and Digital Sciences, Communication and Information Science, Tilburg University, Tilburg, Netherlands, [2] Open Evidence Research, Barcelona, Spain, [3] Tilburg School of Humanities and Digital Sciences, Culture Studies, Tilburg University, Tilburg, Netherlands, [4] imec-mict-UGent, Department of Communication Sciences, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:
Mohamed Elgendi,
University of British Columbia, Canada

Reviewed by:
Hamid Reza Marateb,
Universitat Politecnica de Catalunya, Spain
Colin K. Drummond,
Case Western Reserve University, United States
Yongbo Liang,
Guilin University of Electronic Technology, China

***Correspondence:**
Frans Folkvord
fransfolkvord@gmail.com

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

Received: 11 May 2021
Accepted: 29 June 2021
Published: 26 July 2021

Citation:
Folkvord F, van Breugel A, de Haan S, de Wolf M, de Boer M and Abeele MV (2021) A Protocol Study to Establish Psychological Outcomes From the Use of Wearables for Health and Fitness Monitoring. Front. Digit. Health 3:708159. doi: 10.3389/fdgth.2021.708159

Background: The last few decades people have increasingly started to use technological tools for health and activity monitoring, such as tracking apps and wearables. The main assumption is that these tools are effective in reinforcing self-empowerment because they support better-informed lifestyle decision-making. However, experimental research assessing the effectiveness of the technological tools on such psychological outcomes is limited.

Methods and Design: Three studies will be conducted. First, we will perform a systematic review to examine the experimental evidence on the effects of self-tracking apps on psychological outcome measurements. Second, we will conduct a longitudinal field experiment with a between subject design. Participants ($N = 150$) begin a 50-day exercise program, either with or without the aid of the self-tracking app Strava. Among those who use Strava, we vary between those who use all features and those who use a limited set of features. Participants complete questionnaires at baseline, at 10, 25, and 50 days, and provide details on what information has been tracked via the platform. Third, a subset of participants is interviewed to acquire additional qualitative data. The study will provide a rich set of data, enabling triangulation, and contextualization of the findings.

Discussion: People increasingly engage in self-tracking whereby they use technological tools for health and activity monitoring, although the effects are still unknown. Considering the mixed results of the existing evidence, it is difficult to draw firm conclusions, showing more research is needed to develop a comprehensive understanding.

Trial registration: Netherlands Trial registration: NL9402, received on 20 April 2021; https://www.trialregister.nl/trial/9402.

Keywords: self-tracking, exercise, body empowerment, sports, psychological outcomes, self-determination theory, psychological effects of health tracking technologies

INTRODUCTION

People increasingly use technological tools for health and activity monitoring (1–4). These activity tracking apps and wearables are considered self-empowering because they can help users make better-informed lifestyle decisions based on their data (5, 6). Recent research, however, suggests that self-tracking technology use may lead to bodily *alienation* rather than empowerment, because it could encourage users to trust technology more than what their own body tells them (7). Because these technologies make explicit all sorts of bodily experiences and processes that would otherwise remain unnoticed (heart rate, burned calories, etcetera), people might develop a more objectified and distant instrumental view on their body as something that needs to be monitored and corrected—at the cost of a tacit, smooth reliance on it (5).

To date, however, there is no empirical research, at least to our knowledge, that examines the mechanisms leading to either bodily empowerment or alienation. This protocol describes the uses of different methodological approaches that we will conduct in a project to investigate under which conditions the use of a self-tracking app fosters bodily empowerment or bodily alienation. More specific, the main objective that this project will study is whether and under which conditions the use of self-tracking technologies empowers or alienates people while doing physical activity.

Collecting health related data with technological tools and platforms is referred to as "self-tracking for health" (5, 7). Self-tracking for health involves a self-monitoring process that relies on the quantification of bodily processes (e.g., heart rate, calories burned) and activities (e.g., step counts, types of sports). People have been self-monitoring their body and life for self-improvement since ancient times (8), although with the introduction of digital technologies, users can (1) collect personal and health data in real-time and at a much larger scale, (2) gain information on parameters that are difficult—if not impossible—to track otherwise, (3) obtain personalized feedback and gamified targets derived from processing of collected data (7), and (4) link their data to other data streams, such as social networking platforms, enabling them to compare with others (4).

There is a widespread belief among technology developers, health professionals and scholars that self-tracking technologies can empower users to make healthier life choices (7–9). For example, following Deci and Ryan's (10). Self-Determination Theory, it is argued that self-tracking technologies fulfill the three needs for empowerment: (1) autonomy, (2) competence, and (3) belongingness. Self-tracking technologies allow users to regulate their physical activity without having to rely on external parties (autonomy), help them reach personal targets and compare their performance with others (competence), and generate belongingness through online endorsements and commenting features (11).

Given their assumed empowering potential, one might expect spectacular usage rates for self-tracking technologies. Recent intervention studies, however, report significant numbers of drop out (11–14), with over 50% of new users dropping out in <2 weeks (6). An explanation for this drop-out might be that some users experience *bodily alienation* rather than bodily empowerment (15). This study adopts a multi-method approach to examine if the use of self-tracking technologies lead to bodily alienation rather than bodily empowerment, and whether these in turn predict future use of self-tracking technology as well as future exercise behavior.

Classical phenomenological theories (16–18) shed light on this possibility. They emphasize that in our normal interactions with the world, our bodies function "transparently": we are typically not aware of our bodies functioning, such as our hearts beating or our feet moving. Scholars argue that paying conscious attention to these tacit bodily processes may have a disruptive, alienating effect on individuals, because such "hyperreflectivity" (16) disturbs our effortless, smooth flow of acting, like paying attention to your feet disrupts your dancing moves. In reflectively engaging with our bodies, then, we may come to see our bodies as objects or instruments to be tweaked and altered, rather than functioning in the background of our everyday existence.

METHODS AND ANALYSIS

In total, three separate studies will be conducted in this project, see for a detailed timeline and phases of the study in **Figure 1**. First, we will perform a systematic review to examine the experimental evidence on the effects of self-tracking apps on psychological outcome measurements linked to bodily empowerment and alienation to have a state of the art understanding in this area. Second, we will conduct a longitudinal field experiment with a between subject design. Third, a subset of participants is interviewed to acquire additional qualitative data. The study will provide a rich set of data, enabling triangulation, and contextualization of the findings to provide an answer to our research questions. We will provide some more detailed information on the three separate studies here.

Study 1: Systematic Review

First, to have a state-of-the-art understanding of current existing causal evidence we will conduct a systematic review focusing on experimental studies. As a result of the systematic review we will develop detailed hypotheses that we will test in Study 2: An experimental study and Study 3: Interviews. Participants will be blinded to the hypotheses. We will investigate the following bibliographic databases: PubMed, PsychInfo, and ISI Web of Science, whereby we will restrict our search to peer-reviewed papers or dissertations written in English in the last 10 years (>2010). Next, we will do a manual search in selected articles' reference lists to make sure we do not miss any relevant studies. Studies will be considered eligible if they; (1) report experimental effects of the usage of self-tracking technologies on psychological outcomes, (2) include self-tracking among healthy people and people of any gender and age diagnosed with any type of (chronical) disease, and (3) use self-tracking technologies on one of the following devices for self-tracking; mobile (smart)phones and applications, and/or wearables. Studies will be excluded if they are published before 2010, are not peer-reviewed, are written in a non-English language and have a non-experimental design. The primary outcomes from this systematic review will be used

	STUDY PERIOD						
	Enrolment	Allocation	Post-allocation			Close-out	
TIMEPOINT	$-t_1$	0	t_1	t_2	t_3	t_4	t_x
ENROLMENT:							
Eligibility screen	X						
Informed consent	X						
Pilot Study	X						
Allocation		X					
INTERVENTIONS:							
Control			←	—	—	→	
Self-tracking			←	—	—	→	
ASSESSMENTS:							
Gender			X				
Age			X				
BMI			X				
Primary outcomes							
Motivation for PA			X	X	X	X	
Psychological need satisfaction			X	X	X	X	
Secondary outcomes							
Self-reported PA			X	X	X	X	
Engagement in PA			X	X	X	X	
Enjoyment of PA			X	X	X	X	
Bodily alienation/ -empowerment			X	X	X	X	
Physical body experience			X	X	X	X	
Objectified body consciousness			X	X	X	X	
Self-objectification			X	X	X	X	
Wearable technology trust			X	X	X	X	
Wearable technology Motivation				X	X	X	
Trust in Strava						X	

FIGURE 1 | Enrolment, interventions, and assessment in the study period.

to develop the exact experimental design. For the systematic review, we will follow the PRISMA standard guidelines whereby we provide a detailed explanation to improve transparency. In addition, we will present a flow diagram with all the selection steps of the process and the exclusion criteria.

Study 2: Experimental Study

Second, we will conduct a longitudinal (50 days) field experiment with a between subject design, aiming for 150 participants that we will collect based on convenience sampling at the Tilburg University campus, see also Recruitment section. We will pre-register our trial (www.trialregister.nl) before we start collecting data and we will include the CONSORT and SPIRIT-statements in our reporting of the outcomes of this study. In this study, we will create two groups that we will randomly allocate to using Strava or not for their physical activity tracking, randomly assigned based on enrolment. An eligibility criterion to participate in the study is that participants must not have used self-tracking technologies for at least 2 months before starting the experiment. Participants complete questionnaires at baseline, after 10 days, 25 days, and a final one after 50 days. We will ask participants to provide details on what information has been tracked via the platform. Next, a subset of participants is interviewed (Study 3).

Power analyses (G*Power) (19) with an F-test (MANCOVA), suggest that a sample size of $N = 148$ is expected to be sufficient to detect significant ($\alpha = 0.05$), medium effects ($f = 0.25$) of conditions (power = 0.80). We have based the medium effect size on the study of Seo et al. (20). Considering that some participants might drop out before the last measurement after 50 days, the plan is to include more than 150 participants in the study. We will conduct Multivariate analysis of (Co-)Variance to analyse the primary and secondary outcomes of the experiment.

In the experiment we will repeatedly collect different information (see also **Figure 2**), for example the primary outcomes the motivation to conduct physical activities and

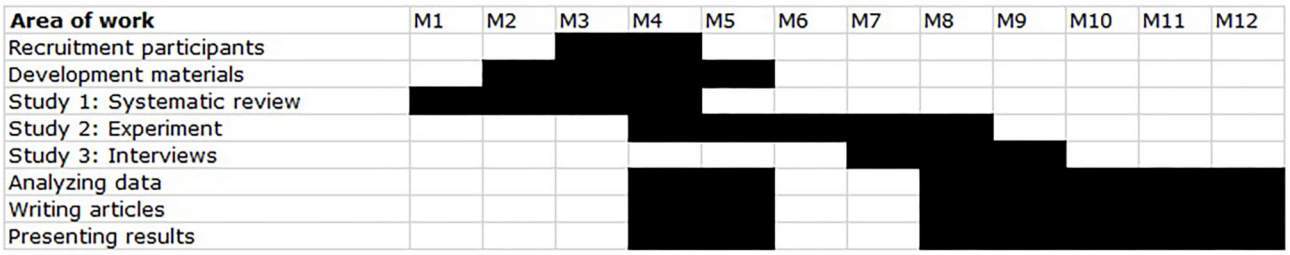


FIGURE 2 | Timetable.

psychological need satisfaction, and the secondary outcomes self-reported physical activity, engagement in physical activity, enjoyment of physical activity, experiences of bodily alienation-empowerment, physical body experience, objectified body consciousness, self-objectification, wearable technology trust, wearable technology motivation, and trust in Strava. We will only use validated scales to assess these factors, and will conduct Factor Analysis and Reliability Analyses to make sure the scales we will use are valid and reliable in our dataset as well. All functionalities within Strava, using GPS for recording your physical activities, connecting to people to share recordings of physical activity, comment, or like physical activities of others or see reactions of others on your recordings, share and detect routes, trainings, races, and challenges with your network, connect to a (local) club or community, share photo's, and track and receive feedback on level of fitness and progress, will be part of the study. No specific assessment is considered to be more relevant than other information because the current study will mainly be conducted to test the differences between participants who use a self-tracking app compared to participants who do not use a self-tracking app. In future studies more attention could be given to different functionalities and see the effects of these on psychological or physiological outcomes.

Study 3: Interviews

Third, we will conduct an in-depth qualitative interview study. For this study, we will invite 10–15 participants to take part in this study. Based on the literature on in-depth qualitative interviewing (19), this number is large enough to allow for saturation of the themes that come up in the interviews while still allowing heterogeneity in findings, also in the event some interviewees will drop out. Recruitment will take place by selecting a subset of participants from the experimental study (study 2) who have volunteered to take part in the interview study. To ensure a heterogeneous sample for our exploratory study, selection criteria include a representation of various gender-identities (the questionnaire provides options for "male," "female," and "prefer not to say") and various levels of user intensity. By using a thematic list for interviewing, including themes such as experiences of health and well-being, exercise routine, and use of the communicative aspect of the Strava app, we provide an overview of the underexplored phenomenon of bodily empowerment and alienation in using Strava, while also providing sufficient focus into bodily experiences of self-trackers.

To analyze the data acquired through this interview study, we will use Interpretative Phenomenological Analysis. This method aims at clarifying how people make sense of their bodies within a larger context (21), and as such, it enables us to identify patterns of bodily alienation and empowerment in the data on using the self-tracking app Strava.

Recruitment

We will recruit participants at the Tilburg University Campus through online participation platforms for scientific research. Inclusion criterium is that participants should not use a self-tracker already for their physical activities and are between 18 and 35 years old. Participant outside this age range and who already use a self-tracker to record physical activities will be excluded from the study. Participants will receive credits for participating in the study.

Ethical Approval and Informed Consents

The ethical committee of the Tilburg School of Humanities and Digital Sciences has provided ethical approval for the current study (reference is REDC 2020.198). Explicit written consent to participate will be collected from all participants.

Confidentiality

Participants will participate in the survey study with a participant number, making them pseudonymous in the survey data, so that they are not directly identifiable in the survey data. There will be a key file (necessary, among others, to assign credits). This key file linking participant number to their name will be stored digitally in a password protected folder. This file will be destroyed once the data have been collected. Identifiable information of the participants participating in the interview study will be pseudonymized when transcribing the data. The recordings made of the interviews will be destroyed immediately after the transcription of each interview. After completion of the study, we will provide all participants with a written debriefing to fully inform them about the study details. We will inform participants to the maximum extent on the usage of the Strava app by providing them with the privacy policy of Strava and asking for them to sign when they have read it and that they comply. In addition, active consent will be collected from all participants.

Data Management

Considering that a large amount of data will be collected among participants, a specific data management plan and strict protocol has been written in close collaboration with the data steward of the host institute (Tilburg School of Humanities and Digital Sciences, the Netherlands). The host institute is very strict in the treatment of the collected data with the highest level of confidentiality to assure good management of data, which has been tested and evaluated during the ethical approval process.

Data Monitoring and Auditing

The host institute is very strict in the treatment of the collected data with the highest level of confidentiality to assure good management of data, which has been tested and evaluated during the ethical approval process.

Risk and Benefits to Participants

We do not consider any risks for the participants in the study. Since we will conduct the study with a student population, the benefits of participating in the study for participants is that they receive credits.

Knowledge Translation and Dissemination

Considering both the societal and academic relevance of the study, we will proactively communicate and disseminate the outcomes of the proposed studies within this protocol as much as possible. For example, we will aim to publish the outcomes of the study in (open access) scientific international journals and present the results on (inter)national symposiums and conferences. Next, we will seek opportunities to publish the outcomes in non-scientific outlets and present the results at public events as well, in order to disseminate the outcomes of the current project to a wider audience.

DISCUSSION

This study describes the protocol to systematically test whether and under which conditions the use of self-tracking technologies empowers or alienates, using a multi-method approach. This question thus far remains unexplored.

Digital technologies provide opportunities to the ever more detailed measurement and monitoring of people's activities, bodies, and behaviors in real time. In addition, they offer possibilities for data archiving related to physical activity and for sharing data with peers, directly, or through the social network of users. Because peers can immediately see and react on someone's physical activity that was tracked, the perceived locus of causality, the extent to which individuals perceive their actions as a result of their external or internal reasons, could become externalized. Without using a self-tracker, an individual might be motivated to conduct physical activity because this person likes the activity in itself, while when using a self-tracker, an individual might be motivated to conduct physical activity because of the reactions this person will receive from the personal community afterwards. Via mobile digital devices, for instance, users are continuously connected to the internet, which enables them to immediately inform peers about physical activities and performances. Devices and wearables are typically fitted with digital sensors that assess not only activity and achievements, but also personal and health related information. Such technologies are considered to be a major source of potential revenue for digital developers and entrepreneurs, but empirical evidence showing positive effects remains scarce.

Considering the increasing use and popularity of digital technologies, using these technologies on a day-to-day basis raises important questions concerning their specific psychological effects, and how these may in turn impact future technology use and exercise behavior. Several studies show the effect of self-tracking technologies on endurance (22), well-being and self-determination to conduct physical activity (23), belongingness (20), and sense of competence (23). Bodily empowerment and alienation, however, appear areas that have thus far been largely overlooked.

Generalizability

Although the proposed experiment and interviews focus on the effects of using Strava, we believe the results of the current study will be generalizable to more self-trackers that have similar functionalities. An increasing number of self-tracker apps and programs are being developed that are able to tracking workouts and provide statistics of these workouts, discover, and connect to local clubs, events and challenges, and interact with personal communities. On the other hand, considering we will mainly focus our study on a student population, we believe the outcomes might be different for older (or younger) target groups because a student population might be more susceptible to social influences than older target groups, or might be more vulnerable to external factors influencing their bodily perceptions.

STRENGTH AND LIMITATIONS

One of the strengths of the current study is that until now there have been limited studies examining the effects of using self-trackers on bodily perceptions. To our knowledge, this is the first systematic study that will examine the effects of using self-trackers. Robust scientific evidence is needed, in particular from large-scale Randomized Clinical Trials, with large sample sizes, longitudinal data collection, and predetermined outcome measurements that provide a comprehensive understanding of the usage of self-trackers. Considering the enormous societal success of self-tracking apps, it is essential to improve our understanding and be able to provide evidence-based recommendations about how these technologies influence people, in particular in the long-term.

On the hand, the current study might face some limitation and biases, such as selection bias, given that we will include mostly students that are willing to participate in a study of at least 50 days. Moreover, participants will receive credits for participation in the study, which might be a different motivation than other people who do not receive any rewarding incentive for using the self-tracker. Based on the interviews we will gather motivational aspects of using a self-tracker that we cannot achieve with only an experimental study. In addition, Strava might not be representative for

all self-trackers, considering that it is a very popular app with several functionalities that other self-tracking apps do not have.

AUTHOR CONTRIBUTIONS

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

FUNDING

The current study is financed by an internal funding scheme of Tilburg School of Humanities and Digital Sciences. The funding body will not have any role during the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Contact person is: Marjet van Loo, m.j.vanloo@tilburguniversity.edu.

REFERENCES

1. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *J Med Internet Res.* (2017) 19:e125. doi: 10.2196/jmir.5604
2. Henriksen A, Sand AS, Deraas T, Grimsgaard S, Hartvigsen G, Hopstock L. Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach. *BMC Public Health.* (2020) 20:1300. doi: 10.1186/s12889-020-09406-w
3. Kay M, Santos J, Takane M. *mHealth: New horizons for Health Through Mobile Technologies*. Geneva: World Health Organization (2011).
4. Stragier J, Abeele MV, Mechant P, De Marez L. Understanding persistence in the use of online fitness communities: comparing novice and experienced users. *Comput Hum Behav.* (2016) 64:34–42. doi: 10.1016/j.chb.2016.06.013
5. Sharon T. Self-tracking for health and the quantified self: Re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. *Philos. Technol.* (2017) 30:93–121. doi: 10.1007/s13347-016-0215-5
6. Duus R, Cooray M, Page NC. Exploring human-tech hybridity at the intersection of extended cognition and distributed agency: a focus on self-tracking devices. *Front Psychol.* (2018) 9:1432. doi: 10.3389/fpsyg.2018.01432
7. Lupton D. Self-tracking, health and medicine. *Health Sociol. Rev.* (2017) 26:1. doi: 10.1080/14461242.2016.1228149
8. Swan M. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data.* (2013) 1:85–99. doi: 10.1089/big.2012.0002
9. Maturo A, Setiffi F. The gamification of risk: how health apps foster self-confidence and why this is not enough. *Health Risk Soc.* (2016) 17:477–94. doi: 10.1080/13698575.2015.1136599
10. Deci EL, Ryan RM. Self-determination theory. In: Van Lange PAM, Kruglanski AW, Higgins ET, editors. *Handbook of Theories of Social Psychology*, Vols. 1 and 2. Thousand Oaks, CA: Sage Publications Ltd. (2011). p. 416–36. doi: 10.4135/9781446249215.n21
11. Stragier J, Vanden Abeele M, De Marez L. Recreational athletes' running motivations as predictors of their use of online fitness community features. *Behav Inform Technol.* (2018) 37:815–27. doi: 10.1080/0144929X.2018.1484516
12. Anastasiadou D, Folkvord F, Serrano-Troncoso E, Lupiañez-Villanueva F. Mobile health adoption in mental health: user experience of a mobile health app for patients with an eating disorder. *JMIR mHealth uHealth*. (2019). 7:e12920. doi: 10.2196/12920
13. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. *Lancet Psychiatry.* (2015) 2:942–8. doi: 10.1016/S2215-0366(15)00268-0
14. Endeavour Partners. *Inside Wearbles: How the Science of Human Behavior Change Offers to Secret to Long-Term Engagement*. Wandina WA: Endeavour Partners (2014). p. 17.
15. Fuchs T. The psychopathology of hyperreflexivity. *J Specul Philos.* (2010) 24:239–55. doi: 10.1353/jsp.2010.0010
16. Leder D. *The Absent Body*. Chicago, IL: University of Chicago Press (1990).
17. Merleau-Ponty M. *Phenomenology of Perception. Transl by C. Smith*. London: Routledge (1945/2002). doi: 10.4324/9780203994610
18. Faul F, Erdfelder E, Lang AG, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* (2007) 39:175–91. doi: 10.3758/BF03193146
19. Smith JA, Flowers P, Larkin M. *Interpretative Phenomenological Analysis. Theory, Method and Research*. London: Sage (2009).
20. Seo MW, Kim Y, Jung HC, Kim JH, Lee JM. Does online social connectivity promote physical activity in a wearable tracker-based intervention? A pilot randomized controlled study. *Sustainability.* (2020) 12:8803. doi: 10.3390/su12218803
21. Gittus M, Fuller-Tyszkiewicz M, Brown HE, Richardson B, Fassnacht DB, Lennard GR, et al. Are Fitbits implicated in body image concerns and disordered eating in women? *Health Psychol.* (2020) 39:900–4. doi: 10.1037/hea0000881
22. Stiglbauer B, Weber S, Batinic B. Does your health really benefit from using a self tracking device? Evidence from a longitudinal randomized control trial. *Comput Human Behav.* (2019) 94:131–9. doi: 10.1016/j.chb.2019.01.018
23. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. *J Med Internet Res.* (2015) 17:e174. doi: 10.2196/jmir.4086

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 September 2021
doi: 10.3389/fdgth.2021.630273

A Real-Time Wearable System for Monitoring Vital Signs of COVID-19 Patients in a Hospital Setting

Mauro D. Santos[1†], Cristian Roman[1†], Marco A. F. Pimentel[1], Sarah Vollam[2], Carlos Areia[2], Louise Young[2], Peter Watkinson[2] and Lionel Tarassenko[1]*

[1] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom,
[2] Critical Care Research Group, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:
Mohamed Elgendi,
University of British Columbia, Canada

Reviewed by:
Yan Wang,
The University of Tokyo, Japan
Colin K. Drummond,
Case Western Reserve University, United States
Maria Vittoria Bulgheroni,
Ab.Acus, Italy

***Correspondence:**
Mauro D. Santos
mauro.dsantos@gmail.com

[†]*These authors have contributed equally to this work*

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

Received: *17 November 2020*
Accepted: *16 August 2021*
Published: *07 September 2021*

Citation:
Santos MD, Roman C, Pimentel MAF, Vollam S, Areia C, Young L, Watkinson P and Tarassenko L (2021) A Real-Time Wearable System for Monitoring Vital Signs of COVID-19 Patients in a Hospital Setting. Front. Digit. Health 3:630273. doi: 10.3389/fdgth.2021.630273

The challenges presented by the Coronavirus disease 2019 (COVID-19) pandemic to the National Health Service (NHS) in the United Kingdom (UK) led to a rapid adaptation of infection disease protocols in-hospital. In this paper we report on the optimisation of our wearable ambulatory monitoring system (AMS) to monitor COVID-19 patients on isolation wards. A wearable chest patch (VitalPatch®, VitalConnect, United States of America, USA) and finger-worn pulse oximeter (WristOx2® 3150, Nonin, USA) were used to estimate and transmit continuous Heart Rate (HR), Respiratory Rate (RR), and peripheral blood Oxygen Saturation (SpO_2) data from ambulatory patients on these isolation wards to nurse bays remote from these patients, with a view to minimising the risk of infection for nursing staff. Our virtual High-Dependency Unit (vHDU) system used a secure web-based architecture and protocols (HTTPS and encrypted WebSockets) to transmit the vital-sign data in real time from wireless Android tablet devices, operating as patient data collection devices by the bedside in the isolation rooms, into the clinician dashboard interface available remotely *via* any modern web-browser. Fault-tolerant software strategies were used to reconnect the wearables automatically, avoiding the need for nurses to enter the isolation ward to re-set the patient monitoring equipment. The remote dashboard also displayed the vital-sign observations recorded by the nurses, using a separate electronic observation system, allowing them to review both sources of vital-sign data in one integrated chart. System usage was found to follow the trend of the number of local COVID-19 infections during the first wave of the pandemic in the UK (March to June 2020), with almost half of the patients on the isolation ward monitored with wearables during the peak of hospital admissions in the local area. Patients were monitored for a median of 31.5 [8.8, 75.4] hours, representing 88.1 [62.5, 94.5]% of the median time they were registered in the system. This indicates the system was being used in the isolation ward during this period. An updated version of the system has now also been used throughout the second and third waves of the pandemic in the UK.

Keywords: electronic track & trigger, COVID-19, wearable devices, isolation wards, vital signs, continuous monitoring, e-obs, remote patient monitoring

INTRODUCTION

The Coronavirus disease 2019 (COVID-19) was declared a global health emergency by the World Health Organisation (1) at the beginning of March 2020. In its early stages, this pandemic presented several challenges for in-hospital patient care in the United Kingdom (UK): the fear of nosocomial infections in clinical environments, the lack of knowledge about the dynamics of virus transmission and initial shortages of Personal Protective Equipment (PPE). These rapidly led to a reduction in hospital admissions for non-COVID-19 disease, e.g., for cancer treatment (2), and the need for rapid adaptation of existing infectious disease protocols across the hospital (3, 4).

While the most severely ill COVID-19 patients were admitted to the Intensive Care Unit (ICU), those who did not meet ICU admission criteria were placed under observation on isolation wards (5). Our research group, a collaboration between clinical staff and biomedical engineers, was tasked by hospital management in Oxford at the end of February 2020 with supplying them with the most appropriate vital-sign monitoring system for these isolation-ward patients. Six priority requirements for the system were established (by order of appearance in the manuscript):

1. Because COVID-19 is primarily a disease which affects the cardio-respiratory system, it was decided that the three most important physiological parameters to monitor, ideally on a continuous basis, should be peripheral blood Oxygen Saturation (SpO_2), Respiratory Rate (RR), and Heart Rate (HR).
2. The patients should not be confined to bed but should ideally be ambulatory, an important factor in the recovery from respiratory disease, for those patients stepping down from Intensive Care.
3. The patients were to be remotely monitored within the hospital, in the sense that they would be in individual rooms on the infection wards, with the nursing staff caring for them situated in another location nearby (the "nurse bays"), on the same hospital floor.
4. Any additional continuous monitoring from wearables should be fully integrated with the periodic nurse observations of the full set of vital signs, comprising SpO_2, RR, HR as well as Blood Pressure (BP), Temperature (Temp), level of consciousness and the corresponding Early Warning Score (EWS) (6), in use throughout UK hospitals. The totality of the patient information (continuous data from wearables and periodic nurse observations) should be made available on real-time displays on a central station in the nurse bays, away from the isolation ward.
5. The amount of contact between the infected patients and the nursing staff was to be minimised, which meant that the frequency of nurse observations, which required the use of PPE, could not be increased, even though it was already known that the COVID-19 disease could lead to *rapid* patient deterioration.
6. The system should work within the hospital cyber-security infrastructure, compliant with patient confidentiality standards.

It soon became clear to us that the wearable ambulatory monitoring system (AMS) which we had been developing to monitor high-risk patients continuously on general wards, and create a virtual High-Dependency Unit (vHDU), could be adapted to meet the above six requirements.

In the next section, we review state-of-the-art wearable ambulatory monitoring systems, before providing an overview of our vHDU system, based on commercial off-the-shelf components, indicating how we had previously assessed their wearability, accuracy, and reliability. We then describe how this system was optimised to ensure that it met all of the above six requirements and deployed to monitor patients infected with the COVID-19 virus during the first wave of the pandemic in the UK, from mid-March to June 2020. Finally, we discuss preliminary results of the system usage in the isolation ward during this period.

WEARABLE AMBULATORY MONITORING SYSTEMS

State-of-the-art wearable AMS present a combination of mechanical (e.g., accelerometer), physiological (e.g., Electrocardiogram, ECG), and biochemical sensors (e.g., glucose monitors) (7) and include adhesive patch, clothing, chest-strap, upper-arm band, wristband, and finger-worn monitors (8). ECG, HR, RR, SpO_2, BP, Temp, and patient activity are the clinical parameters most commonly tracked by current AMS deployed for in-hospital monitoring (9).

The wireless technologies found in these wearable AMS range from Wi-Fi and Bluetooth-Low-Energy (BLE) to cellular and Radio-Frequency (RF) technologies. Data are usually transmitted from the wearable(s) to an intermediate wireless client (e.g., a tablet or smartphone, in a 1-to-1 routing configuration) and then to an intranet or cloud server *via* Wi-Fi routers. Alternatively, several wearables can transmit data to a single access point (N-to-1 routing configuration), which then relays them to servers *via* a wired network. Data from several patients are processed for display and are typically reviewed by clinicians on workstations or ward screens (often wall-mounted) and mobile applications, in parallel and in real-time.

Examples of certified AMS with published validation or feasibility studies in the hospital setting include (10):

- The Vista Solution™ (VitalConnect, United States of America, USA), which uses a disposable 5-day battery chest-patch to collect ECG, HR, RR, body Temp, and activity data into a client tablet *via* BLE (1-to-1 routing). The tablet, in turn, retransmits the data over Wi-Fi into the Vista Solution™ cloud platform. BP and SpO_2 can also be collected *via* third-party BLE devices.
- The ViSi Mobile System (Sotera Wireless, USA), which uses one Wi-Fi enabled wrist-worn module, connecting to one finger-probe pulse oximeter (for SpO_2), to one 3- or 5-lead

ECG chest module (for ECG/HR/RR/skin Temp/activity) and to one upper-arm cuff module (for BP). The module transmits the data to an intranet server, *via* the hospital Wi-Fi routers (N-to-1 routing).

- The Sensium® System (Sensium Healthcare Ltd, UK), which uses a re-usable RF chest patch (disposable electrodes), connected to an axilla probe, to measure HR, RR and body Temp. A proprietary RF wireless router collects data from several patches (N-to-1 routing) and retransmits them to a central server *via* the hospital Wi-Fi (or wired connexion).

For all three examples, the centralised server/cloud data are then made available to remote clinicians *via* a web-, desktop application- or mobile app- based dashboards, on which it is possible to review numerical data such as HR, RR, BP, referred to also as "numerics" thereafter, waveforms such as the ECG, and alert/notifications. Other in-hospital AMS broadly follow similar architectures.

To date there have been no randomised controlled trials with wearable based AMS that show significant clinical benefit/cost-effectiveness, and clinical studies are still needed before these systems can be adopted in large-scale clinical practise (10).

THE OXFORD VIRTUAL HDU AMBULATORY MONITORING SYSTEM

Background Work

When we originally set out to identify suitable wearable devices for our vHDU system, we focused on devices which also provide the raw waveforms, i.e., the Photoplethysmogram (PPG) and the ECG *via* wireless-transmission mode (e.g., BLE). The availability of waveforms, not only enables further biomedical signal processing research, but also allow clinical staff to review these waveforms on bedside monitors to confirm the correct application of the sensors to the patients.

In a first study, we assessed the wearability of a selection of commercially-available wearables for monitoring the vital signs of ambulatory patients (11). Our study, which used a prospective observational cohort design, was reviewed and approved by the Oxford University Research and Ethics Committee and Clinical Trials and Research Governance teams (R55430/RE003). Participants in the study were required to wear up to four different AMS for up to 72 h each to mimic in-hospital use.

Next, a clinical accuracy study was conducted, in which up to 33 healthy participants undertook six different motion tasks followed by an hypoxia exposure phase (study approved by the East of Scotland Research Ethics Service REC 2 (19/ES/0008), study number ISRCTN61535692). During exposure to hypoxia, participants were desaturated under controlled conditions to seven SpO_2 targets: {100, 95, 90, 87, 85, 83, 80}%, controlled *via* an oxygen mask (overseen by an anaesthetist), whilst wearing four wearable pulse oximeters, one standard Philips MX450 monitor pulse oximeter, one adhesive chest patch, and were fitted with an arterial line. The protocol of this study can be found in Areia et al. (12).

The accuracy of the pulse oximeter devices was assessed by comparing their SpO_2 and Pulse Rate (PR) estimates with simultaneous gold-standard arterial blood oxygen saturation (SaO_2), measured from arterial blood gas sampled *via* the arterial line, and with ECG-derived HR, respectively (we make the assumption that PR and HR can be used interchangeably). The accuracy of the chest patch was evaluated by comparing its RR and HR estimates with simultaneous RR and HR estimates derived from the reference capnography and the ECG. Full results for both of these evaluations can be found in Santos et al. (13) and Morgado et al. (14), respectively.

As a result of our wearability and accuracy evaluations, the WristOx® 3150 OEM BLE (Nonin Medical Inc., USA) (15) finger-based pulse oximeter, named Nonin hereafter, and the VitalPatch® (VitalConnect, USA) (16) adhesive chest-patch were ultimately selected as our wearable devices. From the Nonin, we collect the PR, SpO_2 (both at a sampling rate of 1 Hz) and near-infrared PPG waveform (at a sampling rate of 75 Hz). From the VitalPatch®, we collect the HR and RR (both at a sampling rate of 0.25 Hz), patient posture (e.g., standing, sitting, lying down, etc.), number of steps (at a sampling rate of 1 Hz) and the single-lead ECG and 3-axis accelerometer waveforms (at sampling rates of 125 and 62.5 Hz, respectively). Both devices compute signal-quality indices and display battery status. BLE technology allows these wearable devices to stream the combination of HR, RR, and SpO_2 data from active patients into a BLE-enabled tablet, continuously for ∼48 h and 5 days, respectively (meeting the first and second requirements from section Introduction).

While Nonin's BLE protocol documentation was provided by the vendor, so that it could be implemented by a third-party, VitalConnect only provided their Software Development Kit. The latter limits the VitalPatch® device to connect to only one Android Operating System (OS) tablet at a time (1-to-1 routing). As a result, BLE and Wi-Fi enabled Android tablets were the only choices for the patient data collection device required in our system use case. Wi-Fi technology allows the vital-sign data recorded with the wearable devices data to be transmitted from the Android tablets into remote servers *via* the hospital's Wi-Fi routers. The Samsung Tab A 2016 (SM-T585) and 2019 (SM-T515), 10-inch, models were selected for data collection (see **Figure 1**).

Patient Data Collection App

Figure 2 illustrates the data-flow diagram of the vHDU system. The Android Java app was the first component of the system to be designed. Our 1-patient to 1-tablet data routing approach presented a cost limitation, but also an opportunity: given that each tablet's computational resources are available for each patient, the first design decision was that as much of the patient wearable data pre-processing as possible would occur on the tablet app (rather than on the back-end server). Clinical staff from the COVID-19 wards were interested in reviewing remotely: (a) the patient vital-sign data and the connexion and battery status of the devices *in real-time*; and (b) the vital-sign times series trend *retrospectively*. Whilst (a) requires the processing of high-rate data, (b) requires that the data are summarised so as to make it feasible to both store it and allow the clinical staff to browse hourly to daily vital-sign trends (low-rate

FIGURE 1 | Hardware used to monitor COVID-19 patients on the isolation wards (from left to right): the Nonin finger-worn pulse oximeter, the Samsung Tab A 2019, 10-inch, with front and back protective casing, and the VitalPatch®.

vital-sign data). The pre-processing strategies applied to the data collected in the app are described next.

Raw-Data Collection

To allow the detailed analysis and evaluation of the system retrospectively, all the numerical and waveform (raw) data from the devices were recorded in comma-separated values (csv) files on the tablet. Each 5 min, a batch of up to 10 files were compressed in ZIP format (a lossless data compression approach) and then sent to the back-end server (all files sent successfully to the server are then subsequently deleted from the app after 24 h). The raw csv files are recorded up to a maximum of 30 MB of data, corresponding to ZIP files of about 1.5 MB. We note that for each hour of patient monitoring the app creates about 7.5 MB of compressed files. The app's raw-data upload to the back-end is represented by datastream #3 in **Figure 2**. The corresponding web-service is described in section Web-Application Back-End.

High-Rate Data Collection

The synchronisation of the high-rate vital-sign data, and the connexion and battery status from the wearable and tablet devices, were implemented using the ReactiveX library (17). The latter uses a reactive programming paradigm, in which each sensor data (i.e., patch, pulse oximeter, and tablet data) is accumulated in its respective "Observable" variable, asynchronously, and for a given time window. For each variable and time window, either the last available data point or a summary statistic was determined (the median was used in our case, as it determines the most representative value in a given window, being less influenced by motion artefact than the mean). All the variables' simultaneous window values were then combined into a single (synchronised) data structure, using an appropriate "datastream operator" (see the "Zip" operator, from the ReactiveX library).

The following high-rate numerics datastreams (illustrated as datastream #1 and datastream #2 in **Figure 2**, respectively, and detailed in **Table 1**) were pre-processed on the tablet (as described) and then sent to the back-end:

i. for each 2-s window, the most recent available numerics (HR/PR, RR, SpO_2, number of steps and patient posture);
ii. for each 10-s window, high-rate numerical estimates, i.e., median HR/PR, RR, and SpO_2, the most recent number of steps and patient posture, and the wearable/tablet device status data (i.e., the available storage space, battery status,

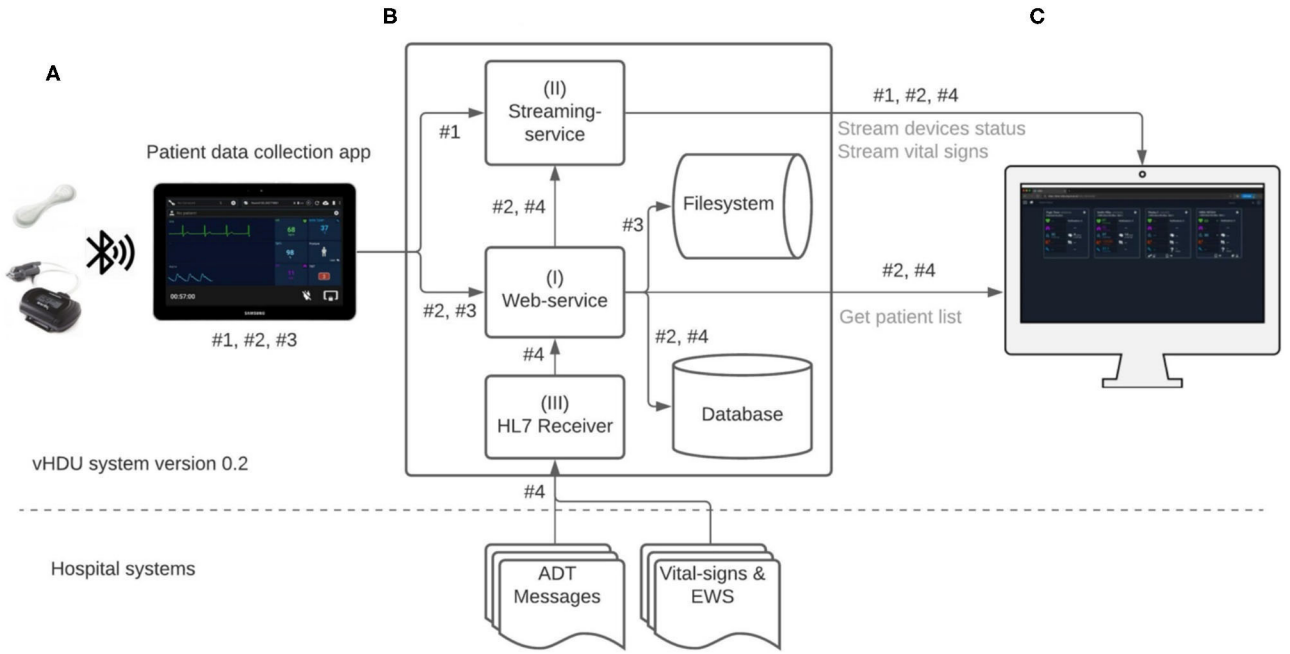


FIGURE 2 | Virtual High-Dependency Unit (vHDU) system version 0.2 data-flow diagram. Components: **(A)** the Patient data collection tablet app, which retransmits the wearables data to the back-end; **(B)** the Web-application back-end, which stores and relays that information to **(C)** the Clinician Dashboard client, which displays the data from remote locations (*via* a web-browser). The main app datastreams of this system are: #1 - high-rate vital-sign values sent from the app to the streaming-service, each 2 s; #2 - high-rate 10-s median vital-sign values and the most recent wearable status values, sent each 10 s to the web-service, and low-rate 5-min vital-sign median values, sent each 5 min to the web-service; #3 - compressed files (ZIP format) with the wearables raw-data, sent to the web-service; #4 - HL7 (Health Level Seven standard) messages with e-obs chart data and patient Admission-Discharge-Transfers (ADT) information, which are sent from the hospital systems into a HL7 receiver application on the vHDU back-end. Data transmission from both the VitalPatch® and Nonin devices to the tablet app is encrypted using the Advanced Encryption Standard (AES) algorithm, over Bluetooth-Low-Energy. Data communication between the tablet app and the back-end web- and streaming-services, and between the latter and the Clinician Dashboard is encrypted using the Secure Sockets Layer (SSL) algorithm over the HTTPS and the secure Websockets protocols (for the web- and streaming-services, respectively). EWS, (local) Early Warning Score.

BLE and Wi-Fi connexion status), pre-processed into a (10-s) "data-packet".

The 2-s data-rate requirement came from feedback from staff on the Cardiology wards (where our vHDU system was also to be deployed), who need to track the vital signs of cardiac patients in (near) real time. At such a high rate, it is possible to replicate the real-time view from the Android tablet, located at the patient bedside, on remote interfaces. This was also felt to be an important feature for monitoring COVID-19 patients on the isolation wards as it was known that the virus could cause very rapid deterioration of a patient's cardio-respiratory system. Similarly, clinical research and ward staff confirmed that updating the device connexion and battery status every 10 s would be sufficient. We note that the high-rate vital-sign data are not filtered (e.g., for motion artefacts) as the objective was to display remotely real-time data from the devices.

Low-Rate Data Collection

To allow browsing large periods of vital-sign data from a remote client, the research team decided to limit the resolution of the low-rate data to 5-min windows. The low-rate data are observed retrospectively and may be significantly influenced by motion artefact. Therefore, to remove as many as possible of the motion artefacts and ensure that each data channel buffer would not consume too much memory, the vital-sign data was first summarised for each 10-s window and subsequently summarised further for each 5-min window, using the median estimator in both cases. Five-minute median values of HR/PR, RR, and SpO_2, and the most recent number of steps and patient posture in those windows, were therefore estimated from the (non-overlapping) 10-s medians aggregated from datastream described in item (ii), section Raw-Data Collection, and finally sent to the back-end server.

The 10-s and 5-min datastreams were recorded both in the app and the back-end databases, generating ∼200 KB of data per patient monitoring hour (both detailed in **Table 1** and illustrated as datastream #2, in **Figure 2**, as the second is derived from the first on the app, and then sent and stored on the same back-end configuration, see also **Table 2**). This is negligible when compared with the amount of raw-data recorded by the app. Finally, the app database is cleared every time a patient is disconnected from the vHDU system by the clinical staff.

TABLE 1 | Technical specification of the datastreams sent from the app or from the e-obs hospital system into the vHDU system version 0.2.

Datastreams[a]	**#1**		**#2**	**#3**	**#4**
Protocol	Socket.io	HTTPS	HTTPS	HTTPS	MLLP
Data source	Wearables	Wearables	Wearables	Wearables	e-obs
Data frequency	2 s	10 s	5 min	5 min	∼ 4 h
Data format	JSON	JSON	JSON	ZIP	HL7 v2.6
Data window size[b]	2 s	10 s	5 min	variable (up to 30 MB)	
Synchronised	✓	✓	✓[c]		✓
On clinician dashboard	✓	✓[d]	✓		✓
Wearable Status					
Connexion		Most recent			
Battery		Most recent			
Vital Signs					
Heart rate	Most recent	Median	Median	Raw	✓
Respiratory rate	Most recent	Median	Median	Raw	✓
SpO_2	Most recent	Median	Median	Raw	✓
Blood pressure					✓
Temperature					✓
Oxygen therapy					✓
AVPU/GCS					✓
Steps	Most recent	Most recent	Most recent	Raw	
Posture	Most recent	Most recent	Most recent	Raw	
Waveforms					
Electrocardiogram				Raw	
Photoplethysmography				Raw	
Acceleration (3-axis)				Raw	

"most recent" - the latest data point available in that window is used. "median" - the median value is determined over a given window. "raw" - all the (raw) data from the wearables are collected.

[a] *The datastreams numbers match those used and illustrated in* ***Figure 2****.*

[b] *The data windows do not overlap.*

[c] *The 5-min medians are computed from the 10-s median values, which are synchronised as detailed in section High-Rate Data Collection.*

[d] *The vital signs from the 10-s data-packet are only visible in the Clinician Dashboard in case the 2-s data are absent.*

AVPU, Alert, Voice, Pain, Unresponsive scale; GCS, Glasgow Coma Scale; e-obs, hospital electronic observation system used by nurses to enter vital-sign observations; MLLP, Minimal Lower Layer Protocol; JSON, JavaScript Object Notation.

App Interface

Figure 3 shows exemplar test data displayed in the patient data collection app interface. It was designed to mimic the functionality of bedside patient monitors. Once the tablet registers with the web-application back-end, clinical staff can enrol a patient in the AMS *via* their wristband Medical Record Number (MRN) and location (i.e., ward, bay and bed, when the last two are available). A monitoring session is then created on the back-end, and data from one VitalPatch® and one Nonin devices can be linked and collected *via* the app. From top to bottom in **Figure 3**: the VitalPatch® and Nonin connexion status and battery (in hours) are displayed on the top left and right corners, respectively; the patient MRN and location are displayed next; the waveform-grid (ECG in green and PPG in blue) and the numerics-grid (HR, RR, SpO_2, skin temperature—removed when deployed at the isolation wards as it was not felt by the clinical staff to be a reliable indicator -, number of steps and patient posture) can be observed in the middle left and right of the display, respectively.

Web-Application Back-End

The vHDU system back-end (**Figure 2B**) was developed to receive and store the patient data collected by the app and make them available to client interfaces on remote locations. It consists of three main services (referenced in **Figure 2B**, using the same numerals):

I The web-service, which uses the CakePHP framework Model-View-Control (MVC) software architecture and the Representational State Transfer (REST) approach for its Application Programming Interfaces (APIs). This web-service receives the 10-s and the 5-min app data, both illustrated as datastream #2, in **Figure 2**, as both are further processed on the back-end and stored on a relational database (PostgreSQL). While the first is relayed to the clinician dashboard *via* the streaming-service, the second is made available to the same client interface *via* the REST API (e.g., when a particular patient chart is reviewed). The raw-data compressed files are also received by the web-service, each 5 min, and stored on the back-end filesystem (illustrated by

TABLE 2 | Technical specification of the system deployed in the isolation ward.

Software	Framework (language)	Server 1	Server 2
Patient data collection app	Android v28 (Java v8)	-	-
Streaming-service[a]	ExpressJS; Socket.io (Javascript)	Node.js v12	
Web-service[b]	CakePHP v3 (PHP v7.3)	IIS v10	PostgreSQL v12
File upload web-service[c]	CakePHP v3 (PHP v7.3)	-	IIS v10; PostgreSQL v12
HL7 receiver	Apache Camel v2; Spring Framwork v4 (Java v8)	Apache Karaf v4	PostgreSQL v12
Clinician dashboard	ReactJS v16 (Javascript)	IIS v10	-

Servers 1 and 2 are both virtual machines with 8 GB of RAM, 4 CPUs each and the main application containers running the vHDU system components are identified for each server.

[a] *Streaming-service configuration to relay datastream #1, and the 10-s data-packets from datastream #2, to the Clinician Dashboard.*

[b] *Web-service configuration to receive and process datastream #2 in* ***Figure 2****, which includes both the 10-s and 5-min data-packets processed in Server 1 and stored in Server 2.*

[c] *Web-service configuration to receive and process datastream #3 in* ***Figure 2****, related with the raw-data files (ZIP format) that are uploaded and stored in Server 2 each 5 min.*

IIS, Internet Information Services; HL7, Health Level Seven standard.

the datastream #3, in **Figure 2**). Secure access to the web-services resources is maintained *via* the use of Java Web Tokens (JWT).

II The streaming-service: the Node.js framework was used to develop the real-time communication of the 2-s high-rate numerical data from the app to the remote interfaces. This real-time communication is illustrated as datastream #1 in **Figure 2**, being present in the arrow from the app component to the streaming-service, and then in the arrow from the latter to the Clinician Dashboard, in that figure. Note that the 2-s data are not stored on the back-end. Additionally, the 10-s data stored in back-end, are also relayed from the web-service (I) to the remote interfaces *via* the streaming-service. This real-time communication is also represented in **Figure 2**, first in the arrow from the web-service to the streaming-service, and then in the arrow from the latter to the Clinician Dashboard, in that figure. The socket.io protocol is used for the real-time, bi-directional communication. JWT is also used to authenticate the initial socket.io handshake request.

III The HL7 receiver: the development of this component was based on prior work from the research group in consuming HL7 (Health Level Seven standard) messages from the hospital systems (18). Apache Java Camel is used to receive (*via* Minimal Lower Layer Protocol, MLLP), store and process the HL7 messages from the hospital's Admission-Discharge-Transfers (ADT) system and from the hospital system responsible for the electronic notification and documentation of vital-sign observations (e-obs) by clinical staff (19), also known locally as the electronic Track-and-Trigger (e-T&T) chart. Finally, a Java Spring REST API makes these data available to the web-service back-end, so they can be stored alongside the continuous wearable data transmitted by the tablet app. The data-flow of this component is illustrated as datastream #4 in **Figure 2**.

Clinician Dashboard

The Clinician Dashboard (**Figure 2C**) was designed to allow the review of both the real-time vital-sign and device status data, from multiple patients in parallel, and the 5-min vital-sign trend charts, for each patient, on the remote ward. The ReactJS (JavaScript) framework was used to develop an interface that could display both the high- and low-rate datastreams *via* a web-browser (usually available from any hospital computer), which includes:

i The staff accounts (secured *via* username and password credentials) and patient details administration pages. Staff accounts are pre-configured with a set of hospital wards from which they are able to access data.

ii A homepage, which allows authorised clinical staff accounts to browse a list of patient cards from their pre-configured set of hospital wards. The web-service APIs allow this list to be queried, whilst the socket.io protocol allows the streaming of the device status data every 10 s, and of the real-time vital-sign values, every 2 s, to each patient card. For each patient card, the most recent values from the continuous datastream (HR, RR, SpO2, number of steps, and patient posture) are displayed alongside the most recent complementary nurse observations, collected *via* the HL7 receiver (BP, Temp and level of consciousness). Temp, the core body temperature, is recorded by the nurses during their observations using a tympanic thermometer, when they also assess the patient's level of consciousness using the Alert, Verbal, Pain, Unresponsive (AVPU) Scale, or the Glasgow Coma Scale (GCS).

iii An *augmented e-obs* chart, with the full set of vital-sign data for each patient, combining the (typically) 4-hourly vital-sign observations from the nurse, and corresponding (local) Early Warning Score, with the 5-min median vital-sign values (HR, RR, SpO_2) from the wearable AMS (the wearable data are not scored at this point).

Figure 4 shows the implemented Clinician Dashboard homepage, which allows authorised staff accounts to review the active patient monitoring sessions from a computer screen located at the nurse bay (the combination of the vHDU system components described up to this point fulfils the third requirement from section Introduction). Each card corresponds to one patient monitoring session with data transmitted by the patient data collection app. The patient is identified at the top by their first/last names, MRN number and location (ward short code/bay/bed). Real-time wearable data are displayed alongside the complementary intermittent vital-sign observation data.

The vital-sign timing is shown below its value in case the values are older than a minute. The wireless connectivity of the wearable devices and their battery status are shown at the bottom

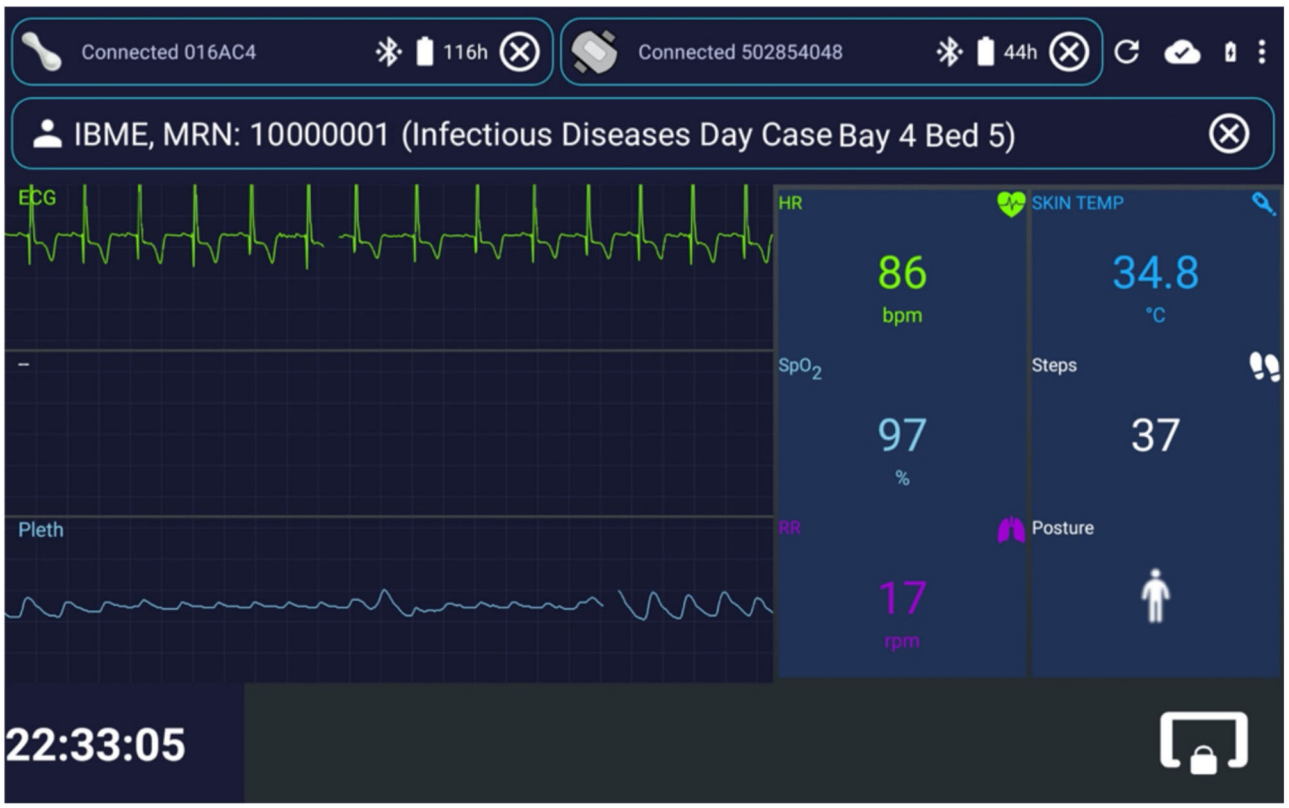


FIGURE 3 | Interface for the patient data collection app, implemented in Android Java. Exemplar test data are shown. The status of the wearable device can be observed at the top of the image, with both the VitalPatch® and the Nonin devices in "Connected" state, and with serial numbers "016AC4" and "502854048," respectively. The top right icons enable (from left to right): restarting the app; checking the connexion status with the back-end server; checking the charging and battery status; and entering the app settings. The patient can be unregistered by pressing the outlined "×" icon on the right of the patient information field. The live vital-sign and waveform data (ECG, Electrocardiogram, and Pleth, Photoplethismography) values are displayed on the right side and left side of the app, respectively. The lower right icon locks the app screen, leaving only the lower left clock displayed. In this case, the monitoring and streaming of the vital-sign data occur in the background, likely to be a less distracting mode for patients (particularly useful at night, because of the light emitted by the tablet). IBME, Institute of Biomedical Engineering; MRN, Medical Record Number; bpm, beats per minute; rpm, respirations per minute; HR, Heart Rate; RR, Respiratory Rate; TEMP, Temperature; SpO_2, peripheral blood Oxygen Saturation.

of each patient card. Finally, the patient card list can be filtered by searching for their details, from the search box, at the top right corner of the page header, or filtering by starred patients (note the yellow/white stars at the top right corner of each card).

Figure 5 shows the augmented e-obs chart for an exemplar COVID-19 patient, presenting the vital-sign observation sets recorded by the nursing staff alongside the 5-min vital-sign data estimates acquired using the wearables (meeting the fourth requirement from section Introduction). These two different data sources can be identified on the chart by the absence or presence of the BLE symbol at the top of the temperature row, respectively. Note that the column of the nurse-recorded observation set, at 18:04, is selected and the respective Temp, BP, HR, RR, SpO_2, oxygen therapy, AVPU, and (local) EWS values recorded at that time, are shown in the rightmost column.

Figure 6 shows another COVID-19 patient data example in which a 5-min median values, determined from the wearable data, are selected, at 19:12 (a data-packet with only HR and SpO_2 from the Nonin, in this case). Displaying the wearable data interleaved with the nurse observations facilitates the review of the patient physiology trend by staff. Wearable data estimates can be collapsed to show only the (most recent 5-min) estimates at 15-min, 30-min or hourly intervals, to make it possible to review longer time-series more easily, when needed.

Fault-Tolerant Software Strategies for Continuous Monitoring

To optimise the wearable AMS prior to patient use, several rounds of testing and troubleshooting were performed in 2019, with the ward staff wearing the devices during working hours. This allowed for software cheques, connectivity assessment, and integration of the wearable AMS into the clinical environment. The testing was also intended to encourage AMS familiarisation amongst ward staff.

Reliability of the AMS was an important feature, as it is critical that the nurses do not spend time investigating problems caused

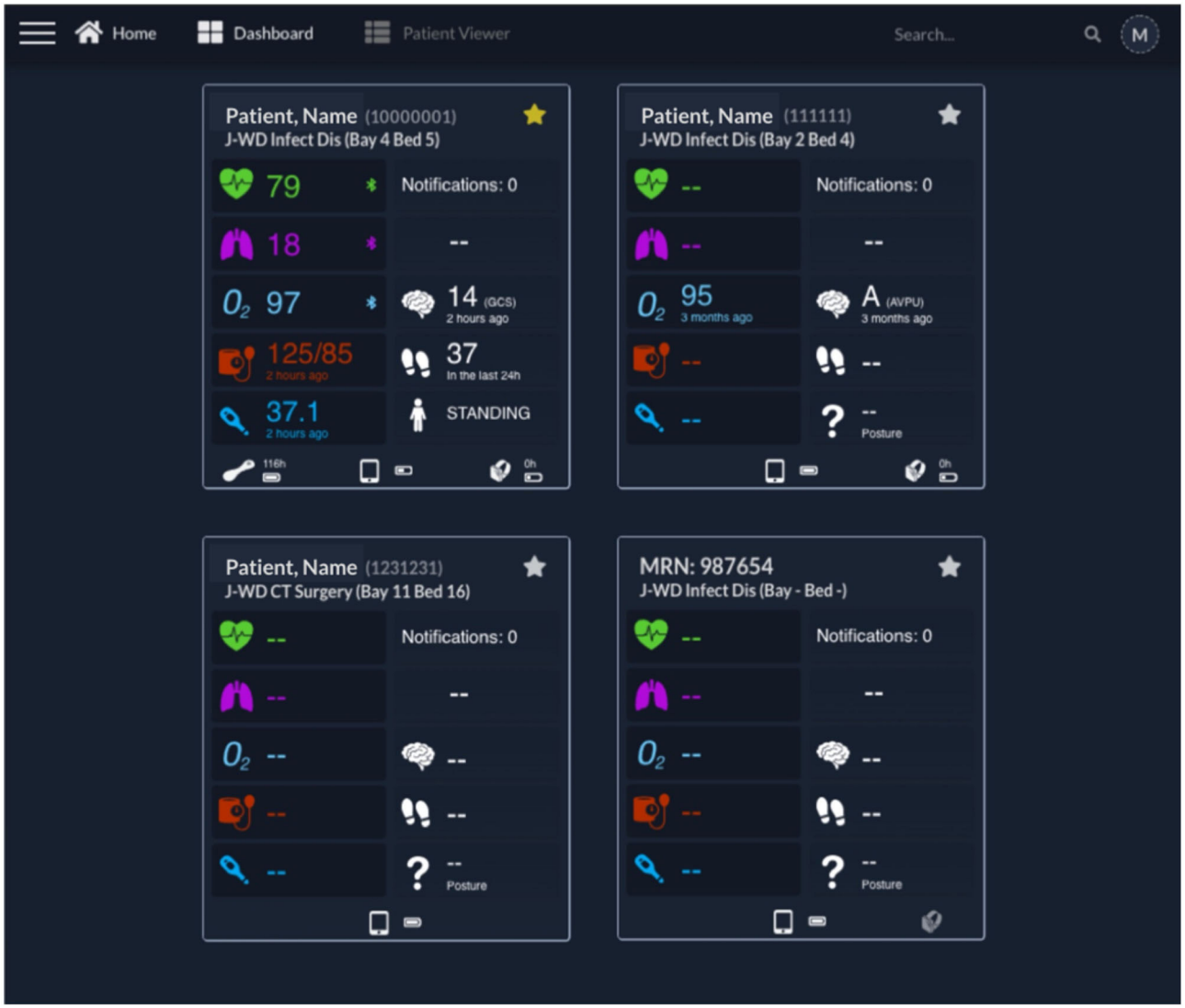


FIGURE 4 | Clinician Dashboard homepage with four exemplar test patient cards. Each card presents on the left: the Heart Rate, Respiratory Rate, peripheral blood Oxygen Saturation ("O_2"), Blood Pressure, and Temperature; and on the right: the Notifications (disabled at this time), continuous Early Warning Score (disabled at this time), Level of consciousness, Steps and Posture. The patient name, Medical Record Number and location are shown at the top of each card. At the bottom of each card, the status of the wearables, including connectivity and battery status, can be observed (from left to right, the VitalPatch®, tablet, and Nonin icons are shown when present—such as in the first card). It can be noted that the upper left patient is "Starred" (by the active yellow star), with a VitalPatch® device connected, in range and with a remaining battery of 118 h; the tablet has a battery charged at 50% of capacity; and the Nonin device is connected to the tablet but has no battery left (0 h). The battery icon flashes in this case, to prompt for a change of batteries. The upper right patient had their last observation 3 months ago, showcasing the system's ability to keep the last patient state until they are discharged from the system (the Nonin also has no battery left in this case). The lower left patient has been pre-registered but the wearable devices have not yet been attached to the patient, as indicated by the lack of icons for the wearables, and thus no vital-sign data are available yet. The lower right patient has no "Patient, Name," being identified only by the MRN in this case. The right-side Nonin icon is also greyed, representative of a Nonin pulse oximeter associated with the patient but out of range from the patient's assigned vHDU tablet. AVPU, Alert, Voice, Pain, Unresponsive scale; GCS, Glasgow Coma Scale; J-WD, John Radcliffe Hospital Ward Code; MRN, Medical Record Number.

by loss of sensor data, as described in the fifth requirement from section Introduction. The following fault-tolerant software functionalities were implemented in the app to ensure its reliability for continuous monitoring of the patients' vital signs:

- a single app lockdown mechanism built into the Android OS, the "kiosk mode," which "pins" the vHDU app to the foreground and limits the access to other tablet functionalities; this mode allows the Android OS to re-open the app automatically every time it closes;
- a background service that reconnects automatically to the BLE wearables, whenever intermittent connexion losses occur, up to 10 min, which can happen e.g., if the patient is very active or temporarily moves away from the tablet;

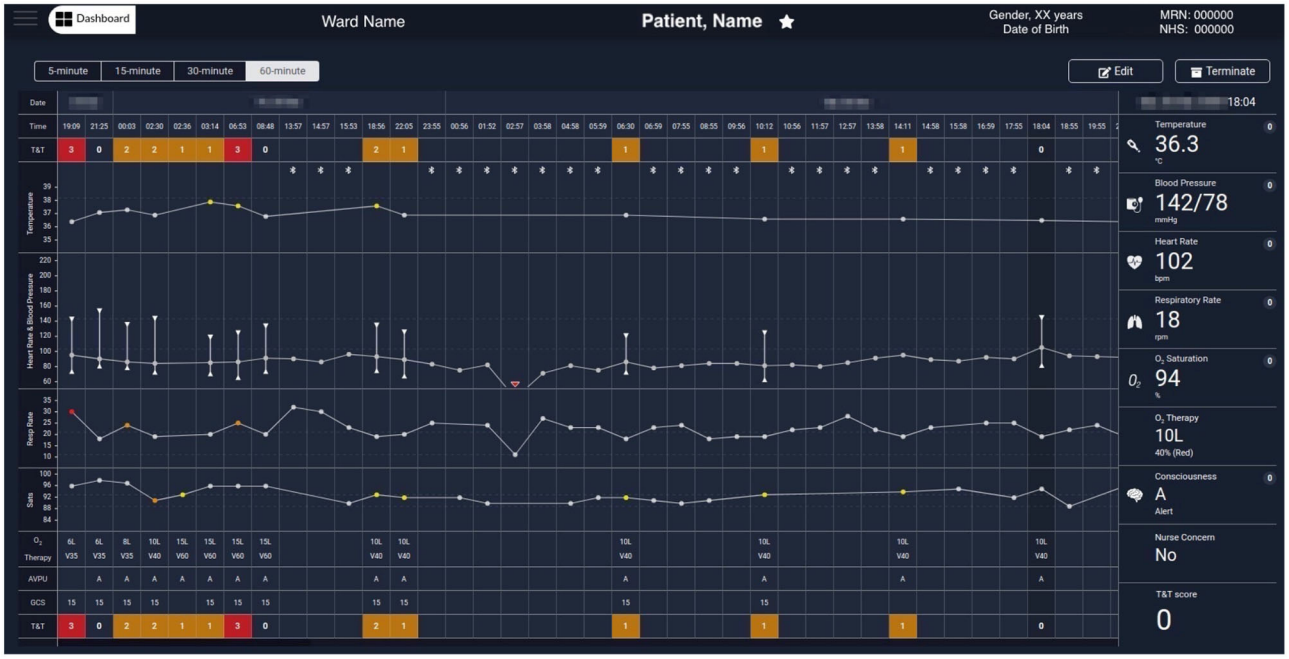


FIGURE 5 | Exemplar augmented e-obs chart from the Clinician Dashboard, showing wearable and e-obs vital-sign data from a COVID-19 patient in the isolation room. The header provides the patient's demographics (from left to right): the ward, bay and bed location, the name, gender, age (with date of birth below), and the MRN and NHS codes. The clinicians have the option to "Edit" the patient information, or "Terminate" the patient's session in the vHDU system (automatically unregistering them from the assigned tablet). The clinicians can review the periodic wearable data, i.e., the 5-min median estimates, using 15-, 30-, or 60-min windows (the most recent 5-min estimates in each window are shown), alongside the intermittent e-obs data (usually entered 4-hourly by the nurses), to facilitate tracking of the patient state. Wearable data is distinguished from the e-obs by using the Bluetooth sign at the top of the Temperature row. Only the e-obs data have scores associated with it, presented in the T&T row at the top and bottom of the chart. In this hospital setting, an Early Warning Score greater or equal to 3 was coloured red, denoting high risk. In this exemplar patient chart, only nurse observations (e-obs data) are present at the start, showing low SpO_2 (89% at 2:30 AM) and high RR, high Temp, and the use of oxygen therapy (26 rpm, 37.9°C and Venturi mask 60, V60, at 15 litres per minute, respectively, at 6:35 AM). The wearable data is summarised in hourly windows to allow the review of a longer monitoring period. It is possible to observe that during this period the VitalPatch® and the Nonin devices were fitted to the patient from 13:00 and then temporarily removed between 16:00 and 23:00. Periodic VitalPatch® data was then consistently available for remote monitoring of the patient status from the nurse bays; however, the Nonin was still removed often by the patient during that period (staff continued to try to use it to monitor the patient's continuous oxygen therapy, which can be confirmed by the presence of the Venturi mask, V40, in the 4-hourly e-obs data present during this period). MRN, Medical Record Number; NHS, National Health Service; T&T, Track & Trigger; Sats, peripheral blood Oxygen Saturation; AVPU, Alert, Voice, Pain, Unresponsive scale; GCS, Glasgow Coma Scale; RA, Room Air; O_2, Oxygen.

- a background service that forces the app to restart and reconnect if any connected wearable device is missing data for more than 10 min; this action resets the BLE connexion state at the OS level, and the app resumes the previously persisted patient- and wearable-registration state automatically.

These strategies guarantee that data loss is never >10 min from when an unaccounted software issue occurs in the app. Finally, the Clinician Dashboard was configured to reload automatically each 30 min (i.e., when loaded on a web-browser e.g., from the nurse bay). This ensures that its socket.io client regains/maintains connectivity with the streaming-service, and consequently keeping the patient cards data updated.

DEPLOYMENT OF THE AMS ON THE ISOLATION WARD

In the previous section we detailed the Oxford vHDU AMS functionalities supporting the first five requirements deemed essential to allow the remote monitoring of the vital signs of COVID-19 patients in their isolation rooms, from the remote nurse bays. Next, we describe its deployment in the hospital's isolation ward. Our initial adaptation of the AMS for monitoring COVID-19 patients on this ward was completed in just 3 weeks, during March 2020. This isolation ward has a maximum capacity of 19 isolation rooms, with two nursing bays outside the isolation rooms. Four to five nursing staff and three to four medical doctors oversee the ward, in each of its two shifts (staff numbers being adjusted as required, in particular during the week-ends).

Software Configuration

The patient data collection app was installed on 20 Android tablets, configured in "kiosk" mode and with Wi-Fi Protected Access 2 (WPA2) accounts. The web-applications back-end was deployed on one virtual server (except the compressed-files upload web-service), and an additional virtual server supported the databases and the compressed-files storage. Their technical specifications can be reviewed in **Table 2**.

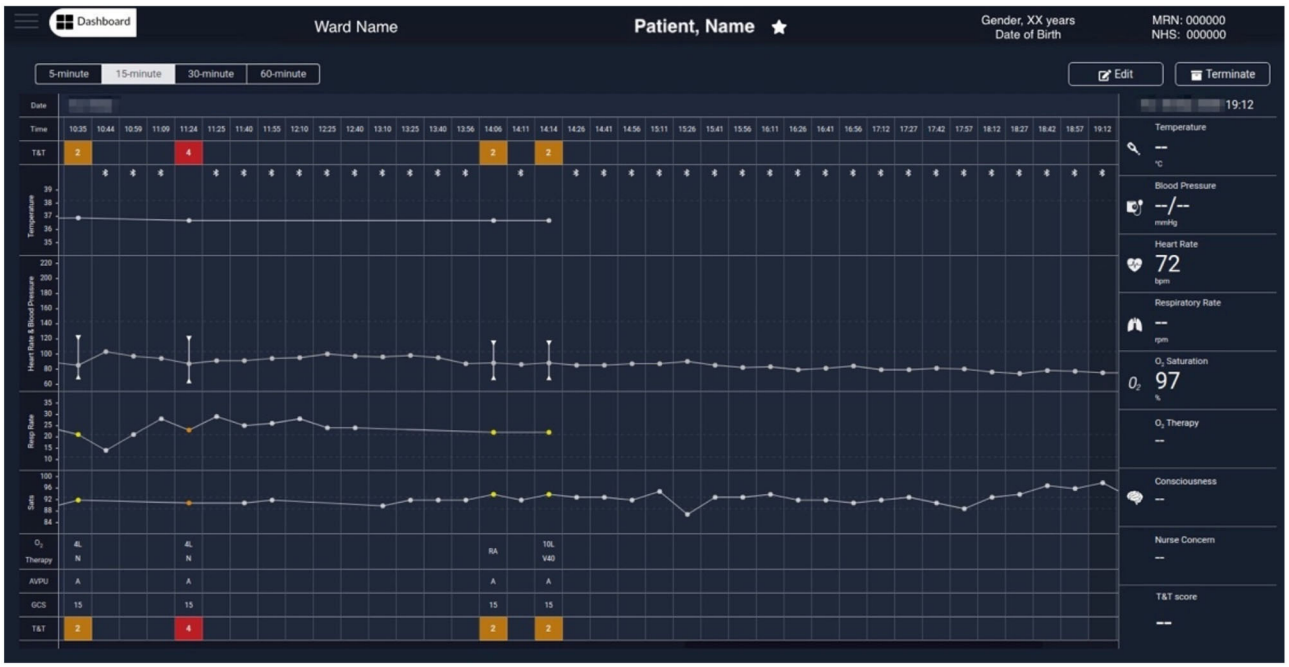


FIGURE 6 | Exemplar augmented e-obs chart from the Clinician Dashboard, showing wearable and e-obs vital-sign data from a COVID-19 patient in the isolation room. In this example, the wearable data from 10:44 to 19:12 is summarised in 15-min windows (i.e., the "15-min" view was activated, and only the most recent available 5-min median estimates within each window are displayed), and shown alongside the intermittent e-obs data. Only the e-obs data had scores associated with it, presented in the T&T row at the top and bottom of the chart. In this hospital setting an Early Warning Score (EWS) greater or equal to 3 was coloured red, denoting high risk. In the period displayed in the chart, the VitalPatch® was already fitted. The increase in the VitalPatch® RR, might have triggered the nurse review of the patient 1 h after the previous review (usually done 4-hourly when the previous total EWS is 2), as an additional e-obs set was entered by the nurse at this time, with high RR (24 rpm), low SpO_2 (90%) values and a total EWS of 4 (1 h after the previous e-obs set, at 10:35 AM). We note that at this point the patient was receiving oxygen therapy *via* a nasal cannula (N) at 4 litres per minute. It is possible to observe that, in the subsequent 3 h, the VitalPatch® was removed and the Nonin pulse oximeter fitted to the patient to monitor the SpO_2 while the oxygen therapy was escalated, by applying a Venturi Mask 40 (V40) at 10 litres per minute, at 14:14. A recovery trend can be observed, with HR and SpO_2 estimates from the Nonin pulse oximeter stabilising in the evening at 72 bpm and 97%, respectively (displayed at the end of the chart). Note that a nurse e-obs set is absent during this period in the chart (i.e., between 14:14 and 21:23, the time at which the patient was removed from the vHDU system; the chart only shows up to 19:12). MRN, Medical Record Number; NHS, National Health Service; T&T, Track & Trigger; Sats, peripheral blood Oxygen Saturation; AVPU, Alert, Voice, Pain, Unresponsive scale; GCS, Glasgow Coma Scale; RA, Room Air; O_2, Oxygen.

Both tablets and servers were behind the Oxford University Hospitals (OUH) National Health Service (NHS) Foundation Trust firewall. The transmission of HL7 v2.6 messages over MLLP is not encrypted. Therefore, an additional requirement to having our servers only available from the hospital intranet, was that they ran a bespoke anti-virus software and were managed by the hospital cyber-security team. All data communication between the tablet app and the servers was encrypted using the Secure Socket Layer (SSL). The VitalPatch® and Nonin data communication was also encrypted using the Advanced Encryption Standard (AES) over BLE. Finally, the Clinician Dashboard accounts were pre-configured to show data from the isolation ward only. These configurations helped in achieving the essential data security and patient confidentiality standards of the sixth requirement from section Introduction. Finally, we note that patient data collected from an electronic system that is part of a research study, such as the vHDU system, is required to be deleted from the hospital servers within a period of 5 years from the completion of the study.

Hardware Configuration

The AMS hardware (tablets and wearables) was configured to minimise the risk of the vHDU equipment spreading infection and to facilitate its setup in the patients rooms:

i. the VitalPatch® is a single-use device, and safely disposed of after usage;
ii. one Nonin pulse oximeter is paired exclusively with one Android tablet, and its serial number tagged on the tablet. The serial number is broadcast to the tablet app, *via* BLE, allowing staff to easily connect both. After use, the casing and finger-probe of the device are cleaned with appropriate wipes and the wrist-straps are replaced with newer ones, for the next patient (the old straps were disposed in this case);
iii. tablets are covered with a back and front protective casing; their app pre-configured with the ward, bay and bed information; while in use, tablets are deployed near a power-socket and within the reach of the Wi-Fi signal in each isolation room; after use, they are cleaned with appropriate wipes (or the casing replaced, as required) and, when not in use, they are secured in bags at the nurse bays;

iv. a dedicated computer screen is configured in the main nurse bay to display the Clinician Dashboard, which allows authorised staff to review the vital signs of multiple patients, as well as the status of their wearable devices and tablet. The dashboard is a web-application front-end, available from any ward computer screen with a modern web-browser within the hospital intranet.

Training and Maintenance

The vHDU system optimised for vital-sign monitoring of patients on the isolation ward was deployed on 20th March 2020 and the first patient was registered in the system on 23rd March (20). Two one-hour training sessions per week were given by a research nurse (LY) and a research engineer (CR), for the 1st month following the start of the first wave of the pandemic, to cover the entire staff. Between March and August 2020, the biomedical research engineers (MDS, CR and MAP) and one software engineer (RL), worked continuously to provide equipment, system configuration and software improvements. The clinical research staff (SV, CA, and LY) received weekly feedback from the nurses managing the COVID-19 patients.

The ability to calibrate the VitalPatch® RR estimates manually, *via* the tablet App, was added 2 weeks after the initial deployment. This device needs to identify the patient posture—standing vs. lying down—to select the correct accelerometer axis and start deriving RR (21). Prior to that, the patient had to take about 30 steps for the RR estimation to be calibrated, which did not happen if the device was fitted while the patient was in bed, and remained in bed.

Finally, the integration of the 4-hourly nurse observations vital-sign values into the Clinical Dashboard was completed and deployed from the middle of June 2020 (i.e., only continuous data was shown in the patient cards and augmented chart up to this point), allowing staff to review both sources of vital-sign data in one interface. This led to the system architecture in **Figure 2**: vHDU version 0.2.

RESULTS

System Uptake on the COVID-19 Isolation Ward

Preliminary feedback regarding the uptake of the vHDU AMS on the isolation wards was received mainly from clinical staff during the training and system maintenance sessions offered by the research team. It was not possible to interact with/interview the patients during the pandemic. Feedback on the AMS usage during the first wave can be categorised as follows:

i VitalPatch®–The disposable cardiac patch was well-received by both the clinical staff and the patients. Staff found it easy to fit on the patient and connect to the tablet App, and that it would not fall easily from the patient's chest. Clinical research staff observed that HR and RR signal quality deteriorated (i.e., noticeable by intermittent HR and absence of RR estimation) when the patch was applied on patients from whom it was not possible to clear the chest hair completely.

ii Nonin—Clinical staff reported that some patients found it difficult to comply with using this device continuously. This confirmed our wearability study findings (11), in which finger-based pulse oximeters such as the Nonin, were found to be less comfortable and tolerated than the cardiac patches, and ring-shaped or wrist-only pulse-oximeter devices. However, the Nonin was the only one capable of regaining connectivity without nurse intervention and with confirmed clinical-grade accuracy. Clinical staff also reported to have compared the Nonin SpO_2 estimates with those from a Dinamap vital-sign monitor used in the ward, in the first couple of weeks. Our hypoxia study (13) showed that although the Nonin was within the clinical-grade accuracy guidelines (i.e., Root-mean-squared error $\leq$ 4%), it showed a negative bias of -1.92 (± 2.73)%, and motion significantly deteriorates its estimates. The device was ultimately compulsory to use for those COVID-19 patients requiring oxygen therapy (usually less active and lying in the room beds), and therefore requiring remote monitoring of their oxygen levels.

iii Linkage with the e-obs system data—displaying the 4-hourly nurse observations alongside the wearables estimates in the Clinician Dashboard was well-received by staff. However, the manual introduction of the MRN (from the tablet App or from the Dashboard) was found to be prone to error, as in a couple of instances the incorrect number was introduced, which prohibited the capture of the intermittent vital-sign data from the HL7 receiver system. Vital-sign data capture systems usually use a barcode scanner to acquire the patient MRN (or other hospital identifier) automatically from the patient wrist-band (avoiding error). It was not possible to have such a solution ready for the AMS, as the patient enrolment was often made *via* the tablet app and it would have potentially required a barcode scanner per isolation room. To avoid this error, future approaches will use a combination of visual cues and a camera-based barcode scanner approach (*via* camera-enabled tablets).

Figures 5, **6** show wearable and e-obs data collected by the vHDU system from two different COVID-19 inpatients in isolation rooms. The wearable data are shown for hourly- and 15-min windows, respectively. Both patients received oxygen therapy to compensate for the low SpO_2 (<90%), and high RR (>20 rpm). The first patient maintained a low SpO_2 throughout the displayed period, requiring additional 4-hourly nurse observations to adjust therapy. In contrast, the second patient showed two relevant events in the wearable data: (a) an increase in the VitalPatch® RR estimate (displayed at 10:09), which might have triggered a nurse review at 11:24 (resulting in a total EWS of 4 at that point, indicating deteriorating physiology), 1 h after the previous nurse review; (b) a recovery trend towards the end of the period displayed on the chart, the Nonin pulse oximeter reporting SpO_2 and HR values of 97% and 72 bpm, respectively. For the latter patient, the nurse did not perform additional observations during the afternoon period (between 14:14 and 21:23, the time at which the patient was removed from the vHDU system—the figure only shows up to 19:12), indicating that, as

TABLE 3 | Statistics on the vHDU system usage in the John Radcliffe Hospital isolation ward, between the 20th of March and the 2nd of August 2020.

Metric	Period: 20/03 – 02/08/2020
# Patients monitored	59
# Monitored hours	2,938.8
Median monitoring hours [Q1, Q3]	31.5 [8.8, 75.4]
Median monitoring hours [Q1, Q3] (%)	88.1 [62.5, 94.5]
# Nonin SpO_2/Pulse Rate hours	1,500.5
# VitalPatch® Heart Rate hours	1,930.0
# VitalPatch® Respiratory Rate hours	1,649.9
Max number of patients simultaneously monitored	9

Q, Quartile.

long as the patient showed a recovery trend *via* the wearable system, the nurse could make the decision not to enter (with full PPE) the isolation room.

The system has now also been used during the second and third waves of the pandemic in the UK (November 2020—April 2021). Since April 2021, 20 semi-structured interviews regarding its use have been held with members of the nursing staff with experience of the system. These interviews were held *via* telephone or face-to-face, and used purposive sampling to gain as wide a range of views as possible. The findings from these interviews will be reported in a subsequent paper.

Continuous Data Available for Remote Monitoring on the COVID-19 Isolation Ward

Preliminary results regarding the continuous data captured by the AMS, and available for the remote monitoring of COVID-19 patients on the isolation rooms, during the first wave of the pandemic in the UK, was reviewed as part of the quality improvement project (Audit Datix ID Number 5973), approved by the OUH Trust on 8th April 2020. These results were assessed by calculating the total amount of data acquired by each device, the median amount of data collected per patient and the number of active monitoring sessions per day. The latter has been compared with the number of new COVID-19 cases per day in England (22) and the number of COVID-19 related hospital admissions in England (23).

A total of 59 patients were monitored *via* wearable devices between 20th March 2020 and 2nd August 2020. The system monitored patients for a total of 2,938.8 h, corresponding to 88.1 [62.5, 94.5]% of the median time the patients were registered in the system (**Table 3**). This corresponds to a median of 31.5 [8.8, 75.4] hours of vital-sign data per patient (minimum was 20 min and maximum was 10 days). The VitalPatch® HR contributed the most data with 1,930 h. The amount of RR data was about 16% lower, mainly because of the 2-week delay in the introduction of the manual re-calibration step required to start its estimation from nurse input. The amount of SpO_2 data is about 21% lower than that of the HR because some patients found it difficult to be compliant with the wearing of the finger-worn Nonin probe. Staff would only enforce the wearing of the probe if the patient was given oxygen therapy. When the VitalPatch® data was not available, the Nonin PR estimates would provide the HR measurements on the dashboard.

From observation of **Figure 7**, we can infer that at the peak of the first wave of the pandemic in the Oxford region, staff placed half of the ward patients on continuous monitoring (i.e., nine out of a total of 19 available rooms), using our wearable vHDU system.

Work is underway to analyse retrospectively the continuous vital-sign data collected during the COVID-19 pandemic, including the second and third waves (between March 2020 and April 2021) and corresponding patient adverse events (e.g., cardiac arrest calls, escalation of care, and mortality). This analysis will include (a) the frequency of nurse observations before and during the use of the vHDU system; (b) the relationship between instances of oxygen desaturation and associated respiratory and heart rate patterns; and (c) the relationship between the patterns of the vital-sign collected by the system and patients adverse events. We will also be reporting on the results of these analyses in a subsequent paper.

DISCUSSION

Comparison With Other AMS

The challenges created by the COVID-19 pandemic to health services throughout the developed world has led to the accelerated deployment and acceptance of many remote monitoring technologies in the clinical setting (2). The work presented in this paper describes the adaptation of a wearables based AMS for real-time remote monitoring of the vital signs of COVID-19 patients being cared for on an isolation ward. The major difference between our prototype system and those reviewed earlier, is the assumption that nurse observations will be recorded using third-party interoperable software. These data are sent to the hospital Electronic Patient Record and made available to our system *via* HL7. Our system then displays it alongside the wearable continuous data.

We note that while the ViSi Mobile System only displays wearable device data, the Vista Solution® and Sensium®'s "E-Obs" package, allow staff to enter additional patient data in order to compute the EWS, generate alerts and suggest a clinical action. Our approach allowed clinical staff to retain their current e-obs system, but to review those data alongside the wearables data in one single chart (as illustrated in **Figures 5**, **6**).

The 5-day battery life of the VitalPatch® adopted in our system, also helps minimise contact with patients in isolation rooms, when compared with the ViSi Mobile System modules which require daily charging. Finally, the Sensium® System does not include the continuous measurement of SpO_2, the key parameter for managing COVID-19 patients, many of whom require oxygen therapy.

Usage of the Oxford vHDU AMS During the First Wave

The need to undertake patient care on isolation wards while keeping clinical staff safe, and the focus of the local hospital teams (including clinical, engineering, and information technology) on helping the UK NHS cope with the COVID-19 pandemic created

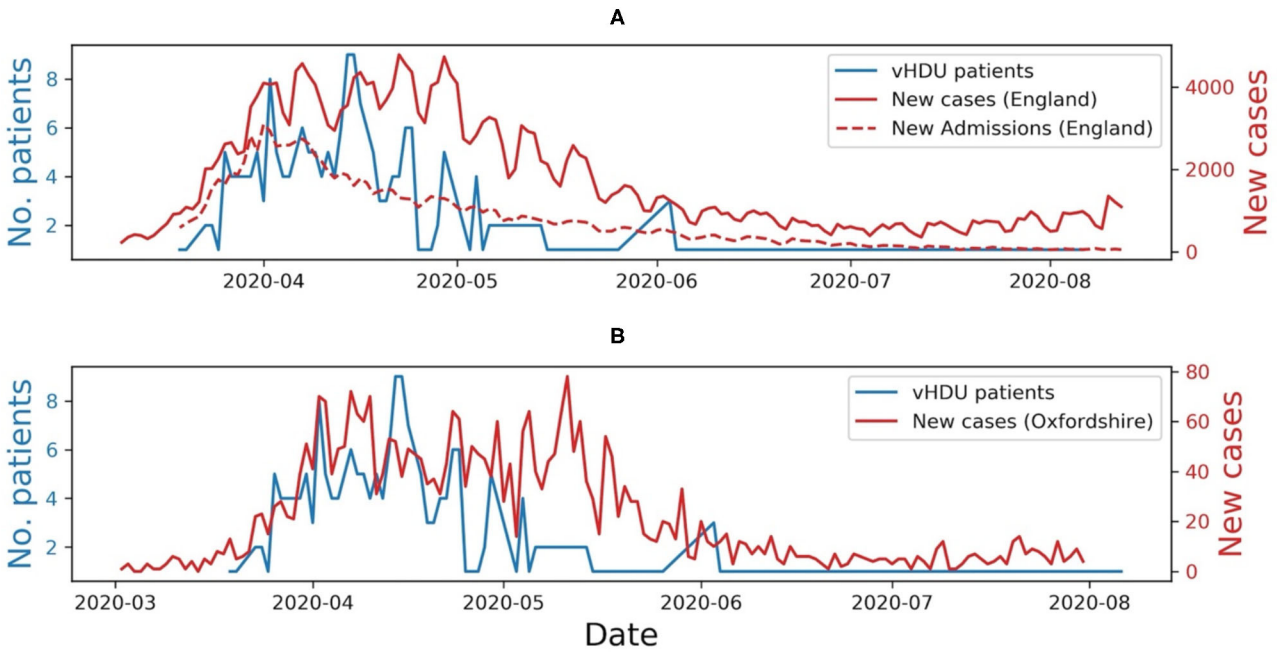


FIGURE 7 | (A) Number of COVID-19 patients monitored *via* the virtual High-Dependency Unit (vHDU) system in the John Radcliffe Hospital isolation ward vs. the number of new daily COVID-19 cases and new daily COVID-19 hospital admissions, in England, between the 20th of March and the 2nd of August 2020. **(B)** The number of new daily COVID-19 cases in Oxfordshire, is shown for comparison. It can be observed that the system usage followed the trend of the first wave of the pandemic in the UK.

an environment which enabled the rapid adoption of our vHDU system into clinical practise. A number of factors helped in deploying the system in such a short time:

a The prior selection of wearable devices that avoid the need for constant adjustment by nursing staff: amongst all the wearable pulse oximeters which we had previously investigated, the Nonin device was the only one that demonstrated good clinical accuracy, allowed consistent BLE communication and activated automatically once the patient's finger was positioned within the probe. Similarly, the results obtained with the VitalPatch® confirmed the conclusions from our previous wearability and functionality tests (11), as the patch was rapidly adopted on the isolation ward, with the advantage that it is a disposable device, thereby avoiding the risk of spreading the infection.

b The use of interfaces familiar to OUH clinical staff: with minimal training, nursing staff when making their observations were able to review the outputs of the data collection app on the Android tablet as if it were a "bedside monitor;" additionally, the charts on the remote Clinician Dashboard were modelled on those used in the electronic observation system, widely used throughout the hospital.

c The inclusion of fault-tolerant software mechanisms, e.g., to automatically recover from wireless disconnection of the wearable sensors and unaccounted software issues (with minimal data loss), to avoid nursing staff entering the isolation ward in order to make adjustments.

d The close collaboration of the biomedical engineering and the clinical research teams with the ward staff (and vice-versa), enabled rapid feedback on the usage of the system, so that it could continually be iterated. During this period, biomedical expertise was required to improve the patient data collection app, e.g., in facilitating the setup of the VitalPatch® RR estimation, and to design and develop the patient chart displaying both the continuous and intermittent vital-sign data. The clinical research team validated the AMS functionalities with the ward staff.

The usage of the AMS on the isolation ward followed the trend of the first wave of the pandemic in the UK. Usage decreased when the number of COVID-19 related hospital admissions decreased (as shown in **Figure 7A**) and when other hospital wards were reorganised to make them more prepared to care for patients with COVID-19. We also observed that the vital-sign data availability followed that which we had previously seen in our wearability study (11), the Nonin capturing 21% less data than the VitalPatch®. The analysis of the amount of data captured per patient was limited to the time they were registered in the system (i.e., ward admission/discharge times were absent from our dataset). Nevertheless, vital-sign data was available to be reviewed by clinical staff, i.e., from the remote nurse bays, for a median of 88.1 [62.5, 94.5]% of the patients' time in the system. The latter, allied to the fact that half of the ward was using at least one wearable device at the peak of the first wave, and that the system was also used during the second and third waves of the

pandemic in the UK, are good indicators that such systems are needed to monitor patients in isolating rooms while keeping the hospital staff safe.

Future Work

The Oxford vHDU AMS has now also been used during the second and third waves of the pandemic in the UK. Therefore, the next steps in evaluating and improving the AMS for the isolation wards, include:

i the analysis of the human factors that might influence the usage of the system in the ward, e.g., *via* focus groups interviews with clinical staff;
ii the analysis of patterns in the continuous and intermittent vital-sign data, collected by the vHDU system, vs. COVID-19 patients adverse events (occurring in the isolation ward);
iii the ability to combine the continuous HR, RR, and SpO2 data from the wearables with the complementary data from the nurse observations (BP, Temp and level of consciousness), available from the hospital e-obs system (19), to compute the EWS for each patient on the vHDU system;
iv the implementation of a notification system for clinical events linked to abnormal physiology (e.g., alerts for low SpO_2 values for COVID-19 patients), and for technical events, such as missing wearable data (the validation of the scoring and alerting system in a dedicated clinical study is currently being planned);
v and finally, exploring the use of the system outside of the hospital, as some patients could continue to be monitored remotely after they have been discharged home.

CONCLUSION

We have developed and deployed a wearables system based on commercial off-the-shelf components, that enables the remote, real-time review of the vital signs of ambulatory COVID-19 patients on a set of individual rooms within the isolation ward of our local hospital. The system was optimised to meet six different requirements which had been established for reliable continuous monitoring of the cardio-respiratory physiology of the patients nursed on this ward. System usage increased when ward occupancy increased during the peak of the first wave of the pandemic in the UK, demonstrating the clinical usefulness of the system.

Beyond the pandemic, we aim to conclude the evaluation of our vHDU wearables system, in which we will evaluate whether the addition of automated alerts to the AMS can help nursing staff identify patient deterioration earlier (between their regular vital-sign observations) in high-dependency or step-down units.

AVAILABILITY AND REQUIREMENTS

Project: vHDU (virtual High-Dependency Unit).
Home Page: https://www.ndcn.ox.ac.uk/research/critical-care-research-group-kadoorie-centre
Operating System(s): Android OS for the patient data collection app; the remainder of the system is platform-independent.
Programming language: HTML, CSS, JavaScript, PHP, Java, and SQL.
Other requirements: Modern web-browser, Android Tablets, VitalPatch® (VitalConnect, USA), WristOx2® 3150 BLE OEM (Nonin, USA) devices.
Licence: Proprietary Licence.
Any restriction to use by non-academics: licence needed.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because these data requires the permission of the Oxford University Hospitals National Health Service (NHS) Foundation Trust to be used. Requests to access the datasets should be directed to Peter Watkinson, peter.watkinson@ndcn.ox.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Oxford University Hospitals National Health Service Foundation Trust. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MS, CR, MP, SV, CA, and LY designed the vHDU software. MS, CR, and MP implemented, tested, and deployed the vHDU software for use in the John Radcliffe Hospital isolation ward. CR, SV, CA, and LY provided training and support to the clinical staff. MS and CR drafted the manuscript. MP and LT helped draught it. LT and PW co-lead the project. All authors reviewed the manuscript and approved it.

FUNDING

MS, CR, SV, CA, LY, PW, and LT were funded/supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. MP was funded by a Drayson Research Fellowship.

ACKNOWLEDGMENTS

The authors would like to acknowledge: Roman Lyshak and Yuan Chi for their help in developing the web-service back-end; Catarina Santos for helping with the design of the app interface; Christopher Biggs for providing system support to isolation ward clinical staff; Anne Thompson for helping with adoption of the system in the isolation ward; finally, the isolation ward clinical staff for all their work in caring for COVID-19 patients.

REFERENCES

1. World Health Organization. *WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV)*. (2020). Available online at: https://www.who.int/dg/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov) (accessed October 12, 2020).
2. Behar J, Liu C, Kotzen K, Tsutsui K, Corino VDA, Singh J, et al. Remote health diagnosis and monitoring in the time of COVID-19. *Physiol Meas*. (2020) 41:10TR01. doi: 10.1088/1361-6579/abba0a
3. Willan J, King AJ, Jeffery K, Bienz N. Challenges for NHS hospitals during covid-19 epidemic. *BMJ*. (2020) 368:m1117. doi: 10.1136/bmj.m1117
4. Ghanchi A. Adaptation of coronavirus disease (COVID-19) protocols to a parisian maternity unit during the 2020 pandemic: a managerial perspective. *Disaster Med Public Health Prep*. (2020) 2020:1–4. doi: 10.1017/dmp.2020.234
5. National Center for Immunization and Respiratory Diseases - Division of Viral Diseases. *Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)*. Centers for Disease Control and Prevention. (2020). Available online at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html (accessed November 15, 2020).
6. Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party*. London (2017).
7. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Personal Med*. (2018) 2018:44. doi: 10.2217/pme-2018-0044
8. Soon S, Svavarsdottir H, Downey C, Jayne DG. Wearable devices for remote vital signs monitoring in the outpatient setting: an overview of the field. *BMJ Innovations*. (2020) 2:e000106. doi: 10.1136/bmjinnov-2019-000354
9. Dias D, Cunha JPS. Wearable health devices—vital sign monitoring, systems and technologies. *Sensors*. (2018) 18:2414. doi: 10.3390/s18082414
10. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. *J Medical Internet Res*. (2020) 22:18636. doi: 10.2196/preprints.18636
11. Areia C, Young L, Vollam S, Ede J, Santos M, Tarassenko L, et al. Wearability testing of ambulatory vital sign monitoring devices: prospective observational cohort study. *JMIR mHealth uHealth*. (2020) 23:20214. doi: 10.2196/preprints.20214
12. Areia C, Vollam S, Piper P, King E, Ede J, Young L, et al. Protocol for a prospective, controlled, cross-sectional, diagnostic accuracy study to evaluate the specificity and sensitivity of ambulatory monitoring systems in the prompt detection of hypoxia and during movement. *BMJ Open*. (2020) 10:e034404. doi: 10.1136/bmjopen-2019-034404
13. Santos M, Vollam S, Pimentel M, Morgado Areia C, Young L, Roman C, et al. Wearable pulse oximeters in the prompt detection of hypoxaemia and during movement: a diagnostic accuracy study. *JMIR Prepr*. 2021:28890. doi: 10.2196/preprints.28890
14. Morgado Areia C, Santos M, Vollam S, Pimentel M, Young L, Roman C, et al. Chest patch for continuous vital-sign monitoring: a clinical validation study during movement and controlled hypoxia. *J Med Internet Res*. (2021) 2021:27547. doi: 10.2196/preprints.27547
15. Nonin Medical Inc, Plymouth U. *WristOx2® Model 3150 with Bluetooth® Low Energy*. Available online at: https://www.nonin.com/products/wristox2-model-3150-with-ble/ (accessed October 13, 2020).
16. VitalConnect. *VitalPatch®*. Available online at: https://vitalconnect.com/solutions/vitalpatch/ (accessed October 13, 2020).
17. ReactiveX. *ReactiveX*. Available online at: http://reactivex.io/ (accessed July 27, 2021).
18. Dahella SS, Briggs JS, Coombes P, Farajidavar N, Meredith P, Bonnici T, et al. Implementing a system for the real-Time risk assessment of patients considered for intensive care. *BMC Med Inform Decis Mak*. (2020) 20:161. doi: 10.1186/s12911-020-01176-0
19. Wong D, Bonnici T, Knight J, Morgan L, Coombes P, Watkinson P. SEND: a system for electronic notification and documentation of vital sign observations. *BMC Med Inform Decis Mak*. (2015) 15:68. doi: 10.1186/s12911-015-0186-y
20. Oxford University Hospitals. *New Wearable Technology to Monitor COVID-19 Patients*. (2020). Available online at: https://www.ouh.nhs.uk/covid-19/news/article.aspx?id=1215 (accessed October 13, 2020).
21. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (2013). p. 4058–61. doi: 10.1109/EMBC.2013.6610436
22. GOV.UK. *Coronavirus (COVID-19) in the UK*. (2020). Available online at: https://coronavirus.data.gov.uk/cases (accessed October 14, 2020).
23. GOV.UK. *Coronavirus (COVID-19) in the UK*. (2020). Available online at: https://coronavirus.data.gov.uk/healthcare (accessed October 14, 2020).

Conflict of Interest: LT and PW reports significant grants from the National Institute of Health Research (NIHR), UK and the NIHR Biomedical Research Centre, Oxford, during the conduct of the study; modest grants and personal fees from Sensyne Health, outside the submitted work and hold share options in the company. LT works part-time for Sensyne Health. PW was Chief medical Officer for Sensyne Health until March 2020.

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 04 November 2021
doi: 10.3389/fdgth.2021.736418

Calibration-Free Gait Assessment by Foot-Worn Inertial Sensors

Daniel Laidig [1], Andreas J. Jocham [2], Bernhard Guggenberger [2], Klemens Adamer [3], Michael Fischer [3,4,5] and Thomas Seel [6]*

[1] Control Systems Group, Technische Universität Berlin, Berlin, Germany, [2] Institute of Physiotherapy, FH JOANNEUM University of Applied Sciences, Graz, Austria, [3] Vamed Rehabilitation Center Kitzbuehel, Kitzbuehel, Austria, [4] Ludwig Boltzmann Institute for Rehabilitation Research, Vienna, Austria, [5] Hannover Medical School MHH, Clinic for Rehabilitation Medicine, Hannover, Germany, [6] Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Walking is a central activity of daily life, and there is an increasing demand for objective measurement-based gait assessment. In contrast to stationary systems, wearable inertial measurement units (IMUs) have the potential to enable non-restrictive and accurate gait assessment in daily life. We propose a set of algorithms that uses the measurements of two foot-worn IMUs to determine major spatiotemporal gait parameters that are essential for clinical gait assessment: durations of five gait phases for each side as well as stride length, walking speed, and cadence. Compared to many existing methods, the proposed algorithms neither require magnetometers nor a precise mounting of the sensor or dedicated calibration movements. They are therefore suitable for unsupervised use by non-experts in indoor as well as outdoor environments. While previously proposed methods are rarely validated in pathological gait, we evaluate the accuracy of the proposed algorithms on a very broad dataset consisting of 215 trials and three different subject groups walking on a treadmill: healthy subjects ($n = 39$), walking at three different speeds, as well as orthopedic ($n = 62$) and neurological ($n = 36$) patients, walking at a self-selected speed. The results show a very strong correlation of all gait parameters (Pearson's r between 0.83 and 0.99, $p < 0.01$) between the IMU system and the reference system. The mean absolute difference (MAD) is 1.4 % for the gait phase durations, 1.7 cm for the stride length, 0.04 km/h for the walking speed, and 0.7 steps/min for the cadence. We show that the proposed methods achieve high accuracy not only for a large range of walking speeds but also in pathological gait as it occurs in orthopedic and neurological diseases. In contrast to all previous research, we present calibration-free methods for the estimation of gait phases and spatiotemporal parameters and validate them in a large number of patients with different pathologies. The proposed methods lay the foundation for ubiquitous unsupervised gait assessment in daily-life environments.

Keywords: inertial sensors, IMU, human motion analysis, gait analysis, gait assessment, gait phases, rehabilitation, walking

OPEN ACCESS

Edited by:
Mohamed Elgendi,
University of British Columbia, Canada

Reviewed by:
Thibault Bernard Warlop,
Catholic University of Louvain, Belgium
Matthew R. Patterson,
Data Scientist Shimmer Research Ltd Dublin, Ireland

****Correspondence:***
Daniel Laidig
laidig@control.tu-berlin.de

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

Received: *05 June 2021*
Accepted: *24 September 2021*
Published: *04 November 2021*

Citation:
Laidig D, Jocham AJ, Guggenberger B, Adamer K, Fischer M and Seel T (2021) Calibration-Free Gait Assessment by Foot-Worn Inertial Sensors. Front. Digit. Health 3:736418. doi: 10.3389/fdgth.2021.736418

1. INTRODUCTION

Walking is a central activity of daily life, and restrictions of this ability lead to a reduction in the quality of life (1, 2). Therefore, gait analysis is an important tool in different medical and therapeutic fields (3, 4). The measurement of various gait characteristics can either facilitate diagnosis or be used to track the progress of rehabilitation. Gait can be measured by spatial (e.g., step or stride length) and temporal (e.g., stride time, cadence) parameters, relative durations of gait phases, and kinematic and kinetic gait variables (5). These parameters are used to quantify gait deviation in both clinical practice and research, and their use varies with the medical field, the research question, and the analysis options. While gait assessment in clinical practice is mostly based on visual observation by medical experts (6), it is desirable to support expert knowledge and time by objective measurements. This is also important because relevant gait changes are often too subtle to be detected by the naked eye (7).

Traditionally, sensor-based gait assessment is performed with stationary systems like marker-based optical motion tracking, instrumented treadmills, or pressure-sensitive walkways (6, 8). Besides being expensive, one major drawback of those systems is that they are limited to a small capture space or require the subject to walk on a treadmill (4, 9–12). Furthermore, the use of walking aids is often not possible or restricted in combination with such systems.

A promising, more ambulatory, and less restrictive alternative is inertial gait analysis, i.e., gait analysis with inertial sensor technology. Lightweight and battery-powered inertial measurement units (IMUs) are used, which transmit the data wirelessly.

The transition from expensive stationary systems to small wearable sensors opens up possibilities that go beyond replacing the measurement technology used for gait assessment in a clinical setting. Integrating objective long-term gait monitoring in day-to-day life—as illustrated in **Figure 1**—could provide more powerful tools for clinicians to help patients in rehabilitation but also to gain further insights into disease progression. Furthermore, non-obtrusive wearable plug-and-play systems facilitate applications in neuroprosthetics (13) or exoskeletons and can be used to provide biofeedback (14). In the last years, wireless battery-powered IMUs have become smaller, lighter, more accurate, and at the same time cheaper and more energy-efficient, and it is to be expected that this development is going to continue. For those new trends, it is important to develop methods that can provide a wide variety of gait parameters that are useful to medical experts. At the same time, the methods need to be robust so that the system can be used by patients in unsupervised settings, outdoors as well as indoors.

It has been shown by previous contributions (15–17) that major gait parameters can be determined with two IMUs that are placed on the feet or the shoes, as illustrated in **Figure 1**. This includes stride length, gait phase durations (e.g., stance and swing percentage), and also the cadence and walking speed.

Our aim is to propose methods for gait assessment that meet the requirements for day-to-day life monitoring in unsupervised settings and that are validated on a broad group of subjects including patients with various gait pathologies. The proposed methods do not assume any fixed orientation of the sensor with respect to the foot and do not require the subject to perform dedicated calibration movements. Furthermore, magnetometers are not used since the magnetic field is known to be severely disturbed in indoor environments (18). This makes the use of inertial gait analysis easy and practical in clinical settings and facilitates future applications of ubiquitous gait analysis in home environments.

The remainder of the article is structured as follows. In section 2, we briefly review existing methods for IMU-based spatiotemporal gait parameter estimation. In section 3, we describe the proposed methods, which we then validate in section 4 using experimental data of 98 orthopedic and neurological patients, as well as 39 healthy subjects walking at different speeds. The results are discussed in section 5, and section 6 provides conclusions.

2. BRIEF REVIEW OF IMU-BASED SPATIOTEMPORAL GAIT PARAMETER ESTIMATION

Several methods have been proposed that employ inertial sensors to obtain spatiotemporal gait parameters. In the following, we present a brief overview of the current state of the art and summarize the different hardware setups that are used, which parameters are calculated, and how the methods were validated. **Table 1** categorizes 23 publications that provide a range of examples for the variety of existing approaches in the estimation of spatiotemporal gait parameters with inertial sensors.

There are different hardware setups, based on the number of inertial sensors and their placement. The chosen setup has an impact on which and how many parameters can be derived from the measured data. The most commonly used setup consists of two IMUs. As shown in **Table 1**, sensors are typically placed on the feet or shoes and sometimes on the shank. This setup is occasionally extended by adding a third sensor on the pelvis or lumbar spine (34, 35). Note that it has even been shown that temporal gait events can be obtained from a single IMU at the pelvis (38), but the potential for extracting further spatial parameters is limited. Full (lower) body motion tracking opens up additional possibilities, as demonstrated with 7 IMUs on the lower body and pelvis in (37) and with 8–15 IMUs in (36). Another, less common, option consists of combining inertial sensors with further measurement devices, e.g., a camera on one foot and LEDs on the other foot to facilitate the direct measurement of relative positions (39).

Some methods require that a known orientation of the sensor axes with respect to the anatomical foot axes has to be ensured by precise placement. Many methods in the literature are based on such assumptions, including (13, 15–17, 22, 23, 25, 26, 30–33). In practice, however, ensuring a precise placement is a challenge, especially in non-supervised application scenarios and during activities of daily life. Alternatives are to develop methods that are agnostic to the sensor-to-segment orientation—e.g., by

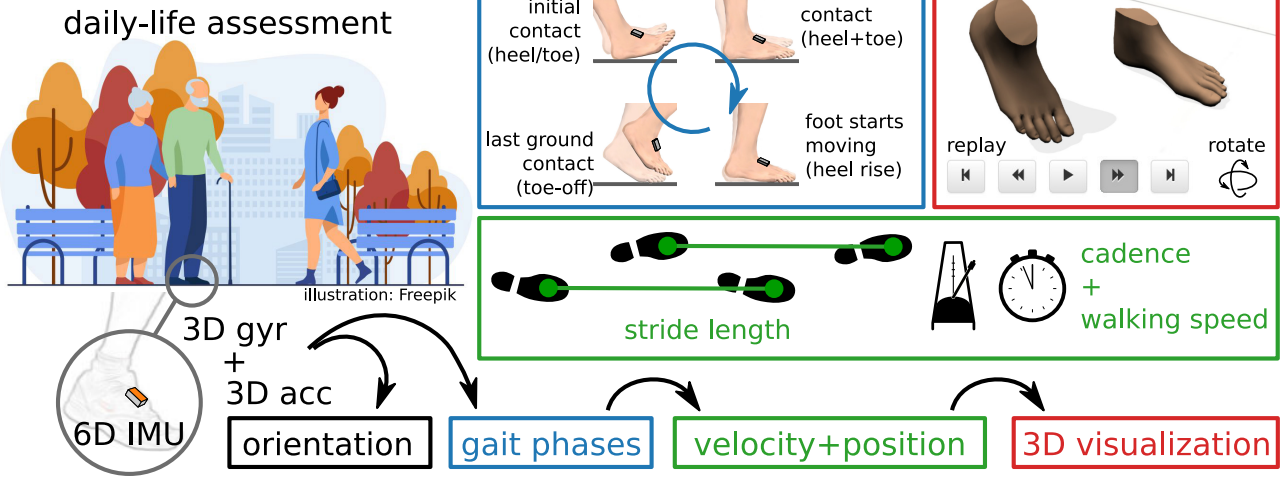


FIGURE 1 | Inertial gait analysis can be realized with two miniaturized IMUs on the shoes, enabling daily-life assessment outside of laboratory environments. From the raw sensor data, orientation, gait phases, and velocity and position trajectories can be estimated. Parameters commonly used in gait analysis, such as stride length, cadence, and walking speed, can easily be derived from this.

only relying on signal norms—or to determine this orientation in a process commonly called anatomical calibration (40). For setups with more sensors, there are recently developed methods that facilitate automatic anatomical calibration by exploiting kinematic constraints of the respective joints without requiring the subject to perform precise calibration movements (41, 42). For those setups, the linking of the sensors to the body segments poses another challenge to a plug-and-play approach, which can be solved by automatic pairing methods (43).

The calculation of spatiotemporal gait parameters is usually implemented in a two-stage approach. In a first step, gait events and corresponding gait phases are detected. In a second step, spatial parameters are calculated.

Existing methods vary in the set of gait events or phases that are detected. In many cases, the focus is only on the separation between stance and swing (cf. **Table 1**), although sometimes additional events, such as mid-swing (33), are also detected. It is also common to detect four events that occur during the gait cycle and are only defined by the ipsilateral (same) foot. Those events are initial contact, full contact, heel rise, and toe-off, although the terminology varies. Despite being common practice in gait analysis (5, 44), employing bilateral information, i.e., combining information from both feet to define the gait phase, is far less common in IMU-based gait analysis. One example is (36) in which single and double limb support durations are calculated.

There are various approaches for detection of gait events using inertial sensors. It has been shown that exploiting features of the angular rate signal in the sagittal plane is sufficient to achieve reliable gait event detection (16, 22, 25, 30, 32). Many other methods use both accelerometers and gyroscopes and detect characteristic signal features in the inertial sensor data, including (13, 15, 17, 19–21, 23, 31, 33). Sometimes automatic adaption mechanisms are used to adjust thresholds based on the subject's walking style (19–21, 33). An alternative to the signal-based methods is to rely on a kinematic model to detect gait events (36, 37). Machine learning methods, often based on hidden Markov models (26, 35), are also used for event detection [cf. (45)].

In addition to the detection of gait events, spatial parameters such as stride length and walking speed are often calculated. Those parameters are obtained by either signal integration,

TABLE 1 | Overview of IMU-based spatiotemporal gait parameter estimation literature.

Employed sensor setup	
2 IMUs on feet/shoes	(15, 17, 19–31)
2 IMUs on shank	(13, 16, 27, 30, 32, 33)
3 or more IMUs	(34–37)
Detected gait phases	
Stance/swing	(13, 16, 17, 22, 23, 27, 30, 32, 33, 35, 37)
4 unilateral events	(15, 19–21, 25, 26)
Single/double support	(29, 36)
Ground truth used for evaluation	
Optical motion capture	(20, 25, 26, 28, 29, 31, 34, 37)
Pressure-sensitive walkways	(16, 17, 23, 32, 33, 35)
Instrumented treadmills	(24, 27)
Pressure insoles	(15, 30)
Others/none	(19, 21, 22, 36)
Non-healthy subjects included in evaluation	
None (healthy only)	(21, 22, 24–28, 34, 36, 37)
≤ 20	(20, 30, 31, 33, 35)
> 20	(15–17, 23, 29, 32)

A total of 23 publications that describe estimation of spatiotemporal gait parameters with IMUs are categorized based on sensor setup, detected gait phases, and the ground truth and number of non-healthy subjects for evaluation.

human gait models, or by machine learning methods (46). By far the most common approach is numerical strapdown integration of the accelerations (16, 17, 22, 27, 28, 31, 32, 37). The cyclic nature of gait and the fact that there is frequent ground contact are exploited to correct for drift that is due to double integration. It has been shown that Fourier-based integration is an alternative to numerical integration (34), that spatial parameters can be obtained from kinematic models (27, 36), and that convolutional neural networks can also be used to estimate spatial parameters (23).

Most publications focus on common spatiotemporal parameters such as stride length, walking speed, and cadence. Other than those spatiotemporal parameters, there is a multitude of spatiotemporal gait parameters that are relevant in a clinical context for various pathologies (6). Examples that can be estimated using inertial sensors include step width (37), swing width (31, 37), incline (22), and foot clearance (47).

Some publications (13, 19–21, 25, 33) focus on real-time detection of events, e.g., to trigger functional electrical stimulation (FES). While the approaches used are usually similar to the ones used in offline gait analysis, this typically implies a focus on minimizing the detection delay rather than the accuracy of the reported values.

As shown in **Table 1**, evaluation is often performed with marker-based optical motion capture as ground truth. Systems based on the detection of pressure, such as pressure-sensitive walkways, instrumented treadmills, and pressure insoles, are a common alternative. In some cases, no validation with respect to a gold standard is performed. Instead, the settings of a (calibrated) treadmill are used for walking speed and incline (22), a manually counted number of steps is combined with the detection of irregularities (21), validation is performed by visual inspection of the results (19), or the focus is only on test-retest reliability (36).

Even though it has been shown that the accuracy of gait analysis methods decreases when applied to non-healthy subjects (45), the evaluation of inertial gait analysis methods is often only based on healthy subjects. When data obtained from non-healthy subjects is part of the evaluation, the number of subjects is often small, for example five transfemoral amputees (20), 10 stroke patients (33), 10 hemiparetic patients and 10 Huntington's disease patients (35), or 10 patients with Parkinson's disease (31).

To the best of our knowledge, few publications (15–17, 23, 29, 32) exist which propose methods for IMU-based spatiotemporal gait parameter estimation *and* validate the methods on a larger set of subjects with gait pathologies. In the following, we briefly summarize those publications.

In (15), sensors are placed on the forefoot in a known orientation, and four different unilateral gait events are detected based on features of the angular velocity in the sagittal plane, the norm of the accelerometer signal, and the derivative of angular velocity norm. Using pressure insoles as reference, the method is validated on 10 healthy and 32 orthopedic subjects.

The commercial Gait Up system is evaluated in (29) with 25 subacute stroke patients as subjects and marker-based optical motion capture as reference.

Gait events and stride length are calculated in (16) based on shank-mounted IMUs. Events are detected based on the angular rate in the sagittal plane, and stride length is obtained via double integration of the accelerations. The latter relies on the proprietary orientation estimation algorithm provided by the sensor manufacturer. Experimental evaluation is performed using the GAITRite pressure-sensitive walkway as reference on 10 healthy elderly and 30 non-healthy subjects.

In (32), the same method is validated on a much larger group of subjects, consisting of 236 community-living older adults, including 31 mild cognitive impaired subjects and 125 Parkinson's disease patients.

In (17), IMUs are placed laterally on the shoe in a fixed orientation, stance, and swing durations are calculated based on characteristic signal features, and the stride length is obtained via double integration. The method is evaluated using a large data set of 101 geriatric inpatients, with reference data obtained from a GAITRite pressure-sensitive walkway.

Using the same gait event detection method and the same data set for evaluation as (17), Hannink et al. (23) estimates stride length, stride width, mediolateral change in foot angle, heel contact times, and toe contact times using deep convolutional neural networks.

In summary, the main shortcoming of existing approaches for the vision of plug-and-play ambulatory gait analysis is that most methods—especially those with broad validation—require a precise attachment of the sensor to the subject's foot. Some methods only focus on gait events and do not provide spatial parameters, and some methods rely on proprietary algorithms of the sensor manufacturers. Furthermore, very few of the proposed methods are validated on a large group of subjects with diverse gait pathologies.

In the following section, we propose a set of methods that combine the valuable achievements of existing methods with additional features that overcome the remaining limitations.

3. METHODS

In the following, we propose a set of methods to determine gait parameters from two IMUs attached to the foot. The proposed methods are based on the following assumptions and requirements: An IMU is attached to each foot (or shoe) in an arbitrary orientation. This implies that the proposed method does not make any assumption about the orientation of the sensor coordinate system, which means it does not require any specific sensor axis to be aligned with an anatomical or functional axis of the foot. In order to avoid artifacts caused by toe or ankle motions, and also to not limit the subject's freedom of movement, we propose to attach the IMU on the instep, i.e., the dorsal side of the midfoot. We obtain the accelerometer and gyroscope readings of both IMUs at a fixed sampling rate (typically in the range 50–1,000 Hz). We assume that data for several steps is processed at once, which allows us to employ non-causal signal processing to increase the accuracy compared to sample-by-sample real-time capable methods. This processing can either be performed in batches while the subject is walking, e.g., for use

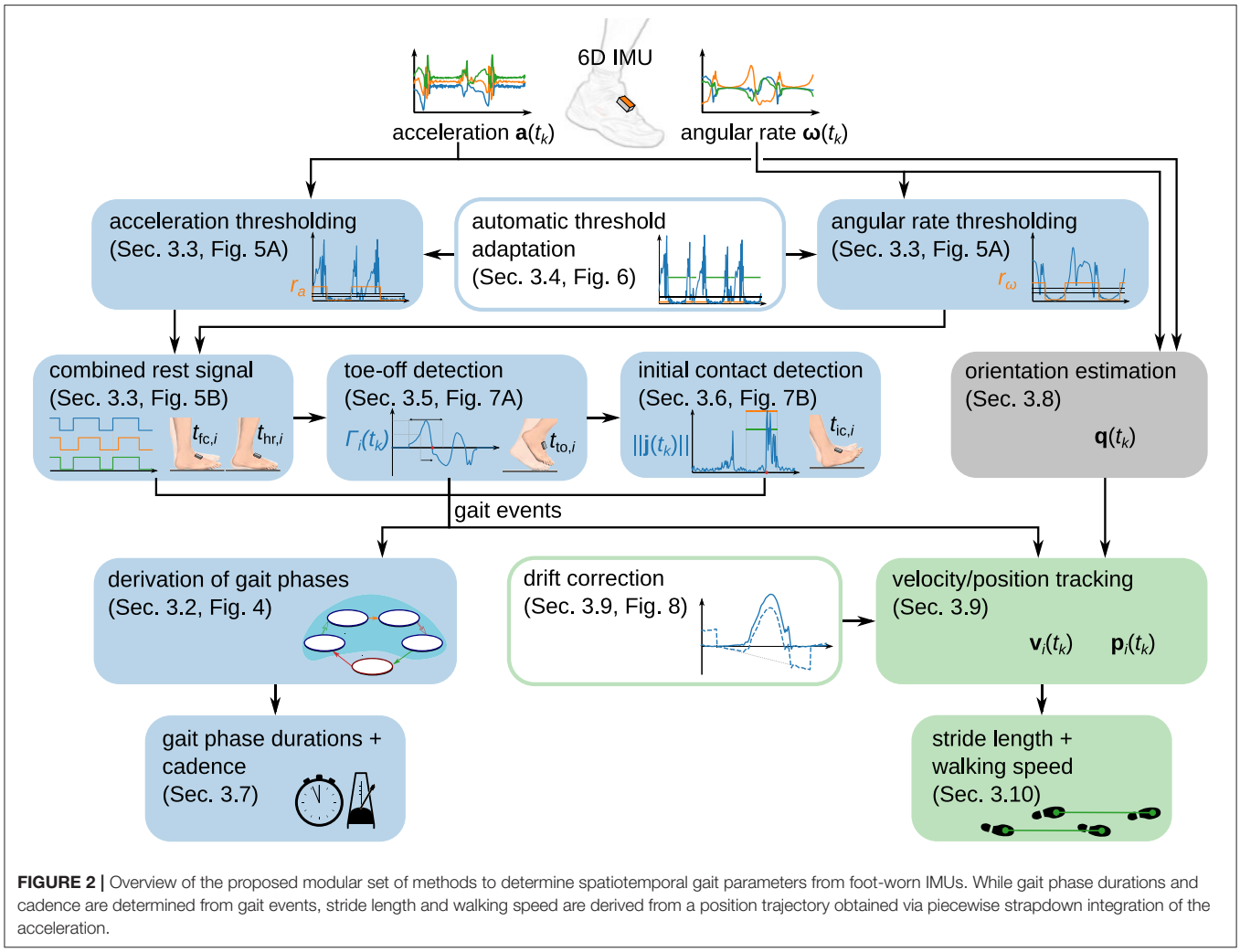


FIGURE 2 | Overview of the proposed modular set of methods to determine spatiotemporal gait parameters from foot-worn IMUs. While gait phase durations and cadence are determined from gait events, stride length and walking speed are derived from a position trajectory obtained via piecewise strapdown integration of the acceleration.

in biofeedback applications, or after the recording is completed. During the recording, the subject walks either on a treadmill or an indoor or outdoor ground.

The set of methods that we propose is explained in the following subsections, and the presentation is structured as follows. Separately for each foot, we use the recorded sensor data to separate phases in which the foot is in full contact with the ground from phases in which the foot moves, i.e., we detect when strides take place (section 3.3). For each detected stride, we then detect toe-off (section 3.5) and initial contact (section 3.6). The gait events from the ipsilateral and contralateral foot are combined to define gait phases. We calculate the relative duration of each gait phase and the cadence (section 3.7). We then estimate the sensor orientation by sensor fusion of the gyroscope and accelerometer readings (section 3.8) and double-integrate the acceleration to obtain a position trajectory (section 3.9). From this position trajectory, we obtain the stride length and the walking speed (section 3.10). **Figure 2** provides an overview of the proposed set of methods.

In the remainder of this section, we define parameters used by the method. For an overview of those parameters and proposed values, please refer to **Table 2** in section 4. Our aim is to define the parameters in a way that they are not sensitive to different gait styles or velocities. In section 4, we demonstrate that this approach works by only employing one common set of parameters for validation on a very broad data set with healthy and non-healthy subjects walking at different speeds.

3.1. Notation

Denote the accelerometer readings $\mathbf{a}(t_k) \in \mathbb{R}^3$ and the gyroscope readings $\boldsymbol{\omega}(t_k) \in \mathbb{R}^3$, sampled at times $t_k = kT_s, k \in \{1 \,..\, N\}$, $T_s \in \mathbb{R}_{>0}$.

In the following, all times t with any index are multiples of T_s. If any calculation yields a time that is not a multiple of T_s, we assume that this value is rounded to the nearest multiple of T_s and do not explicitly write this for the sake of a compact notation. Furthermore, any summation over τ should be interpreted as a summation with a non-integer step size of T_s, i.e., we simply write $\sum_{\tau=t_1}^{t_2} x(\tau)$ instead of the longer but mathematically precise notation $\sum_{k=k_1}^{k_2} x(t_k), k_1 = \frac{t_1}{T_s}, k_2 = \frac{t_2}{T_s}$.

Unit quaternions in vector notation are used to represent rotations and orientations (48). When a quaternion is used

to represent the sensor orientation, it is the rotation from an inertial reference frame with the z-axis pointing up (and arbitrary heading) to the coordinate system of the sensor. In the context of quaternion multiplication, which we denote by $\otimes$, three-dimensional vectors are implicitly regarded as quaternions with zero real part.

Furthermore, $\boldsymbol{v}^\intercal$ denotes the transpose of the vector $\boldsymbol{v}$.

3.2. Gait Events and Gait Phases

According to standard literature (44) and as illustrated in **Figure 3A**, the gait cycle starts at initial contact. Each stride can be separated into *stance* and *swing*. Stance consists of the gait phases *loading response*, *mid-stance*, *terminal stance*, and *pre-swing*. Swing can be separated into *initial swing*, *mid-swing*, and *terminal swing*. The combination of mid-stance and terminal stance is called *single limb support* and corresponds to the swing phase of the contralateral foot. In standard literature (44), the initial contact is commonly considered to be a very short gait phase with a duration of 2 %. As it is common practice in IMU-based gait analysis (15, 19, 21, 26), we define the initial contact as an event without duration. Note that sometimes the initial contact is also called foot strike (26) or heel strike (15).

The separation between stance and swing and the separation of stance into loading response, mid-stance, terminal stance, and pre-swing is defined based on three events that describe a change of ground contact of the feet: *initial contact*, *heel rise*, and *toe-off*. In contrast, the separation of swing into initial swing, mid-swing, and terminal swing is based on positional information of the feet and on the tibia orientation. The gait phases are defined based on bilateral events, i.e., the gait phase of the ipsilateral foot is not only described based on the events of the same (ipsilateral) foot but also based on toe-off and initial contact of the other (contralateral) foot.

We will now describe how we determine five of those gait phases (swing and the four sub-phases of stance) using IMUs in a two-step approach. First, we detect four gait events independently for each foot. We then use this gait event cycle of both feet to derive gait phases for each foot.

To this end, for each stride $i \in \{1 \ldots M\}$, we define the following events that we want to detect independently for the right and left foot from the raw measurement data of the corresponding IMU:

- initial contact – $t_{\mathrm{ic},i}$
- full contact – $t_{\mathrm{fc},i}$
- heel rise – $t_{\mathrm{hr},i}$
- toe-off – $t_{\mathrm{to},i}$.

Note that in addition to the three events used to define gait phase transitions in **Figure 3A**, we introduce an event called *full contact* that indicates that the foot is in full contact with the ground. For various processing steps, such as zero-velocity updates and position integration, we further define a rest instant $t_{\mathrm{rest},i}$ at the middle of the foot flat phase, i.e.,

$$t_{\mathrm{rest},i} \coloneqq \frac{1}{2}\left(t_{\mathrm{fc},i} + t_{\mathrm{hr},i}\right). \tag{1}$$

See **Figure 3B** for a plot of the raw accelerometer and gyroscope data measured during one stride along with a graphical representation of the gait event cycle defined by the introduced events. In the following subsections, we will describe in detail how we determine those time instants from the raw sensor data.

After having determined the gait events for both feet, we use the gait event cycles from both feet to determine the gait phase according to the commonly used definitions by (44). As shown in **Figure 4**, finite automata for the gait phases of the left and right foot are each driven by the gait event cycles of both feet.

Since time instances from both sensors are used for the definition of the gait phase transitions, both feet must be equipped with sensors, and precise time synchronization is required. However, note that the separation into stance and swing directly follows from the gait event cycle (as shown in **Figure 4**) and is independent of the contralateral foot. Therefore, we can determine stance and swing regardless of the synchronization between the sensors. This is also useful if only one foot is equipped with a sensor and facilitates on-chip data processing.

Note that the three sub-phases of stance in the gait event cycle hold further information that is not directly captured by the standard gait phase definitions as given in **Figure 3A**. We denote the phase from $t_{\mathrm{fc},i}$ to $t_{\mathrm{hr},i}$, in which the foot is fully on the ground, as *foot flat*. Note that the other two sub-phases of the stance phase, $t_{\mathrm{ic},i}$ to $t_{\mathrm{fc},i}$ and $t_{\mathrm{hr},i}$ to $t_{\mathrm{to},i}$, are sometimes called loading response and pre-swing (19, 21) but do not correspond to the phases with the same name as defined in standard literature (44).

Furthermore, as also shown in **Figure 4**, time-synchronized events from both feet also allow for the distinction of *double support*, *single support*, and *zero-contact* phases, which occur only during running (44).

3.3. Foot Flat Detection

As the first step of gait phase detection, the phases in which the foot is fully on the ground (foot flat) are detected. When the foot is fully on the ground, the Euclidean norm of the accelerometer readings will be close to $9.81\,\mathrm{m/s^2}$, and the norm of the gyroscope readings will be close to zero. During a stride, we typically will see an increase in the signal norms. However, it is possible that during the motion phase there are long periods with only small changes of velocity or small rotations. To obtain a robust stride detection, we, therefore, first find activity using either the accelerometer or the gyroscope readings and then combine this information.

For an acceleration-based rest signal $r_a(t_k)$, we consider the absolute difference of the norm from $9.81\,\mathrm{m/s^2}$,

$$a(t_k) \coloneqq \left|\|\boldsymbol{a}(t_k)\| - 9.81\right|, \tag{2}$$

and perform acausal thresholding using a threshold a_{th} and a hysteresis factor h_a by applying hysteresis in forward and

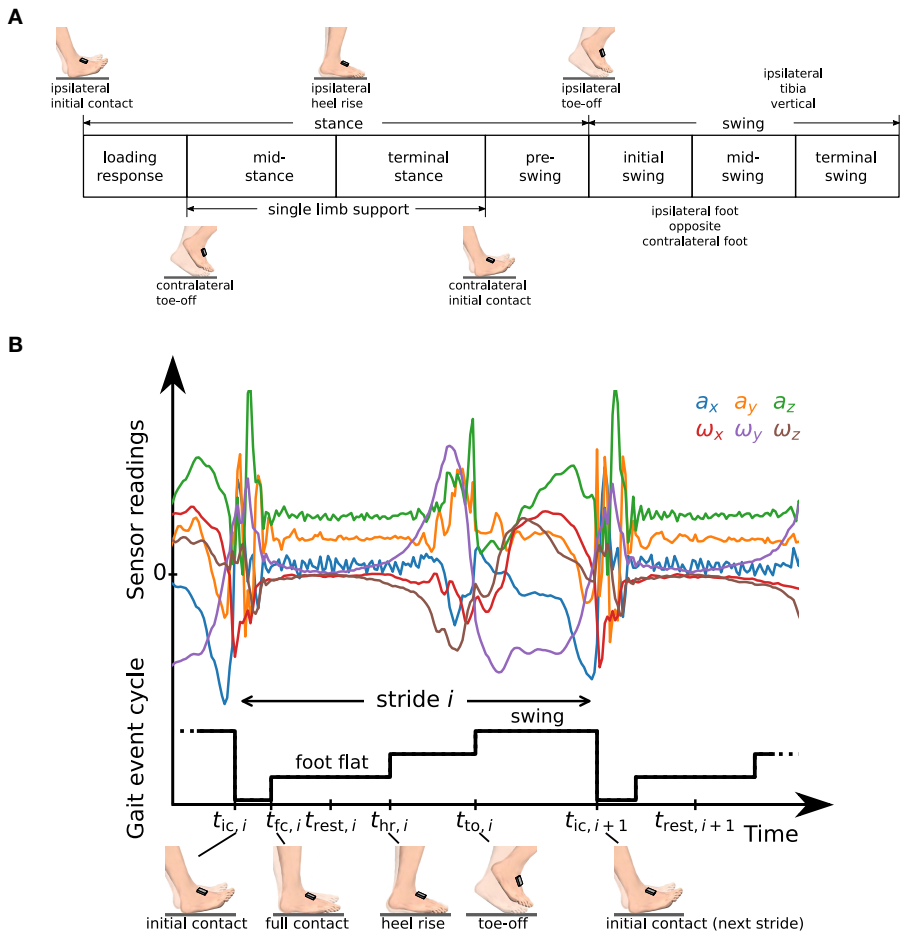


FIGURE 3 | (A) Definition of gait phases as used in standard literature [cf. (44)], and transitions based on gait events of the ipsilateral and contralateral foot. **(B)** Raw accelerometer and gyroscope sensor readings and representation of the gait event cycle with a staircase-shaped signal. We define time instants $t_{ic,i}$, $t_{fc,i}$, $t_{hr,i}$, $t_{to,i}$ that mark characteristic events and a rest instant $t_{rest,i}$ in the middle of the phase in which the foot is fully on the ground (foot flat).

backward direction, i.e.,

$$r_a^*(t_k) := \begin{cases} 1 & a(t_k) > (1+h_a)a_{\text{th}} \\ 0 & a(t_k) < (1-h_a)a_{\text{th}} \\ r_a(t_{k-1}) & \text{otherwise} \end{cases} \quad (3)$$

$$r_a(t_k) := \begin{cases} 1 & r_a^*(t_k) = 1 \\ 0 & a(t_k) < (1-h_a)a_{\text{th}} \\ r_a(t_{k+1}) & \text{otherwise} \end{cases} \quad (4)$$

with $r_a^*(0) = 0$ and $r_a(t_N) = r_a^*(t_N)$. In the resulting signal, zero-phases shorter than $T_{0,\min}$ are set to one, and afterward, one-phases shorter than $T_{1,\min}$ are set to zero.

The same acausal thresholding with the removal of short phases is applied to the gyroscope norm signal $\omega(t_k) := \|\boldsymbol{\omega}(t_k)\|$ using a threshold ω_{th} and hysteresis factor h_ω, which yields a gyroscope-based rest signal $r_\omega(t_k)$. See **Figure 5A** for an illustration of the thresholding method.

Both rest signals, $r_a(t_k)$ and $r_\omega(t_k)$, are combined into $r(t_k)$, which is set to one if at least one of the two signals is one. Afterward, zero-phases shorter than $T_{0,\min}$ are set to one, and then one-phases shorter than $2T_{1,\min}$ are set to zero. This process is illustrated in **Figure 5B**. Each zero-to-one transition of the resulting signal marks a heel rise $t_{hr,i}$, and each one-to-zero transition marks a full contact $t_{fc,i+1}$.

3.4. Automatic Threshold Adaptation

A common issue with thresholding approaches is that the thresholds have to be adapted based on gait velocity and also other gait and sensor characteristics (19, 20). Therefore, instead of performing the thresholding of the accelerometer and gyroscope norm using manually tuned thresholds a_{th} and ω_{th}, we propose an algorithm that automatically determines these thresholds for each trial based on the measured data.

The threshold a_{th} is determined using an iterative algorithm similar to (49), with l being the iteration index and w_a being a

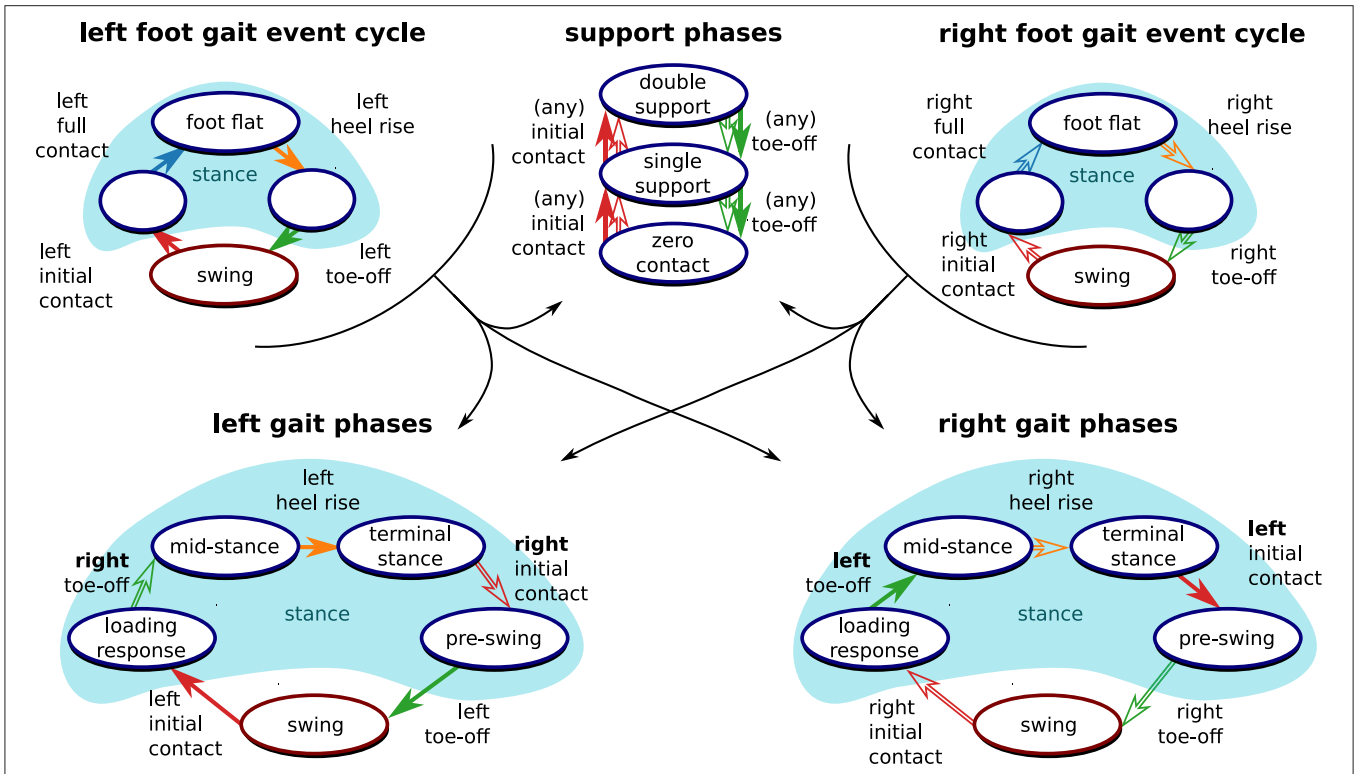


FIGURE 4 | Derivation of clinically relevant gait phases from the gait event cycles. Following standard literature (44), events from both the left foot (filled arrows) and the right root (outlined arrows) are necessary to define the gait phase of each foot. Furthermore, support phases based on the number of feet that are in contact with the ground can be defined based on the gait events.

weighting parameter:

$$a_{\text{th},0} = \frac{1}{2}\left(\max_{t_k\in[t_1,t_N]} a(t_k) + \min_{t_k\in[t_1,t_N]} a(t_k)\right) \tag{5}$$

$$T^+ = \{t_k \in [t_1, t_N] | a(t_k) > a_{\text{th},l}\} \tag{6}$$

$$T^- = \{t_k \in [t_1, t_N] | a(t_k) \leq a_{\text{th},l}\} \tag{7}$$

$$a_{\text{th},l+1} = \frac{w_a}{|T^-|}\sum_{t_k\in T^-} a(t_k) + \frac{1-w_a}{|T^+|}\sum_{t_k\in T^+} a(t_k). \tag{8}$$

We perform 200 iterations to ensure convergence, i.e., $a_{\text{th}} := a_{\text{th},200}$. **Figure 6** illustrates the result of this process. Further, we define a lower bound $a_{\text{th,min}}$ for this threshold.

Similarly, we determine the threshold ω_{th} based on the gyroscope norm $\omega(t_k)$ and a weighting factor w_ω.

3.5. Toe-Off Detection

After determining heel rise and full contact, we want to detect the beginning of the swing phase, i.e., the toe-off. During toe-off, the foot first rotates approximately along the mediolateral axis as the heel rises, then loses contact with the ground and rotates in the opposite direction. An inertial sensor attached to the foot cannot directly measure when the foot fully loses contact with the ground, in contrast to, e.g., pressure-sensitive walkways. Note that the accuracy of toe-off detection using pressure sensors also depends on calibration and the chosen thresholds (12).

As rotation can be measured precisely with IMUs, we exploit the fact that the direction of rotation of the foot changes when transitioning from the phase in which the heel rises while the toe stays on the ground to the phase in which the toe leaves the ground. This approach is commonly used in existing literature, as detailed in section 2. However, most methods directly rely on the angular rate measured in the sagittal plane and thereby require at least one sensor axis to be well-aligned with a functional axis of the foot.

To be independent of the sensor orientation and also to obtain a reliable detection if the subject exhibits strong inversion or eversion during toe-off, we define a signal called tilt-rate $\Gamma_i(t_k)$, from each heel rise $t_{\text{hr},i}$ to the subsequent full contact $t_{\text{fc},i+1}$, as

$$\Gamma_i(t_k) := \boldsymbol{\omega}(t_k)^\intercal \frac{\sum_{\tau=t_{\text{hr},i}}^{t_k}\boldsymbol{\omega}(\tau)}{\left\|\sum_{\tau=t_{\text{hr},i}}^{t_k}\boldsymbol{\omega}(\tau)\right\|}, t_k \in [t_{\text{hr},i}, t_{\text{fc},i+1}]. \tag{9}$$

The rationale behind the definition of the tilt-rate $\Gamma_i(t_k)$ is to identify the main axis of rotation since the last heel rise and compute the current rate of rotation around this main axis. This enables us to detect a zero-crossing of the main rotation without making any assumptions on the orientation of the sensor with respect to the foot.

In general, the tilt-rate $\Gamma_i(t_k)$ will exhibit a change of sign after a distinct peak (cf. **Figure 7A**). As there might be noise, leading to

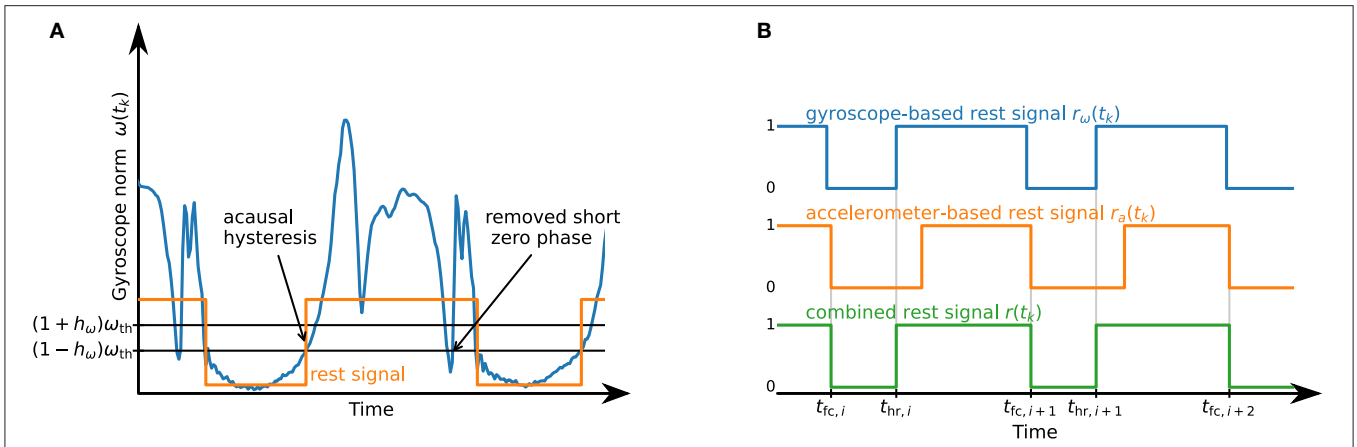


FIGURE 5 | (A) Illustration of the thresholding algorithm. Acausal hysteresis and the removal of short phases ensure the robust detection of the desired rest phase. **(B)** Illustration of the combination of $r_\omega(t_k)$ and $r_a(t_k)$ into $r(t_k)$. By using the *OR* combination of the accelerometer- and gyroscope-based signals, we are able to robustly detect when the foot is not fully on the ground.

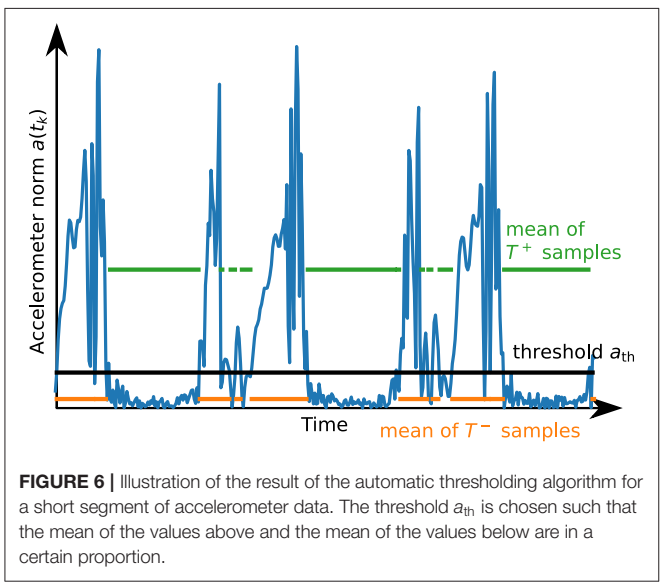


FIGURE 6 | Illustration of the result of the automatic thresholding algorithm for a short segment of accelerometer data. The threshold a_{th} is chosen such that the mean of the values above and the mean of the values below are in a certain proportion.

frequent sign changes right after $t_{hr,i}$, as well as large peaks later during the stride, we propose the following strategy to robustly determine the sign change of interest:

During the first half of the movement phase, let $\Gamma_{max,i}$ denote the maximum value of $\Gamma_i(t_k)$, i.e.,

$$\Gamma_{max,i} := \max_{t_k \in [t_{hr,i}, \frac{1}{2}(t_{hr,i}+t_{fc,i+1})]} \Gamma_i(t_k). \tag{10}$$

We then find the first time instant for which $\Gamma_i(t_k) \geq \frac{1}{2}\Gamma_{max,i}$. Starting from this time instant, we find the first time instant at which $\Gamma_i(t_k) \leq 0$. We assume this time instant to be the toe-off $t_{to,i}$, i.e., the start of the swing phase. **Figure 7A** illustrates this process.

Note that $t_{to,i}$ is defined based on a feature of the rotation of the foot and not directly as the lift-off of the toes. Using the maximum of the tilt rate (or any weighted average of the maximum and zero-crossing time instant) are also plausible approaches.

3.6. Initial Contact Detection

The initial contact marks the beginning of the loading response and can be detected by the jerk, i.e., the change of acceleration, caused by the foot touching the ground. We calculate the jerk using the first-order backward difference approximation, i.e.,

$$\boldsymbol{j}(t_k) := \frac{1}{T_s}\left(\boldsymbol{a}(t_k) - \boldsymbol{a}(t_{k-1})\right). \tag{11}$$

For every stride, we only consider a sub-window of the phase between toe-off and the beginning of the subsequent foot-flat phase and denote the start time of this window as $t_{win,i} := j_{win}t_{to,i-1} + (1 - j_{win})t_{fc,i}$, $j_{win} \in [0, 1]$. In this time window, we first determine the maximum value of the jerk norm, i.e.,

$$j_{max,i} := \max_{t_k \in [t_{win,i}, t_{fc,i}]} \|\boldsymbol{j}(t_k)\|. \tag{12}$$

We then mark the first time instant in this window with $\|\boldsymbol{j}(t_k)\| \geq j_{th}j_{max,i}$ as the start of the loading response $t_{ic,i}$. See **Figure 7B** for an illustration of the initial contact detection.

3.7. Stride and Gait Phase Durations and Cadence

For each detected stride, we calculate the stride duration as the duration from one initial contact to the subsequent initial contact of the same foot, i.e.,

$$T_{stride,i} := t_{ic,i+1} - t_{ic,i}. \tag{13}$$

For each detected stride, the duration of the swing phase is the time between toe-off and initial contact of the subsequent stride, i.e.,

$$T_{swing,i} := t_{ic,i+1} - t_{to,i}. \tag{14}$$

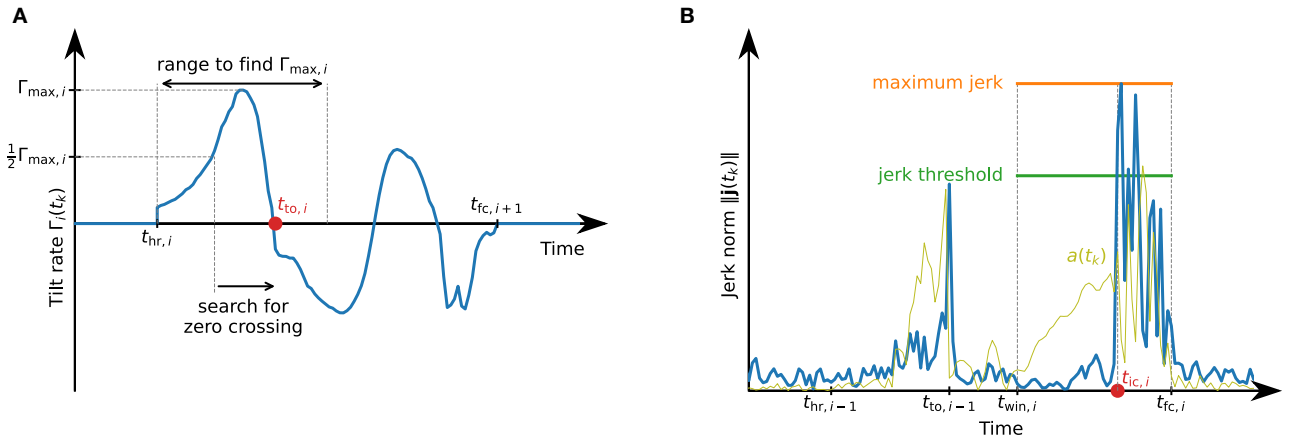


FIGURE 7 | Detection of toe-off and initial contact events that define the swing phase. **(A)** Illustration of the toe-off detection. Between heel rise and full contact, the tilt rate might exhibit multiple local maxima and zero-crossings. For a robust detection of the correct zero-crossing, we first find the maximum value during the first half of the phase from $t_{\text{hr},i}$ to $t_{\text{fc},i+1}$ and search for the first zero-crossing after the tilt rate has reached half of this maximum. **(B)** Illustration of the initial contact detection based on the jerk norm. Note how the jerk norm reflects the sudden change when the foot touches the ground much better than the accelerometer norm signal $a(t_k)$.

The stance duration is the remaining duration of the stride:

$$T_{\text{stance},i} := T_{\text{stride},i} - T_{\text{swing},i}. \tag{15}$$

Since relative gait phase durations are easier to interpret, we calculate

$$T_{\text{swing,rel},i} := \frac{T_{\text{swing},i}}{T_{\text{stride},i}}, \tag{16}$$

$$T_{\text{stance,rel},i} := \frac{T_{\text{stance},i}}{T_{\text{stride},i}}. \tag{17}$$

Similarly, for every stride we calculate relative gait phase durations for loading response $T_{\text{lr,rel},i}$, single limb support $T_{\text{sl,rel},i}$, terminal stance $T_{\text{ts,rel},i}$, and pre-swing $T_{\text{ps,rel},i}$, based on the bilateral gait phases as defined in **Figure 3A**. Note that, analogously, we can also calculate absolute and relative durations for all other gait phases defined in **Figure 3A**.

To calculate the cadence, we multiply the inverse of the stride duration by two in order to express the cadence as the number of steps per minute instead of strides per minute, i.e.,

$$c_i := \frac{2}{T_{\text{stride},i}}. \tag{18}$$

3.8. Orientation Estimation

By fusing the gyroscope and accelerometer measurements, we obtain an estimate of the sensor orientation with respect to a global frame that has a vertical z-axis and an arbitrary heading.

Starting with an arbitrary initial orientation $\mathbf{q}_\omega(0)$, e.g., $\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}^{\mathrm{T}}$, we perform gyroscope strapdown integration

$$\mathbf{q}_\omega(t_k) := \mathbf{q}_\omega(t_{k-1}) \otimes \left[\cos\left(\tfrac{T_s}{2}\|\boldsymbol{\omega}(t_k)\|\right) \quad \frac{\boldsymbol{\omega}^{\mathsf{T}}(t_k)}{\|\boldsymbol{\omega}(t_k)\|}\sin\left(\tfrac{T_s}{2}\|\boldsymbol{\omega}(t_k)\|\right)\right]^{\mathsf{T}}. \tag{19}$$

Using this orientation, we transform the measured acceleration into a (slowly drifting) inertial frame, i.e.,

$$\mathbf{a}_\omega(t_k) := \mathbf{q}_\omega(t_k) \otimes \boldsymbol{a}(t_k) \otimes \mathbf{q}_\omega(t_k)^{-1}. \tag{20}$$

In the rotating sensor frame, the gravitational acceleration can point in different directions depending on sensor orientation. In the inertial frame, however, the gravitational acceleration will point in (almost) the same direction regardless of the sensor orientation, and, when integrating, acceleration and deceleration will cancel out. Exploiting this property, we low-pass filter each component of $\mathbf{a}_\omega(t_k)$ by applying a moving average filter with a window length of T_a in forward and reverse direction. Assuming that the change of velocity over the filter window length is small, the resulting filtered acceleration will be dominated by the gravitational acceleration. This filtered acceleration $\mathbf{a}_{\omega,\text{f}}(t_k)$ is then transferred back to the sensor frame

$$\boldsymbol{a}_{\text{f}}(t_k) := \mathbf{q}_\omega(t_k)^{-1} \otimes \mathbf{a}_{\omega,\text{f}}(t_k) \otimes \mathbf{q}_\omega(t_k). \tag{21}$$

We then correct the inclination of the gyroscope strapdown integration quaternion $\mathbf{q}_\omega(t_k)$ by using the filtered acceleration as a vertical reference. To this end, we transform the filtered acceleration into the global frame

$$\boldsymbol{a}_{\text{r}}(t_k) := \mathbf{q}_a(t_{k-1}) \otimes \mathbf{q}_\omega(t_k) \otimes \boldsymbol{a}_{\text{f}}(t_k) \otimes (\mathbf{q}_a(t_{k-1}) \otimes \mathbf{q}_\omega(t_k))^{-1}, \tag{22}$$

with $\mathbf{q}_a(0) := \begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}^{\mathrm{T}}$, and correct the inclination

$$\boldsymbol{n}(t_k) := \boldsymbol{a}_{\text{r}}(t_k) \times \begin{bmatrix}0 & 0 & 1\end{bmatrix}^{\mathrm{T}} \tag{23}$$

$$\alpha(t_k) := \arccos\left(\begin{bmatrix}0 & 0 & 1\end{bmatrix}^{\mathrm{T}} \frac{\boldsymbol{a}_{\text{r}}(t_k)}{\|\boldsymbol{a}_{\text{r}}(t_k)\|}\right) \tag{24}$$

$$\mathbf{q}_a(t_k) := \mathbf{q}_a(t_{k-1}) \otimes \left[\cos\left(\tfrac{\alpha(t_k)}{2}\right) \quad \frac{\boldsymbol{n}(t_k)}{\|\boldsymbol{n}(t_k)\|}\sin\left(\tfrac{\alpha(t_k)}{2}\right)\right]^{\mathsf{T}}. \tag{25}$$

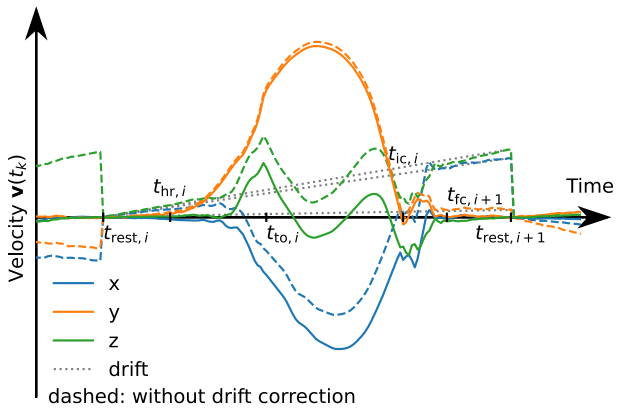


FIGURE 8 | Velocity trajectories with (solid) and without (dashed) linear drift correction. The dotted line represents the subtracted linear drift approximation for stride *i*. For demonstration purposes, the drift has been artificially increased by a factor of 10.

Multiplication of the gyroscope strapdown integration quaternion and the accelerometer correction quaternion yields the sensor orientation,

$$\mathbf{q}(t_k) := \mathbf{q}_a(t_k) \otimes \mathbf{q}_\omega(t_k). \tag{26}$$

3.9. Foot Velocity and Position Tracking

Using the estimated orientation, we perform double integration of the measured accelerations to estimate the length of each stride, i.e., the horizontal displacement between two adjacent foot-flat phases.

To integrate accelerations, they are first transformed into the reference frame

$$\mathbf{a}_\varepsilon(t_k) := \mathbf{q}(t_k) \otimes \boldsymbol{a}(t_k) \otimes \mathbf{q}(t_k)^{-1}. \tag{27}$$

Assuming that the velocity is zero in the middle of the foot-flat phase, i.e., at $t_{\mathrm{rest},i}$, we integrate those accelerations for each stride which yields a velocity

$$\mathbf{v}_i(t_k) := T_\mathrm{s} \sum_{\tau=t_{\mathrm{rest},i}}^{t_k} \left(\mathbf{a}_\varepsilon(\tau) - \begin{bmatrix}0 & 0 & 9.81\end{bmatrix}^\mathrm{T}\right), \quad t_k \in [t_{\mathrm{rest},i}, t_{\mathrm{rest},i+1}]. \tag{28}$$

Due to measurement errors, mainly accelerometer bias, this velocity is usually not zero at $t_{\mathrm{rest},i+1}$ even if the foot is perfectly at rest. Therefore, we correct this drift linearly over the time duration of the stride:

$$\mathbf{v}_{\mathrm{df},i}(t_k) := \mathbf{v}_i(t_k) - \frac{t_k - t_{\mathrm{rest},i}}{t_{\mathrm{rest},i+1} - t_{\mathrm{rest},i}} \mathbf{v}_i(t_{\mathrm{rest},i+1}). \tag{29}$$

See **Figure 8** for an example velocity trajectory with and without drift correction.

By integrating this drift-free velocity over the stride duration, we obtain a position trajectory,

$$\mathbf{p}_i(t_k) := T_\mathrm{s} \sum_{\tau=t_{\mathrm{rest},i}}^{t_k} \mathbf{v}_{\mathrm{df},i}(\tau) =: \begin{bmatrix}p_{i,x}(t) & p_{i,y}(t) & p_{i,z}(t)\end{bmatrix}^\mathrm{T}. \tag{30}$$

3.10. Stride Length and Walking Speed

We calculate the stride length L_i as the horizontal displacement during the stride *i*. Since $\mathbf{p}_i(t_{\mathrm{rest},i}) = 0$,

$$L_i := \sqrt{p_{i,x}(t_{\mathrm{rest},i+1})^2 + p_{i,y}(t_{\mathrm{rest},i+1})^2}. \tag{31}$$

Note that this method does not make any assumption on the orientation in which the sensor is attached to the foot. Also, note that we integrate from $t_{\mathrm{rest},i}$ to $t_{\mathrm{rest},i+1}$ and not from $t_{\mathrm{ic},i}$ to $t_{\mathrm{ic},i+1}$ since this makes the zero-velocity assumption more robust.

By dividing the stride length by the stride duration, we obtain the walking speed,

$$v_i := \frac{L_i}{T_{\mathrm{stride},i}}. \tag{32}$$

3.11. Summary of the Estimated Parameters

After performing all steps presented above, the set of proposed methods provides the time instants of the defined gait events, the sensor orientation quaternion for each time instant, and velocity and position trajectories. From those time-based signals, the following gait parameters are extracted for each stride *i*:

- swing duration $T_{\mathrm{swing,rel},i}$ [%]
- stance duration $T_{\mathrm{stance,rel},i}$ [%]
- analogously, relative durations for the other gait phases as defined in **Figure 4**
- stride length L_i [cm]
- walking speed v_i [km/h]
- cadence c_i [steps/min]

Note that all quantities are calculated separately for each stride of each foot. In many cases, only the mean of those values over multiple steps will be of interest. However, this stepwise calculation also allows for analysis of the variance and the detection of trends.

The accuracy of those gait parameters is validated in the next section.

4. EXPERIMENTAL VALIDATION

In this section, we aim to show that the less restrictive IMU-based setup combined with the methods proposed in section 3 is able to determine the same parameters as stationary systems that are used in clinical practice while providing similar accuracy. To this end, with a large data set consisting of three different subject groups, we compare the parameters calculated by the proposed methods with values reported by instrumented treadmills.

4.1. Setup

One PABLO® Lower Extremity inertial sensor (Tyromotion GmbH, Graz, Austria) was attached to each shoe (cf. **Figure 9A**). The sensors measure angular rate and acceleration at a sampling frequency of 110 Hz. Each sensor has a size of 56 × 34 × 21 mm and transmits the data wirelessly using Bluetooth. The sensors were attached to the subjects' shoes with special Velcro straps.

FIGURE 9 | Experimental setup. **(A)** Patient with inertial sensors attached to the shoe. **(B)** Instrumented treadmill at NTK Kapfenberg. Gait parameters are derived from the measurement data of the inertial sensors with the proposed methods and validated against parameters obtained from the instrumented treadmill serving as ground truth.

Zebris Rehawalk instrumented treadmills (Zebris Medical, Isny, Germany) were used as reference systems. Since the data collection took place in various institutions (FH Joanneum Graz, NTK Kapfenberg, Rehabilitation Center Kitzbühel), different systems with identical function were used. See **Figure 9B** for a picture of the setup at NTK Kapfenberg.

- FH Joanneum (Graz, Austria)
  - Treadmill: h-p-c Mercury Med Treadmill (HP Cosmos, Nussdorf, Germany), walking speed: 0–22 km/h in 0.1 km/h steps, walking surface: 150 × 50 cm
  - Pressure measuring platform: FDM-THM-M-3i (Zebris Medical, Isny, Germany), 120 Hz, sensor area: 108.4 × 47.4 cm, 7,168 sensors.
- NTK (Kapfenberg, Austria)
  - Treadmill: h-p-c Locomotion Med Treadmill (HP Cosmos, Nussdorf, Germany), walking speed: 0–10 km/h in 0.1 km/h steps, walking surface: 150 × 50 cm
  - Pressure measuring platform: FDM-THM-M-2i (Zebris Medical, Isny, Germany), 120 Hz, sensor area: 111.8 × 49.5 cm, 3,432 sensors.
- Rehabilitation Center Kitzbühel (Kitzbühel, Austria)
  - Treadmill: h-p-c Mercury Med Treadmill (HP Cosmos, Nussdorf, Germany), walking speed: 0–22 km/h in 0.1 km/h steps, walking surface: 150 × 50 cm
  - Pressure measuring platform: FDM-THM-M-2i (Zebris Medical, Isny, Germany), 120 Hz, sensor area: 111.8 × 49.5 cm, 3,432 sensors.

4.2. Subjects and Experimental Procedure

The data collection was carried out in three different institutions with different groups of subjects. Approval from the ethics committee of the University of Graz was obtained, and an informed consent form was signed by all participants.

Healthy participants were recorded at three different walking speeds, each for two minutes: 1.5, 3, and 5 km/h. A prerequisite for participation was the ability to walk on a treadmill at different speeds. The healthy participants ($n = 39$) were recruited from the students at the Physiotherapy Institute of FH Joanneum Graz.

Non-healthy participants with affected ability to walk were asked to walk on a treadmill at a self-selected comfortable walking speed. Patients who were unable to walk on a treadmill were excluded during participant selection. The following set of participants were recruited:

- Participants with different neurological diseases ($n = 36$) were recruited from patients who were in neurological inpatient rehabilitation at NTK Kapfenberg at the time of data collection. This comprises 20 post-stroke patients, 6 patients with Parkinson's disease, two with multiple sclerosis, two with meningioma, two after polytrauma, and one patient each with epilepsy, spinocerebellar ataxia, low back pain, and polyneuropathy.
- Participants with various orthopedic diseases ($n = 62$) were recruited from the patients who were in orthopedic inpatient rehabilitation at Rehazentrum Kitzbühel at the time of data collection. Of these, four patients had pathologies in the area of the ankle or lower leg (e.g., ankle joint fractures, tibia fractures), 21 patients at the knee (e.g., osteoarthritis, total knee arthroplasty), 18 patients in the area of the thigh and hip (e.g., osteoarthritis, total hip arthroplasty, femur fractures), 16 patients in the area of the lumbar spine (low back pain, lumbar vertebrae fractures) as well as three patients in whom different body areas were affected (polytrauma, polymyositis).

All participants had time to get used to walking on the treadmill prior to the data collection. All participants were free to use the treadmill support (handrail, fall protection system). For the data collection, two minutes of walking was recorded simultaneously by both systems. IMU data was recorded with a tool of the TyroS software (Tyromotion, Graz, Austria) that allows the export of raw gyroscope and accelerometer data. Zebris data was recorded, analyzed, and exported with the software FDM v1.18.38 (Zebris Medical, Isny, Germany).

4.3. Data Processing

For each trial, we obtain the following gait parameters from the Zebris Rehawalk instrumented treadmill:

- loading response duration

TABLE 2 | Parameter values used for the proposed IMU-based methods.

Symbol	Description	Value
h_a	Hysteresis factor for acceleration	0.23
h_ω	Hysteresis factor for angular rate	0.23
w_a	Factor for a_{th} auto-tuning	0.85
$a_{\text{th,min}}$	Lower bound for a_{th}	1.8 m/s^2
w_ω	Factor for ω_{th} auto-tuning	0.8
$\omega_{\text{th,min}}$	Lower bound for ω_{th}	0 rad/s
$T_{0,\text{min}}$	Minimum duration of zero-phase	120 ms
$T_{1,\text{min}}$	Minimum duration of one-phase	180 ms
j_{win}	Ratio of the window to look for initial contact	0.7
j_{th}	Threshold for jerk norm (relative to maximum)	0.95
T_a	Time constant for acceleration moving average filter	8.0 s

This parametrization is used for the processing of all trials, regardless of gait pathology, walking speed, or style, in order to show that the method works well without tuning the parameters for specific gait characteristics.

- single limb support duration
- pre-swing duration
- swing duration
- stride length
- walking speed
- cadence.

These parameters are reported as averages over the whole trial. The gait phase durations are relative to the stride duration and reported separately for the left and right foot. We add the loading response, single limb support, and pre-swing durations to obtain the stance duration (cf. **Figure 3A**).

From phases in which the treadmill is not moving and the foot is resting on the ground for approximately 5 s at the beginning and end of each trial, gyroscope turn-on bias is automatically estimated and removed. Using the methods described in section 3, each recorded trial is processed with the parameter values given in **Table 2**. Note that we use the same set of parameters for all different subject groups and walking speeds in order to demonstrate that the method works well without adjusting the parameters for the specific gait velocity and style.

The sensor attachment used for recording the data sets, as shown in **Figure 9A**, ensures that one sensor axis is always roughly aligned with the mediolateral axis of the foot. To show that the proposed methods do not make assumptions regarding the sensor orientation, we simulate a random sensor attachment by multiplying all gyroscope and accelerometer measurements with a random rotation matrix that is different for each trial.

Finally, we calculate the same gait parameters as reported by the reference system by averaging the respective parameters, excluding the first and last three strides of each foot, and compare the resulting values to the values reported by the Zebris system. The results are found in the following section.

4.4. Results

For each trial, we first consider the five main parameters stance duration, swing duration, stride length, walking speed, and cadence, and evaluate the difference between the proposed methods (IMU) and the Zebris Rehawalk reference system (REF). The results are presented separately for each of the three subject groups in scatter plots and Bland-Altman plots (50) and can be found in **Figure 10**, for the healthy participants walking at three different speeds; in **Figure 11**, for the participants with orthopedic diseases; and in **Figure 12**, for the participants with neurological diseases.

The error (mean ± standard deviation) for the relative stance duration is 1.04 ± 1.34 % for healthy subjects, −0.29 ± 1.52 % for orthopedic patients, and 2.06 ± 1.63 % for neurological patients. For relative swing duration, the errors are −1.01 ± 1.35 % for healthy subjects, 0.32 ± 1.54 % for orthopedic patients, and −2.02 ± 1.64 % for neurological patients. This means that the average swing/stance duration error is in the range of 1–2 % for all subject groups.

For the stride length, the errors are −1.59 ± 1.53, −1.74 ± 1.63, and 0.51 ± 1.37 cm for healthy subjects, orthopedic patients, and neurological patients, respectively. This means that the average stride length error is below 2 cm for all subject groups.

The mean errors and standard deviations for the walking speed are −0.02 ± 0.05 km/h for healthy subjects, −0.03 ± 0.05 km/h for orthopedic patients, and 0.03 ± 0.03 km/h for means that the average walking speed error is below 0.05 km/h for all subject groups.

The cadence estimates show deviations of 0.68 ± 0.56 steps/min for healthy subjects, 0.55 ± 0.47 steps/min for orthopedic patients, and 0.57 ± 0.51 steps/min for neurological patients. This means that the average cadence error is below 1 step/min for all subject groups.

As an additional evaluation metric, we calculate the mean of the absolute difference (MAD) between the values reported by Zebris and the IMU-based analysis over all trials. **Table 3** summarizes the results for the three subject groups and all 215 evaluated trials.

The MAD of the stance and swing durations are ∼1.3 % for healthy subjects and orthopedic patients and 2.2 % for neurological patients. Note that we also evaluated the differences for the three sub-phases of stance that the Zebris Rehawalk reference system reports, i.e., loading response, single limb support, and pre-swing. **Table 3** shows that we can estimate the duration of those phases with the same accuracy as stance and swing.

To summarize, for all subject groups, the MAD is in the range of 1–2 % for gait phase durations, below 2 cm for the stride length, below 0.05 km/h for the walking speed, and below 1 step/min for the cadence.

5. DISCUSSION

In the present contribution, we have proposed a set of methods for spatiotemporal gait analysis based on two inertial sensors attached to the feet. Our methods allow for the calculation of

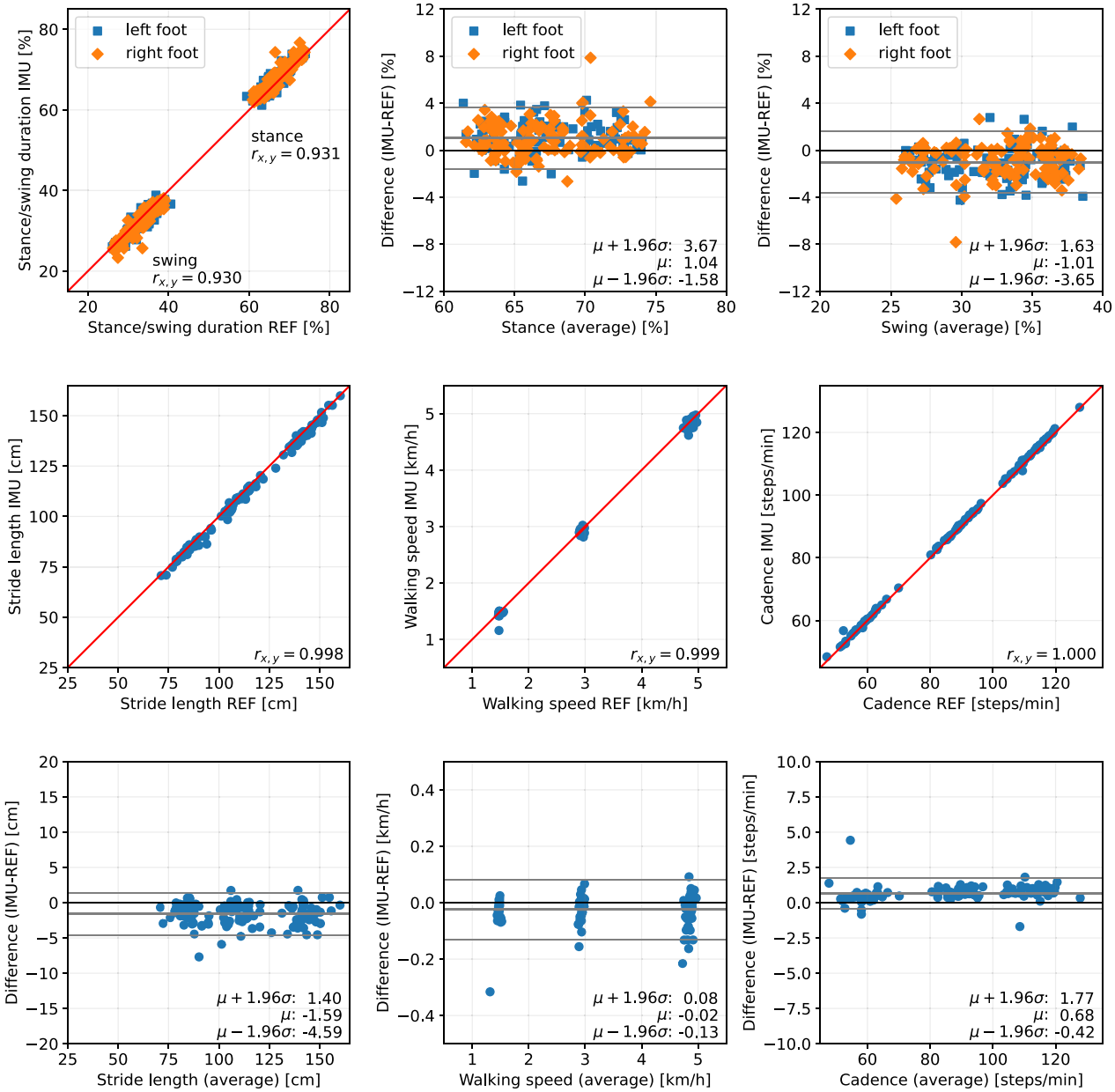


FIGURE 10 | Scatter plots and Bland-Altman plots for stance and swing duration, stride length, walking speed, and cadence of 39 healthy subjects walking at 1.5, 3, and 5 km/h. Red: 45-degree lines ($y = x$). Values obtained with the proposed IMU-based methods (IMU) are compared to the ground truth from the Zebris reference system (REF). The average deviation is ∼1 % for gait phase durations, below 2 cm for the stride length, below 0.05 km/h for the walking speed, and below 1 step/min for the cadence.

the main spatiotemporal gait parameters that are also reported by stationary laboratory systems: gait phase durations, stride length, walking speed, and cadence. Using a large data set consisting of healthy subjects walking at three different speeds, subjects with orthopedic diseases, and subjects with neurological diseases, we have validated the calculation of those parameters, using a Zebris Rehawalk instrumented treadmill as reference. All parameters show a very strong correlation (Pearson's r between 0.83 and 0.99, $p < 0.01$) (51). **Figures 10–12** display consistent results over this large and diverse group of subjects. Averaged over all trials, the MAD with respect to the reference system is 1.4 % for the gait phase durations, 1.7 cm for the stride length, 0.04 km/h for the walking speed, and 0.7 steps/min for the cadence.

In clinical practice and research, the presented parameters are used to quantify gait abnormalities and to document changes in the walking behavior of patients. Associations between spatiotemporal gait parameters and functional capacity, or increased mortality, have been demonstrated (52–54). A positive

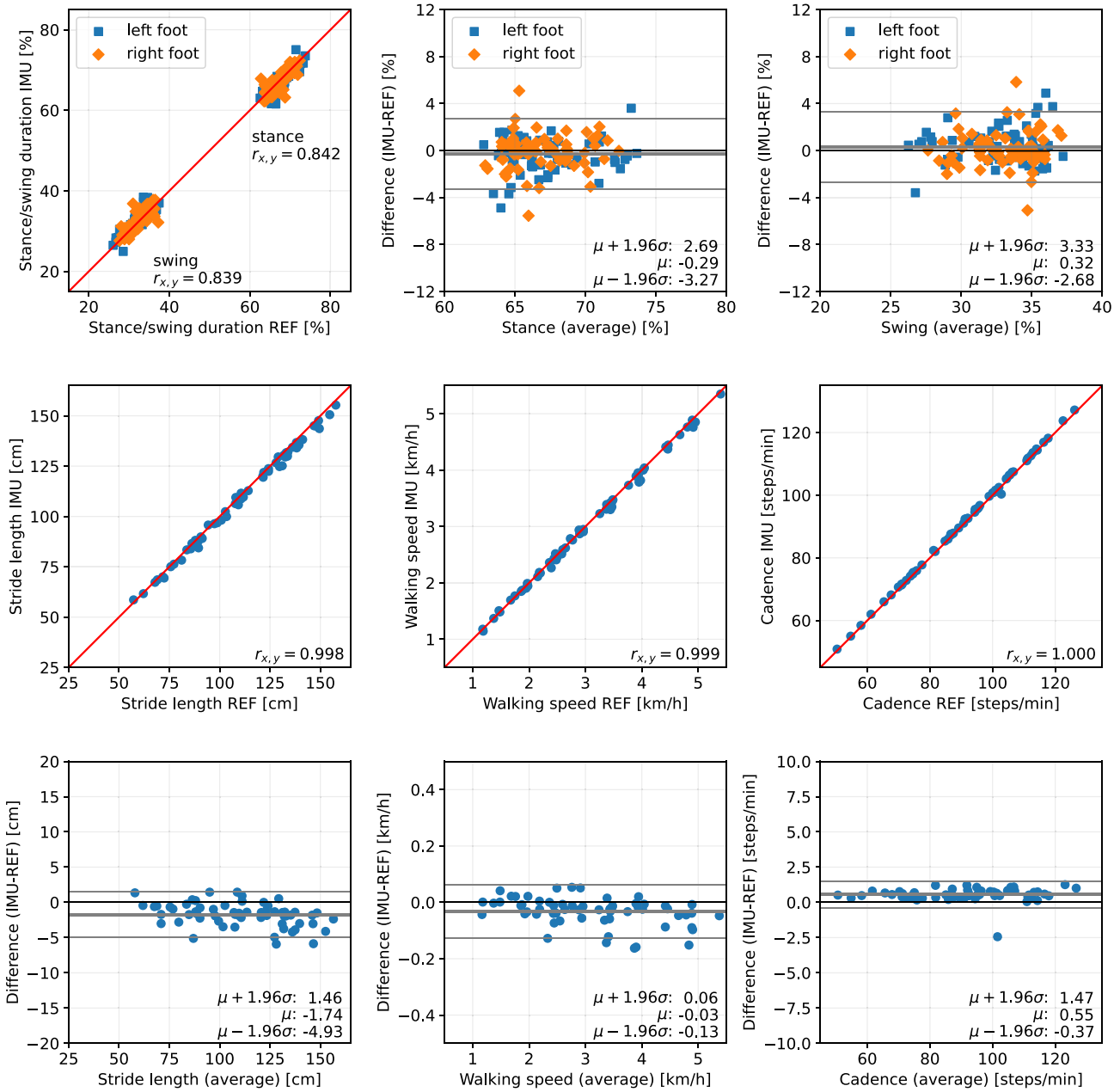


FIGURE 11 | Scatter plots and Bland-Altman plots for stance and swing duration, stride length, walking speed, and cadence of 62 orthopedic patients. Red: 45-degree lines ($y = x$). Values obtained with the proposed IMU-based methods (IMU) are compared to the ground truth from the Zebris reference system (REF). The average deviation is below 1 % for gait phase durations, below 2 cm for the stride length, below 0.05 km/h for the walking speed, and below 1 step/min for the cadence.

correlation with cardiovascular-related mortality was found for cadence (55). A reduction in walking speed has been shown to correlate with fall risk, frequency of hospitalization, and mortality (56–58). Stride length describes a strong correlation with walking speed according to the research of (59). Slower walking speed, altered gait phase duration, and increased variability of walking increase the risk of falls (60). Furthermore, it was found that psychological modalities, such as fear of falling, can also influence stride length and gait phase duration (61). The minimal clinically important difference (MCID) can be used to determine how precisely these changes must be detected in order to make a statement about their relevance. Despite thorough research, specific values for the MCID could only be found for the walking speed, ranging from 0.36 to 0.72 km/h (62–64). For IMU-based measurement with the proposed methods, the smallest detectable change (SDC) for walking speed is 0.21 km/h and clearly within

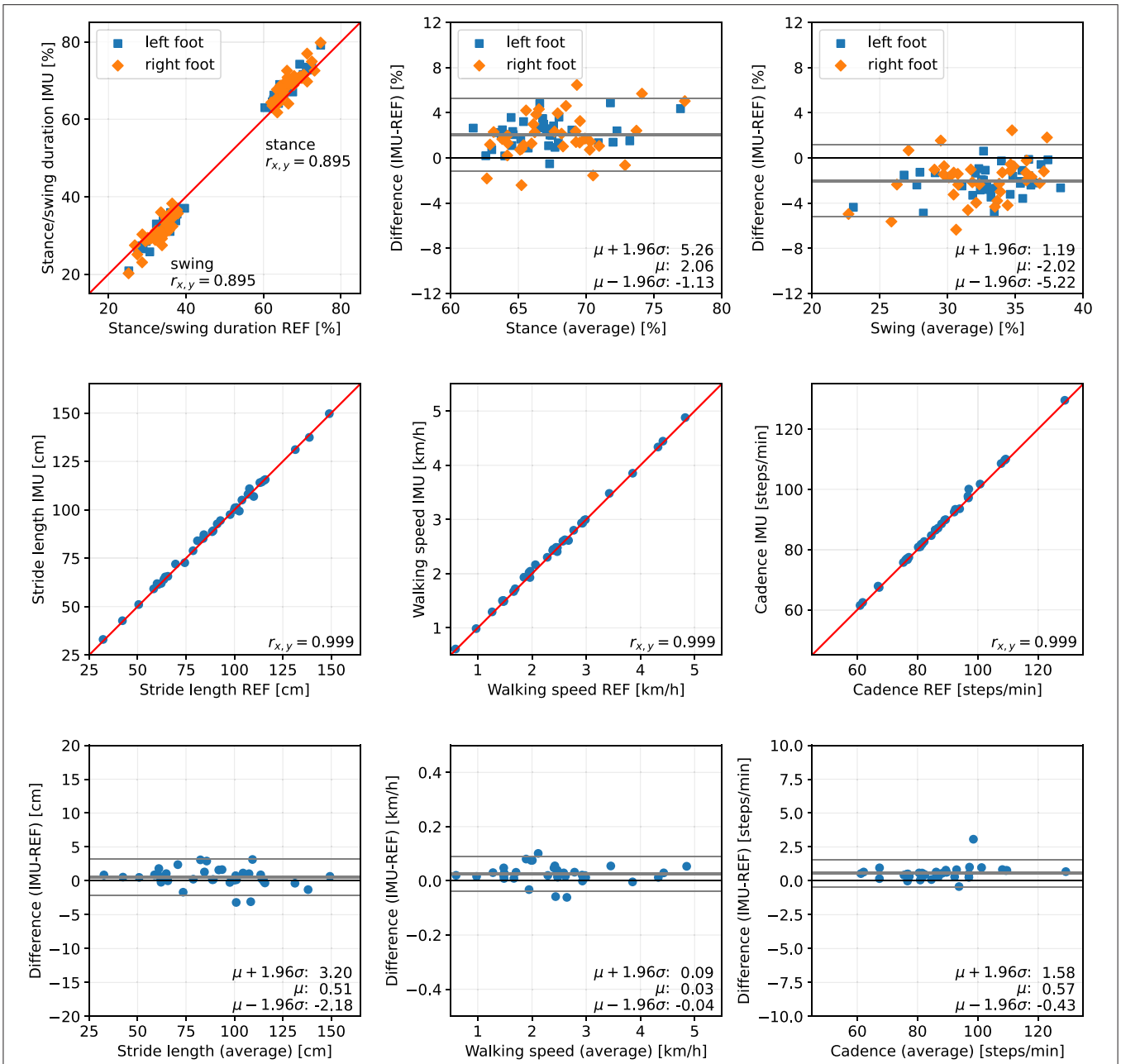


FIGURE 12 | Scatter plots and Bland-Altman plots for stance and swing duration, stride length, walking speed, and cadence of 36 neurological patients. Red: 45-degree lines ($y = x$). Values obtained with the proposed IMU-based methods (IMU) are compared to the ground truth from the Zebris reference system (REF). The average deviation is ~2 % for gait phase durations, below 1 cm for the stride length, below 0.05 km/h for the walking speed, and below 1 step/min for the cadence.

the MCID for all examined groups. For the other parameters, no reported MCID values could be found, which is consistent with the statement of (29).

The SDC for the cadence is 2.01 steps/min across all studied groups of subjects. This allows for much more accurate changes to be detected than those described as relevant in the literature [e.g., reduction in cadence of 10 steps per minute increases mortality by 4 % (65)]. The achieved SDC for stride length of 5.3 cm in the patients with neurological diseases seems to be sufficiently accurate to capture the differences occurring, for example, in Parkinson's disease (66). The stance and swing phase durations show an SDC of 6.5 % across all trials.

Unlike many existing contributions, we showed that the proposed methods reliably work on patients in addition to healthy subjects and still produce accurate results. This is noteworthy since it has been shown that pathological walking deteriorates the accuracy of many gait analysis methods (45)

TABLE 3 | Deviation between IMU-based and Zebris gait parameters.

	Stance [%]				Swing [%]	Stride	Walking	Cadence
		LR [%]	SLS [%]	PS [%]		length [cm]	speed [km/h]	[steps/min]
Healthy subjects ($n = 39$)								
MAD	**1.32**	**1.29**	**1.28**	**1.32**	**1.31**	**1.73**	**0.04**	**0.74**
$\mu \pm \sigma$	1.04 ± 1.34	0.97 ± 1.34	-0.96 ± 1.33	1.03 ± 1.34	-1.01 ± 1.35	-1.59 ± 1.53	-0.02 ± 0.05	0.68 ± 0.56
$r_{x,y}$	0.93	0.93	0.93	0.93	0.93	> 0.99	> 0.99	> 0.99
LoA	−1.58 to 3.67	−1.65 to 3.59	−3.56 to 1.65	−1.60 to 3.66	−3.65 to 1.63	−4.59 to 1.40	−0.13 to 0.08	−0.42 to 1.77
SDC	5.25	5.24	5.21	5.26	5.28	6.00	0.21	2.19
Orthopedic patients ($n = 62$)								
MAD	**1.14**	**1.12**	**1.14**	**1.07**	**1.16**	**1.94**	**0.04**	**0.63**
$\mu \pm \sigma$	-0.29 ± 1.52	-0.33 ± 1.49	0.35 ± 1.49	-0.31 ± 1.44	0.32 ± 1.54	-1.74 ± 1.63	-0.03 ± 0.05	0.55 ± 0.47
$r_{x,y}$	0.84	0.84	0.85	0.85	0.83	> 0.99	> 0.99	> 0.99
LoA	−3.27 to 2.69	−3.26 to 2.60	−2.57 to 3.26	−3.13 to 2.51	−2.68 to 3.33	−4.93 to 1.46	−0.13 to 0.06	−0.37 to 1.47
SDC	5.96	5.86	5.84	5.65	6.02	6.39	0.19	1.84
Neurological patients ($n = 36$)								
MAD	**2.26**	**2.20**	**2.23**	**2.21**	**2.22**	**1.09**	**0.03**	**0.60**
$\mu \pm \sigma$	2.06 ± 1.63	2.04 ± 1.65	-2.04 ± 1.64	2.06 ± 1.65	-2.02 ± 1.64	0.51 ± 1.37	0.03 ± 0.03	0.57 ± 0.51
$r_{x,y}$	0.89	0.89	0.89	0.89	0.89	> 0.99	> 0.99	> 0.99
LoA	−1.13 to 5.26	−1.19 to 5.28	−5.26 to 1.18	−1.16 to 5.29	−5.22 to 1.19	−2.18 to 3.20	−0.04 to 0.09	−0.43 to 1.58
SDC	6.39	6.47	6.44	6.46	6.41	5.38	0.13	2.01
All trials (215 trials)								
MAD	**1.43**	**1.39**	**1.40**	**1.40**	**1.41**	**1.68**	**0.04**	**0.68**
$\mu \pm \sigma$	0.83 ± 1.65	0.78 ± 1.65	-0.76 ± 1.64	0.82 ± 1.64	-0.79 ± 1.66	-1.28 ± 1.73	-0.02 ± 0.05	0.62 ± 0.53
$r_{x,y}$	0.87	0.87	0.88	0.88	0.87	> 0.99	> 0.99	> 0.99
LoA	−2.41 to 4.07	−2.45 to 4.0	−3.98 to 2.46	−2.39 to 4.03	−4.05 to 2.46	−4.68 to 2.11	−0.12 to 0.09	−0.41 to 1.66
SDC	6.48	6.46	6.44	6.42	6.50	6.79	0.21	2.08

LR, loading response; SLS, single limb support; PS, pre-swing.
MAD, mean absolute difference between IMU-based and Zebris values.
$\mu \pm \sigma$, mean and standard deviation of difference between IMU-based and Zebris values.
$r_{x,y}$: Pearson correlation coefficient ($p < 0.01$ for all values).
LoA, limits of agreement, $\mu - 1.96\sigma$ to $\mu + 1.96\sigma$.
SDC, smallest detectable change, range between both LoA.

and specifically the neurologically induced gait abnormalities are challenging for IMU-based gait analysis (29).

A fundamental challenge of IMU-based gait event detection is that IMUs do not directly measure the gait parameters of interest. For toe-off detection, the time instant of load relief cannot directly be measured, and instead, the inversion of the direction of rotation is used. Similarly, initial contact is not detected based on the onset of load but based on the change of acceleration. It is therefore important to properly validate the IMU-based methods by comparing the estimated gait parameters to a reliable ground truth.

As reference system, treadmills instrumented with Zebris pressure measurement platforms were used, which are frequently employed for gait analysis in clinical practice as well as scientific data collection (12). This system shows good reliability (67), but no studies could be found in which the validity of the gait parameters was investigated. It should be noted that due to the length of the pressure sensors (FDM-THM-M-3i: 0.85 cm; FDM-THM-M-2i: 1.27 cm) there may be inaccuracies in the recording of spatial parameters, which may have an effect on the results of the comparative measurements. Moreover, calibration and proper thresholding pose challenges in gait event detection based on pressure measurements (12).

For the neurological patients, the reported duration of stance is, on average, 2 % longer than the reference duration. While this is still a small deviation, it is worth noting because this bias suggests a pattern that is common to this subject group. One likely explanation is that toe-off is being detected later than with the Zebris system. This might be due to a comparatively long phase of load relief that causes the pressure to fall below the threshold too early. Furthermore, the reversal of rotation direction might happen later than for healthy subjects or orthopedic patients. Still, even though both systems measure inherently different phenomena, the observation deviation is only 2 %.

As a replacement for traditional stationary gait analysis systems, which are commonly used in clinical practice, IMU-based gait analysis offers several advantages. Measurement is possible both on treadmills and overground and not restricted to a dedicated laboratory. The small and lightweight IMUs do not restrict the movement of the subject and can be used in conjunction with walking aids such as wheeled walkers.

Furthermore, only a very short setup time is required before starting the actual measurement.

Unlike most existing methods (cf. section 2), the proposed method makes gait analysis easier and faster by not requiring any specific sensor attachment, which we demonstrated by simulating a different random sensor-to-foot orientation in each trial. It does not make use of magnetometers and can therefore be used in both indoor and outdoor environments.

While evaluation was limited to the gait phases reported by the reference system, our proposed set of methods further allows for the calculation of many gait phases (**Figure 4**), i.e., swing and stance for each foot, four unilateral gait phases for each foot, five bilateral gait phases following standard literature (44) for each foot, and finally the distinction between double and single support. To the best of our knowledge, no existing work on IMU-based gait analysis describes the calculation of this set of gait phases.

Besides the more fine-grained gait phases, there are many more parameters that can be extracted, e.g., from the velocity and position trajectories, such as the maximum velocity during swing, foot clearance, and symmetry parameters. While it is not surprising that the prevalence of pressure-based systems has led researchers to focus on features based on ground contact, it is to be expected that the focus of clinical gait analysis will be directed toward other parameters as IMU-based systems become more popular.

Furthermore, miniaturized lightweight sensors with a long battery life open up possibilities for objective gait analysis outside of clinical laboratories. Daily-life gait assessment over the course of multiple days can bring insights that are not possible with short sessions in a laboratory. If patients place the sensors on or in the shoes themselves in an unsupervised telemedicine setting, not requiring the sensor to be oriented in a special way becomes even more important.

Technological advancement also facilitates real-time biofeedback applications. While there are methods for real-time applications that require event detection during a step (41), e.g., to trigger FES, the proposed set of methods is real-time capable in the sense that during walking, sections of data containing a small number of strides can be processed and used to provide feedback to the subject.

The presented work exhibits a few remaining limitations. In the statistical analysis, the gait parameters were averaged over the duration of the trial before comparison with the reference. While it allows for single-stride errors to cancel out, this methodology corresponds well with the use case of clinical gait analysis, in which a subject is asked to walk for several steps, and averaged parameters are then used to assess the gait. An additional stride-by-stride comparison was not performed because the employed reference system can only export averaged gait parameters. In addition, it should be noted that all recordings were made on treadmills and not while walking overground, which has an influence on the movement pattern of gait (68, 69). Despite the known differences between treadmill walking and overground walking, treadmill gait analysis is considered a standard method in clinical practice (70, 71), especially when weight support and handrails are required for safety reasons.

6. CONCLUSION

In the present contribution, we have proposed a set of methods for IMU-based gait analysis. Based on gyroscope and accelerometer measurements from two inertial sensors on the feet, we estimate durations of five gait phases, stride length, walking speed, and cadence. Using a Zebris Rehawalk instrumented treadmill as reference, we validated the proposed methods based on a large data set consisting of healthy subjects ($n = 39$) walking at three different speeds, subjects with orthopedic diseases ($n = 62$), and subjects with neurological diseases ($n = 36$). Averaged over all trials, the MAD with respect to the reference system are 1.4 % for the gait phase durations, 1.7 cm for the stride length, 0.04 km/h for the walking speed, and 0.7 steps/min for the cadence. We also demonstrated that the proposed methods work reliably not only in healthy subjects but also in patients and still provide accurate results under different pathological gait patterns.

This shows that the proposed setup in combination with the proposed methods can accurately calculate relevant gait parameters from the inertial sensor data and thus has the potential to replace traditional stationary gait analysis systems.

Furthermore, we validated that the proposed methods work well regardless of the orientation in which the sensor is attached to the foot, and dedicated calibration movements and magnetometer measurements are completely avoided. The combination of these advantages facilitates long-term ambulatory gait analysis in day-to-day situations without the need for supervision by health professionals.

Future research will focus on the estimation of additional gait parameters, on the validation on stairs and slopes, and the validation against marker-based optical motion capture systems.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because sharing of the data is not covered by the ethical approval. Requests to access the datasets should be directed to Andreas J. Jocham, andreas.jocham@fh-joanneum.at.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Graz (GZ. 39/55/63 ex 2017/18, 28 May 2018). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DL and TS devised and developed the mathematical method. DL implemented the method, performed the data analysis, and drafted the manuscript. AJ, BG, KA, and MF planned and organized the data collection. AJ, BG, and KA conducted the data collection and post-processed the data. DL, AJ, BG, KA,

MF, and TS revised the manuscript. All authors approved the submitted version.

FUNDING

We acknowledge support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

ACKNOWLEDGMENTS

We thank all subjects for their participation, Dr. Matthias König and Andrea Eschbach as well as the whole team of the Neurological Therapy Center Kapfenberg for the support during data collection, and Andrew Cote and Eva Kastenbauer for the skillful support in algorithm development and evaluation.

REFERENCES

1. Hass CJ, Malczak P, Nocera J, Stegemöller EL, Shukala A, Malaty I, et al. Quantitative normative gait data in a large cohort of ambulatory persons with Parkinson's disease. *PLoS ONE*. (2012) 7:e42337. doi: 10.1371/journal.pone.0042337
2. Park J, Kim TH. The effects of balance and gait function on quality of life of stroke patients. *NeuroRehabilitation*. (2019) 44:37–41. doi: 10.3233/NRE-182467
3. Abu-Faraj ZO, Harris GF, Smith PA, Hassani S. Human gait and clinical movement analysis. In: Webster JG, editor. *Wiley Encyclopedia of Electrical and Electronics Engineering*. New York, NY: John Wiley & Sons, Inc. (2015). p. 1–34. doi: 10.1002/047134608X.W6606.pub2
4. Baker R. Gait analysis methods in rehabilitation. *J NeuroEng Rehabil*. (2006) 3:4. doi: 10.1186/1743-0003-3-4
5. Baker RW. *Measuring Walking: A Handbook of Clinical Gait Analysis*. 1st Edn. London: Mac Keith Press (2013).
6. Chen S, Lach J, Lo B, Yang GZ. Toward pervasive gait analysis with wearable sensors: a systematic review. *IEEE J Biomed Health Informatics*. (2016) 20:1521–37. doi: 10.1109/JBHI.2016.2608720
7. Bridenbaugh SA, Kressig RW. Laboratory review: the role of gait analysis in seniors' mobility and fall prevention. *Gerontology*. (2011) 57:256–64. doi: 10.1159/000322194
8. Givon U, Zeilig G, Achiron A. Gait analysis in multiple sclerosis: characterization of temporal–spatial parameters using GAITRite functional ambulation system. *Gait Posture*. (2009) 29:138–42. doi: 10.1016/j.gaitpost.2008.07.011
9. Wren TAL, Tucker CA, Rethlefsen SA, Gorton GE, Ounpuu S. clinical efficacy of instrumented gait analysis: systematic review 2020 update. *Gait Posture*. (2020) 80:274–9. doi: 10.1016/j.gaitpost.2020.05.031
10. Riis J, Byrgesen SM, Kragholm KH, Mørch MM, Melgaard D. Validity of the GAITRite walkway compared to functional balance tests for fall risk assessment in geriatric outpatients. *Geriatrics*. (2020) 5:77. doi: 10.3390/geriatrics5040077
11. Schmitz-Hübsch T, Brandt AU, Pfueller C, Zange L, Seidel A, Kühn AA, et al. Accuracy and repeatability of two methods of gait analysis–GaitRiteTM und mobility labTM–in subjects with cerebellar ataxia. *Gait Posture*. (2016) 48:194–201. doi: 10.1016/j.gaitpost.2016.05.014
12. Wearing SC, Reed LF, Urry SR. Agreement between temporal and spatial gait parameters from an instrumented walkway and treadmill system at matched walking speed. *Gait Posture*. (2013) 38:380–4. doi: 10.1016/j.gaitpost.2012.12.017
13. Kotiadis D, Hermens HJ, Veltink PH. Inertial gait phase detection for control of a drop foot stimulator: inertial sensing for gait phase detection. *Med Eng Phys*. (2010) 32:287–97. doi: 10.1016/j.medengphy.2009.10.014
14. Dvorani A, Wiesener C, Valtin M, Voigt H, Kühn A, Wenger N, et al. Mobil4Park: development of a sensor-stimulator network for the therapy of freezing of gait in Parkinson patients. *Curr Direct Biomed Eng*. (2020) 6:1–4. doi: 10.1515/cdbme-2020-2013
15. Mariani B, Rouhani H, Crevoisier X, Aminian K. Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors. *Gait Posture*. (2013) 37:229–34. doi: 10.1016/j.gaitpost.2012.07.012
16. Trojaniello D, Cereatti A, Pelosin E, Avanzino L, Mirelman A, Hausdorff JM, et al. Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait. *J NeuroEng Rehabil*. (2014) 11:152. doi: 10.1186/1743-0003-11-152
17. Rampp A, Barth J, Schülein S, Gaßmann KG, Klucken J, Eskofier BM. Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. *IEEE Trans Biomed Eng*. (2015) 62:1089–97. doi: 10.1109/TBME.2014.2368211
18. de Vries WHK, Veeger HEJ, Baten CTM, van der Helm FCT. Magnetic distortion in motion labs, implications for validating inertial magnetic sensors. *Gait Posture*. (2009) 29:535–41. doi: 10.1016/j.gaitpost.2008.12.004
19. Seel T, Landgraf L, Cermeño Escobar V, Raisch J, Schauer T. Online gait phase detection with automatic adaption to gait velocity changes using accelerometers and gyroscopes. *Biomed Eng Biomed Techik*. (2014) 59:795–98. doi: 10.1515/bmt-2014-5011
20. Müller P, Seel T, Schauer T. Experimental evaluation of a novel inertial sensor based realtime gait phase detection algorithm. In: *Proceedings of the Technically Assisted Rehabilitation Conference*. (2015).
21. Schicketmueller A, Rose G, Hofmann M. Feasibility of a sensor-based gait event detection algorithm for triggering functional electrical stimulation during robot-assisted gait training. *Sensors*. (2019) 19:4804. doi: 10.3390/s19214804
22. Sabatini AM, Martelloni C, Scapellato S, Cavallo F. Assessment of walking features from foot inertial sensing. *IEEE Trans Biomed Eng*. (2005) 52:486–94. doi: 10.1109/TBME.2004.840727
23. Hannink J, Kautz T, Pasluosta CF, Gaßmann KG, Klucken J, Eskofier BM. Sensor-based gait parameter extraction with deep convolutional neural networks. *IEEE J Biomed Health Informatics*. (2017) 21:85–93. doi: 10.1109/JBHI.2016.2636456
24. Donath L, Faude O, Lichtenstein E, Nüesch C, Mündermann A. Validity and reliability of a portable gait analysis system for measuring spatiotemporal gait characteristics: comparison to an instrumented treadmill. *J NeuroEng Rehabil*. (2016) 13:6. doi: 10.1186/s12984-016-0115-z
25. Mannini A, Genovese V, Maria Sabatini A. Online decoding of hidden Markov models for gait event detection using foot-mounted gyroscopes. *IEEE J Biomed Health Informatics*. (2014) 18:1122–30. doi: 10.1109/JBHI.2013.2293887
26. Mannini A, Sabatini AM. Gait phase detection and discrimination between walking–jogging activities using hidden Markov models applied to foot motion data from a gyroscope. *Gait Posture*. (2012) 36:657–61. doi: 10.1016/j.gaitpost.2012.06.017
27. Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C. Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait Posture*. (2017) 55:87–93. doi: 10.1016/j.gaitpost.2017.04.013
28. Mariani B, Hoskovec C, Rochat S, Büla C, Penders J, Aminian K. 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. *J Biomech*. (2010) 43:2999–3006. doi: 10.1016/j.jbiomech.2010.07.003
29. Lefeber N, Degelaen M, Truyers C, Safin I, Beckwée D. Validity and reproducibility of inertial physilog sensors for spatiotemporal gait analysis in patients with stroke. *IEEE Trans Biomed Eng*. (2019) 27:1865–74. doi: 10.1109/TNSRE.2019.2930751
30. Jasiewicz JM, Allum JHJ, Middleton JW, Barriskill A, Condie P, Purcell B, et al. Gait event detection using linear accelerometers or angular velocity transducers in able-bodied and spinal-cord injured individuals. *Gait Posture*. (2006) 24:502–9. doi: 10.1016/j.gaitpost.2005.12.017

31. Mariani B, Jiménez MC, Vingerhoets FJG, Aminian K. On-shoe wearable sensors for gait and turning assessment of patients with Parkinson's disease. *IEEE Trans Biomed Eng*. (2013) 60:155–8. doi: 10.1109/TBME.2012.2227317
32. Bertoli M, Cereatti A, Trojaniello D, Avanzino L, Pelosin E, Del Din S, et al. Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: multicenter validation among Parkinson, mildly cognitively impaired and healthy older adults. *BioMedical Eng OnLine*. (2018) 17:1–14. doi: 10.1186/s12938-018-0488-2
33. Chia Bejarano N, Ambrosini E, Pedrocchi A, Ferrigno G, Monticone M, Ferrante S. A novel adaptive, real-time algorithm to detect gait events from wearable sensors. *IEEE Trans Neural Syst Rehabil Eng*. (2015) 23:413–22. doi: 10.1109/TNSRE.2014.2337914
34. Sabatini AM, Ligorio G, Mannini A. Fourier-based integration of quasi-periodic gait accelerations for drift-free displacement estimation using inertial sensors. *BioMed Eng OnLine*. (2015) 14:106. doi: 10.1186/s12938-015-0103-8
35. Mannini A, Trojaniello D, Della Croce U, Sabatini AM. Hidden Markov model-based strategy for gait segmentation using inertial sensors: application to elderly, hemiparetic patients and Huntington's disease patients. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan (2015). p. 5179–82. doi: 10.1109/EMBC.2015.7319558
36. Marín J, Blanco T, de la Torre J, Marín JJ. Gait analysis in a box: a system based on magnetometer-free IMUs or clusters of optical markers with automatic event detection. *Sensors*. (2020) 20:3338. doi: 10.3390/s20123338
37. Teufl W, Lorenz M, Miezal M, Taetz B, Fröhlich M, Bleser G. Towards inertial sensor based mobile gait analysis: event-detection and spatio-temporal parameters. *Sensors*. (2019) 19:38. doi: 10.3390/s19010038
38. Sejdić E, Lowry KA, Bellanca J, Perera S, Redfern MS, Brach JS. Extraction of stride events from gait accelerometry during treadmill walking. *IEEE J Transl Eng Health Med*. (2016) 4:1–11. doi: 10.1109/JTEHM.2015.2504961
39. Hung TN, Suh YS. Inertial sensor-based two feet motion tracking for gait analysis. *Sensors*. (2013) 13:5614–29. doi: 10.3390/s130505614
40. Picerno P, Cereatti A, Cappozzo A. Joint kinematics estimate using wearable inertial and magnetic sensing modules. *Gait Posture*. (2008) 28:588–95. doi: 10.1016/j.gaitpost.2008.04.003
41. Seel T, Raisch J, Schauer T. IMU-based joint angle measurement for gait analysis. *Sensors*. (2014) 14:6891–909. doi: 10.3390/s140406891
42. Laidig D, Müller P, Seel T. Automatic anatomical calibration for IMU-based elbow angle measurement in disturbed magnetic fields. *Curr Direct Biomed Eng*. (2017) 3:167–70. doi: 10.1515/cdbme-2017-0035
43. Graurock D, Schauer T, Seel T. Automatic pairing of inertial sensors to lower limb segments–a plug-and-play approach. *Curr Direct Biomed Eng*. (2016) 2:715–8. doi: 10.1515/cdbme-2016-0155
44. Perry J, Burnfield JM, editors. *Gait Analysis: Normal and Pathological Function*. 2nd Edn. Thorofare, NJ: SLACK (2010).
45. Caldas R, Mundt M, Potthast W, Buarque de Lima Neto F, Markert B. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. *Gait Posture*. (2017) 57:204–10. doi: 10.1016/j.gaitpost.2017.06.019
46. Yang S, Li Q. Inertial sensor-based methods in walking speed estimation: a systematic review. *Sensors*. (2012) 12:6102–16. doi: 10.3390/s120506102
47. Mariani B, Rochat S, Büla CJ, Aminian K. Heel and toe clearance estimation for gait analysis using wireless inertial sensors. *IEEE Trans Biomed Eng*. (2012) 59:3162–8. doi: 10.1109/TBME.2012.2216263
48. Kuipers JB. Quaternions and rotation sequences. In: *Proceedings of the International Conference on Geometry, Integrability and Quantization*. Varna: Coral Press Scientific Publishing (2000). p. 127–43.
49. Ridler TW, Calvard S. Picture thresholding using an iterative selection method. *IEEE Trans Syst Man Cybern*. (1978) 8:630–2. doi: 10.1109/TSMC.1978.4310039
50. Giavarina D. Understanding bland altman analysis. *Biochem Med*. (2015) 25:141–51. doi: 10.11613/BM.2015.015
51. Chan YH. Biostatistics 104: correlational analysis. *Singapore Med J*. (2003) 44:614–9.
52. Busch TdA, Duarte YA, Pires Nunes D, Lebr ao ML, Satya Naslavsky M, dos Santos Rodrigues A, et al. Factors associated with lower gait speed among the elderly living in a developing country: a cross-sectional population-based study. *BMC Geriatr*. (2015) 15:35. doi: 10.1186/s12877-015-0031-2
53. Dommershuijsen LJ, Isik BM, Darweesh SKL, van der Geest JN, Ikram MK, Ikram MA. Unraveling the association between gait and mortality–one step at a time. *J Gerontol*. (2020) 75:1184–90. doi: 10.1093/gerona/glz282
54. White DK, Neogi T, Nevitt MC, Peloquin CE, Zhu Y, Boudreau RM, et al. Trajectories of gait speed predict mortality in well-functioning older adults: the health, aging and body composition study. *J Gerontol Ser A*. (2013) 68:456–64. doi: 10.1093/gerona/gls197
55. Jerome GJ, Ko Su, Kauffman D, Studenski SA, Ferrucci L, Simonsick EM. Gait characteristics associated with walking speed decline in older adults: results from the Baltimore longitudinal study of aging. *Arch Gerontol Geriatr*. (2015) 60:239–43. doi: 10.1016/j.archger.2015.01.007
56. Abellan Van Kan G, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) task force. *J Nutr Health Aging*. (2009) 13:881–9. doi: 10.1007/s12603-009-0246-z
57. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. *JAMA*. (2011) 305:50–8. doi: 10.1001/jama.2010.1923
58. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc*. (2003) 51:314–22. doi: 10.1046/j.1532-5415.2003.51104.x
59. Smith AJJ, Lemaire ED. Temporal-spatial gait parameter models of very slow walking. *Gait Posture*. (2018) 61:125–9. doi: 10.1016/j.gaitpost.2018.01.003
60. Verghese J, Holtzer R, Lipton RB, Wang C. Quantitative gait markers and incident fall risk in older adults. *J Gerontol Ser A*. (2009) 64A:896–901. doi: 10.1093/gerona/glp033
61. Maki BE. Gait changes in older adults: predictors of falls or indicators of fear? *J Am Geriatr Soc*. (1997) 45:313–20. doi: 10.1111/j.1532-5415.1997.tb00946.x
62. Bohannon RW, Glenney SS. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review. *J Eval Clin Pract*. (2014) 20:295–300. doi: 10.1111/jep.12158
63. Tilson JK, Sullivan KJ, Cen SY, Rose DK, Koradia CH, Azen SP, et al. Meaningful gait speed improvement during the first 60 days poststroke: minimal clinically important difference. *Phys Ther*. (2010) 90:196–208. doi: 10.2522/ptj.20090079
64. Palombaro KM, Craik RL, Mangione KK, Tomlinson JD. Determining meaningful changes in gait speed after hip fracture. *Phys Ther*. (2006) 86:809–16. doi: 10.1093/ptj/86.6.809
65. Brown JC, Harhay MO, Harhay MN. Walking cadence and mortality among community-dwelling older adults. *J Gen Intern Med*. (2014) 29:1263–9. doi: 10.1007/s11606-014-2926-6
66. Hannink J, Kautz T, Pasluosta CF, Barth J, Schülein S, Gaßmann KG, et al. Mobile stride length estimation with deep convolutional neural networks. *IEEE J Biomed Health Informatics*. (2018) 22:354–62. doi: 10.1109/JBHI.2017.2679486
67. Nüesch C, Overberg JA, Schwameder H, Pagenstert G, Mündermann A. Repeatability of spatiotemporal, plantar pressure and force parameters during treadmill walking and running. *Gait Posture*. (2018) 62:117–23. doi: 10.1016/j.gaitpost.2018.03.017
68. Lee SJ, Hidler J. Biomechanics of overground vs. treadmill walking in healthy individuals. *J Appl Physiol*. (2008) 104:747–55. doi: 10.1152/japplphysiol.01380.2006
69. Watt JR, Franz JR, Jackson K, Dicharry J, Riley PO, Kerrigan DC. A Three-dimensional kinematic and kinetic comparison of overground and treadmill walking in healthy elderly subjects. *Clin Biomech*. (2010) 25:444–9. doi: 10.1016/j.clinbiomech.2009.09.002
70. Meyer C, Killeen T, Easthope CS, Curt A, Bolliger M, Linnebank M, et al. Familiarization with treadmill walking: how much is enough? *Sci Rep*. (2019) 9:5232. doi: 10.1038/s41598-019-41721-0

71. Kalron A, Dvir Z, Frid L, Achiron A. Quantifying gait impairment using an instrumented treadmill in people with multiple sclerosis. *ISRN Neurol*. (2013) 2013:e867575. doi: 10.1155/2013/867575

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

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

frontiers in Digital Health

REVIEW
published: 12 November 2021
doi: 10.3389/fdgth.2021.770096

Smartphones and Video Cameras: Future Methods for Blood Pressure Measurement

Joe Steinman[1], *Andrew Barszczyk*[1,2], *Hong-Shuo Sun*[1,3], *Kang Lee*[2]* *and Zhong-Ping Feng*[1]*

[1] Department of Physiology, University of Toronto, Toronto, ON, Canada, [2] Dr. Eric Jackman Institute of Child Study, University of Toronto, Toronto, ON, Canada, [3] Department of Surgery, University of Toronto, Toronto, ON, Canada

Regular blood pressure (BP) monitoring enables earlier detection of hypertension and reduces cardiovascular disease. Cuff-based BP measurements require equipment that is inconvenient for some individuals and deters regular home-based monitoring. Since smartphones contain sensors such as video cameras that detect arterial pulsations, they could also be used to assess cardiovascular health. Researchers have developed a variety of image processing and machine learning techniques for predicting BP *via* smartphone or video camera. This review highlights research behind smartphone and video camera methods for measuring BP. These methods may in future be used at home or in clinics, but must be tested over a larger range of BP and lighting conditions. The review concludes with a discussion of the advantages of the various techniques, their potential clinical applications, and future directions and challenges. Video cameras may potentially measure multiple cardiovascular metrics including and beyond BP, reducing the risk of cardiovascular disease.

Keywords: blood pressure, imaging, physiology, hemodynamics, cardiovascular digital health

OPEN ACCESS

Edited by:
Mohamed Elgendi,
University of British Columbia, Canada

Reviewed by:
Yongbo Liang,
Guilin University of Electronic Technology, China
Toshiyo Tamura,
Waseda University, Japan

***Correspondence:**
Kang Lee
kang.lee@utoronto.ca
Zhong-Ping Feng
zp.feng@utoronto.ca

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

Received: *03 September 2021*
Accepted: *15 October 2021*
Published: *12 November 2021*

Citation:
Steinman J, Barszczyk A, Sun H-S, Lee K and Feng Z-P (2021) Smartphones and Video Cameras: Future Methods for Blood Pressure Measurement. Front. Digit. Health 3:770096. doi: 10.3389/fdgth.2021.770096

INTRODUCTION

Blood pressure (BP) measurement is necessary in determining an individual's risk for cardiovascular disease and the need for early treatment. Early detection and treatment of BP may delay or prevent conditions related to high BP, such as stroke. This is particularly important in the Covid era, where there has been an increase in the number of virtual consultations with patients (1). Digital or at-home methods where individuals accurately and easily determine BP may improve population health, while minimizing hospital visits.

Methods for measuring BP at home or in the clinic are commonly cuff-based. Cuff-based systems are automated; however, they present difficulty in portability outside the home. Many individuals find application of the cuff awkward, inconvenient, and uncomfortable. This limits the number of daily BP measurements. Since BP varies according to time, season, amount of sleep, and activity, a single measurement over the course of a day, or every few days, does not provide an accurate assessment of cardiovascular changes and BP variation in an individual (2).

Smartphones could serve as alternatives to the cuff. Many individuals possess smartphones and operate their features with ease. Phones are embedded with cameras, microphones, light emitters, and force sensors that can be used to obtain a cardiovascular pulse signal, and ultimately predict BP. Due to their size, they overcome issues of portability, discomfort, or inconvenience.

Methods that utilize video cameras to predict BP are continually undergoing research and development for improved accuracy. Most smartphone techniques utilize the video camera to extract the photoplethysmography (PPG) signal from light reflected from the skin. Due to complexity in relating the PPG signal to BP, methodologies have been developed to determine BP from video PPG. This includes image processing to extract blood flow, machine learning algorithms for calculating BP, and incorporation of smartphone features such as the microphone or force sensors. Mathematical models may be applied to separate hemoglobin signals from melanin and light, and the shape of the pulse or different arrival times of the pulse used to predict BP (3).

This paper focuses on currently published video camera and smartphone methods for BP measurement, highlighting research efforts and experiments from a variety of groups. The two categories of smartphone/video-camera BP measurements, contact and non-contact, are covered. Present use of these techniques is discussed, together with their clinical applicability. We conclude by discussing the role of video cameras in health and BP monitoring. This paper provides an in-depth review of BP-video camera measurement technologies, including their accuracy, image processing methodologies, and machine learning algorithms used for predicting BP. This will provide a resource for researchers in this field to compare the advantages/disadvantages and technical details of various published methods.

Video cameras and smartphones could measure BP non-invasively. Through simultaneously measuring additional cardiovascular properties such as heart rate or blood oxygenation, video camera technologies may provide continuous monitoring of multiple cardiovascular properties from an individual's home.

NON-INVASIVE SMARTPHONE CONTACT MEASUREMENTS OF BLOOD PRESSURE

Contact methods typically involve pressing the fingertip against the rear camera to acquire a PPG signal. The theory of reflection mode PPG in video cameras is initially discussed. This is followed by a description of the three categories of contact-based BP measurements (see **Table 1** and **Figure 1**): (1) Oscillometry; (2) Analysis of pulse waveform features; (3) and Pulse transit time (PTT), calculated as the time delay between two PPG waveforms at different arterial sites.

Reflection Photoplethysmography and Smartphones: An Introduction

The heart generates pulsatile flow, causing blood vessels in the skin to expand and contract. Light absorption by hemoglobin in the blood is maximized when the vessel is fully expanded during systole and minimized during diastole. In reflection mode PPG, light reflected from the skin is detected by a sensor or camera. The PPG waveform has an oscillating "AC" component largely due to arterial pulsation, which is superimposed on a DC component corresponding to fat and blood volume.

In 2010, the smartphone was used to obtain a PPG signal for heart rate assessment by pressing the finger against the rear camera (27). Although data from red, green, and blue color channels are obtained, the green channel is typically used for calculating physiological parameters *via* video camera methods. This is likely due to green light possessing higher absorption by hemoglobin than red, while penetrating deeper into tissue than blue (28). These techniques have been extended to estimate BP (systolic, diastolic, mean) *via* a logarithmic equation relating pressure to heart rate and pulse volume (19).

Application of Oscillometry to Smartphones

Oscillometry techniques produce an automated digital pressure output and determine BP with limited user input. Vibrations produced through opening of the arterial wall travel through air inside the cuff, and into a transducer producing an electrical signal (29). The upper and lower envelopes of the oscillation are traced as cuff pressure varies from above systolic pressure (SP) to below diastolic pressure (DP). Algorithms then estimate the mean, systolic, and diastolic pressures from the oscillogram (plot of the oscillation amplitude vs. cuff pressure) (30, 31).

Oscillometry has been extended to smartphones. Chandrasekhar et al. (16) developed a smartphone-based device to detect varying pressure in a finger artery, similar to that of a changing cuff pressure as with cuff-based oscillometry. A case was attached to the smartphone that contained a PPG sensor overlaying a force transducer, since a sensor capable of detecting force applied by the finger was not present within the phone itself. An infrared LED illuminated the finger pressed against the PPG sensor, and the force sensor detected the varying force applied by the user. An oscillogram was generated, and finger BP related to brachial BP *via* fitting a parametric model to the oscillogram.

Most users found the technique user-friendly and learned it after 1–2 trials. However, whereas the accuracy of this technique was comparable to a finger-cuff, only approximately 60 % of BP measurements were successful with the device. More than half the failures to output a BP value were attributed to a "computation failure." A special case was also required for this method that incorporated sensors for production of a finger oscillogram. An additional study by the same group (17) incorporated the iPhone X's built-in 3D touch force sensor and camera for PPG detection, eliminating the need for a special case.

PPG Waveform Analysis to Predict Blood Pressure

Calculating Blood Pressure From Waveform Analysis

A common approach to calculating BP is to extract features from the pulse waveform related to the shape of the pulse. The user turns on the LED flash, presses their finger against the phone camera, and records a video. The resulting video can be analyzed to produce a pulsatile waveform. Features are then extracted characterizing the waveform such as pulse width, slope of initial upstroke, height, time between pulses, etc. These features are input into machine learning models, such as

TABLE 1 | Contact methods for smartphone blood pressure (BP) measurement.

Publications	Number of subjects in study and additional experimental details	Accuracy of method	Analysis/processing method
Chandrasekaran et al. (4)	5 subjects - 2 phones (1 camera, 1 microphone) OR single phone camera with external microphone	*Still with single mobile:* SP: 98.59% accuracy DP: 97.96% accuracy *Movement with single mobile:* SP: 97.45% accuracy DP: 97.63% accuracy - mean accuracy values calculated by Steinman et al.: original data in paper provided individual but overall mean values	Pulse Transit Time
Lamonaca et al. (5)	5 experiments	Max error in SP: 11 mHg Max error in DP: 12 mmHg	Waveform Analysis
Visvanathan et al. (6)	17 subjects	*Linear regression:* SP: 98.7% detection accuracy DP: 99.7% detection accuracy *SVM:* SP: 100% detection accuracy DP: 99.29% detection accuracy * BP values divided into bins	Waveform Analysis
Visvanathan et al. (7)	156 subjects	SP: 98.81% detection accuracy (cross validation) DP: 98.21% detection accuracy (cross validation) * BP values divided into bins	Waveform Analysis
Banerjee et al. (8)	23 subjects (15 training, 8 testing)	SP: 4 ± 2 mmHg (MAE ± std) DP: 4 ± 2 mmHg (MAE ± std) - values calculated by Steinman et al.; original data in paper only provided for individuals but not averaged	Waveform Analysis
Liu et al. (9)	12 subjects - 2 cameras (1 for fingertip, other for forehead temple)	- correlation 0.86 ± 0.06 between established PTT and OFP * OFP is the time interval between minimum PPG signal from temple and maximum of PPG signal from fingertip	Pulse Transit Time
Peng et al. (10)	32 subjects stethoscope attached to phone	SP: 4.339 ± 6.121 (MAE ± std) DP: 3.171 ± 4.471 (MAE ± std) MP: 3.480 ± 4.961 (MAE ± std)	Heart Sounds Only
Junior et al. (11) Junior et al. (12)	3 subjects - heart sounds and camera	Mean percent error in automated vs. manual calculation of PTT: 2.53% (maximum 3.00%)	Pulse Transit Time
Gao et al. (13)	65 subjects	SP: 5.1 ± 4.3 mmHg (ME ± std) DP: 4.6 ± 4.3 mmHg (ME ± std)	Waveform Analysis
Plante et al. (14)	85 subjects - heart sounds and camera	SP: 12.4 ± 10.5 mmHg (MAE ± std) DP: 10.1 ± 8.1 mmHg (MAE ± std)	Pulse Transit Time
Datta et al. (15)	118 subjects (68 training from oximeter PPG; 50 smartphone PPG for testing)	SP: Mean absolute percentage difference 7.4% Correlation 0.57 with ground truth SBP DP: Mean absolute percentage difference 9.1 % Correlation 0.40 with ground truth DBP	Waveform Analysis
Chandrasekhar et al. (16)	32 subjects (35 subjects originally) - special case used to measure PPG and applied force	SP: 3.3 ± 8.8 mmHg (ME ± std) DP: −5.6 ± 77 mmHg (ME ± std)	Oscillometry
Chandrasekhar et al. (17)	18 subjects (20 subjects originally) - phone camera, iPhoneX 3D touch feature to measure applied force	SP: −4.0 ± 11.4 mmHg (ME ± std) DP: −9.4 ± 9.7 mmHg (ME ± std)	Oscillometry
Dey et al. (18)	205 subjects (160 training, 45 testing)	SP: 6.9 ± 9.0 mmHg (MAE ± std) DP: 5.0 ± 6.1 mmHg (MAE ± std)	Waveform Analysis
Matsumara et al. (19)	13 subjects	SP: 0.67 ± 12.7 mmHg (ME ± std) DP: 0.45 ± 8.6 mmHg (ME ± std) MP: 0.49 ± 9.6 mmHg (ME ± std)	Waveform Analysis
Wang et al. (20)	7 subjects (nine subjects originally) - phone accelerometer and camera	DP: 5.2 ± 2.0 (RMSE ± std)	Pulse Transit Time
Baek et al. (21)	26 subjects - convolutional neural network without feature extraction	SP: 5.28 ± 1.80 (MAE ± std) DP: 4.92 ± 2.42 (MAE ± std)	Waveform Analysis

(Continued)

TABLE 1 | Continued

Publications	Number of subjects in study and additional experimental details	Accuracy of method	Analysis/processing method
OptiBP, Schoettker et al. (22)	50 subjects for training (51 originally), 40 validation (50 originally)	SP: −0.7 ± 7.7 mmHg (ME ± std) DP: −0.4 ± 4.5 mmHg (ME ± std) MP: −0.6 ± 5.2 mmHg (ME ± std)	Waveform Analysis
Nemcova et al. (23)	22 subjects - heart sounds and camera	SP: −0.2 ± 6.7 mmHg (ME ± std) DP: −0.07 ± 8.8 mmHg (ME ± std)	Pulse Transit Time
Tabei et al. (24)	6 subjects - cameras from 2 smartphones	SP: 2.07 ± 2.06 mmHg (MAE ± std) DP: 2.12 ± 1.85 mmHg (MAE ± std)	Pulse Transit Time
Preventicus (app)	Raichle et al. (25): 32 pregnant women Dörr et al. (26): 965 subjects (1,036 subjects originally)	Raichle 2018: SP: 5.0 ± 14.5 mmHg (ME ± std) Dörr 2021: SP: −0.41– ± 16.52 mmHg (ME ± std)	Waveform Analysis

ME, mean error; MAE, mean absolute error; SP, systolic pressure; DP, diastolic pressure; MP, mean pressure; PTT, pulse transit time; PPG, photoplethysmography.

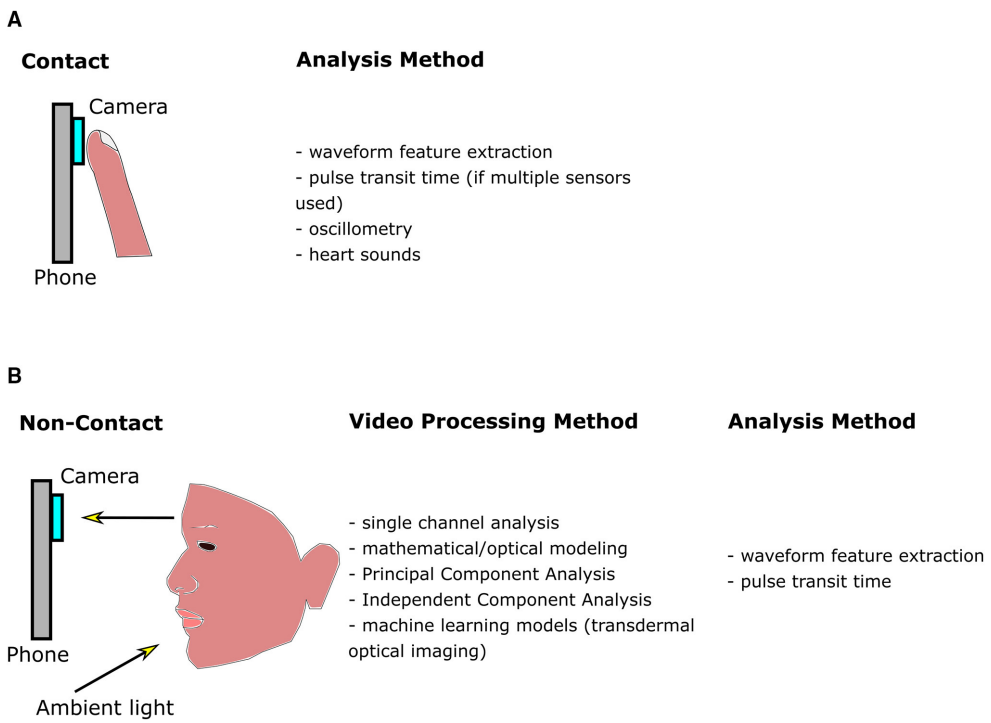


FIGURE 1 | Diagram outlining methods for smartphone blood pressure (BP) estimation. **(A)** Contact methods, which require pressing the finger against the phone camera to obtain a finger blood volume pulse. Pixels within the video are averaged in each frame. The signal may be further processed and filtered, producing a waveform as a function of time. Features are extracted from the waveform and input into machine learning algorithms to calculate BP. Pulse transit time (PTT) may also be correlated with BP, however multiple sensors are required. **(B)** Non-contact methods utilize ambient light reflected from the face. The resulting video is processed to enhance the signal-to-noise ratio of the hemoglobin signal, whose features are be input into a machine learning algorithm to calculate BP (similar to contact methods). Since multiple facial regions or body parts may be imaged simultaneously, PTT may be estimated using only a single camera by estimating the difference between pulse arrival at different body regions.

neural networks or regression models, thereby calculating blood pressure. Multiple features render algorithms less susceptible to data variability, and increase pressure calculation accuracy. Unlike methods such as PTT (see below), only one sensor is needed, and it is less sensitive to motion artifacts because the finger is pressed against the camera. This also produces a stronger signal than with non-contact video PPG methods (section Non-contact Video Camera Measurement of Blood Pressure).

Use of the PPG waveform to predict BP is not limited to smartphone-based PPG. As such, there is a range of machine learning methods that are used, and methods for acquiring

data. For example, datasets may be publicly available, such as through the Multiparameter Intelligent Monitoring in Intensive Care MIMIC database, (32) or frequently acquired *via* pulse oximeter (33).

PPG waveform analysis faces challenges, several of which are outlined in (34). Briefly, PPG waveforms are an indirect measure of pressure. Finger sizes and pressing pressure vary considerably between subjects, affecting the PPG waveform and consequently influencing prediction accuracy. Diseases, such as anemia, reduce hemoglobin concentration and alter the relationship between blood volume and total hemoglobin. Other diseases alter the circulation and body temperature, in turn reducing the correlation between the peripheral pulse measured with PPG and BP. Nevertheless, arterial BP, and PPG signals have a high similarity in their morphology, with potential for determining whether patients are normotensive or hypertensive (35). Consequently, studies employing PPG waveform analysis *via* smartphones (as detailed below) have potential for predicting BP and diagnosing conditions such as hypertension.

Lamonaca et al. (5) trained a neural network on 15,000 PPG pulses with associated pressure from MIMIC. Features from the PPG pulses relating to length of time in portions of the cardiac cycle, systolic upstroke time, diastolic time, and cardiac period were extracted and input into the neural network algorithm for training on the database. This trained network was applied to PPG pulses acquired with a smartphone and compared to pressure measured in the arm with a cuff. Over five experiments, the maximum difference between predicted and reference systolic values was 11 and 12 mmHg between diastolic values. This is above the accuracy threshold of 5 ± 8 mmHg according to the Association for the Advancement of Medical Instrumentation.

Visvanathan et al. (6) analyzed 14 time domain features of the PPG waveform, in addition to height, weight, and age. These were input into a linear regression or support vector machine classification model to estimate BP. In 2014, a similar analysis was performed that included additional features in the time and frequency domain (7).

To reduce noise, Banerjee et al. (8) approximated the PPG signal as a sum of two Gaussian functions. The cardiovascular system was modeled as a circuit, with a peripheral resistance (R) and arterial compliance (C). SP and DP were expressed as an exponential function of R and C. Seven features from the PPG signal modeled as the sum of Gaussians were input into a neural network to calculate R and C, enabling estimation of SP and DP.

Gao et al. (13) applied a discrete wavelet transform to the PPG signal to extract periodic features. Feature selection was performed using a linear support vector machine, followed by training a non-linear support vector machine to predict BP. The mean error for DP was 4.6 ± 4.3 mmHg, and the error for SP of 5.1 ± 4.3 mmHg.

Datta et al. (15) analyzed the ratio of PPG features, systolic upstroke time, the inverse of systolic upstroke time squared, and age and body mass index (six features total). Mean absolute error values of 7.4% (systolic) and 9.1% (diastolic) were obtained. The advantage of measuring the ratio of features is to reduce dependence on use of a particular camera or smartphone. This provides an algorithm that is applicable between phones and manufacturers, and is less likely influenced by sensor or phone properties.

Dey et al. (18) incorporated 233 total features in the time and frequency domain. A mean absolute error for DP of 5.0 ± 6.1 mmHg, and 6.9 ± 9.0 mmHg for SP was calculated.

These studies used feature extraction to predict BP, which may be influenced by sensor and signal quality, and vary between studies. It is possible to eliminate feature extraction, as demonstrated by Baek et al. (21). In this study, a convolutional neural network was applied to PPG signals without feature extraction. They obtained a mean absolute error for DP of 4.92 ± 2.42 mmHg, and 5.28 ± 1.80 for SP.

Large-scale studies in patients with a range of BP values are necessary to assess the accuracy of an app or technique, since inaccuracy may be induced at higher or lower BP ranges. This could be attributable to utilizing training data largely from normotensive populations, healthy/younger individuals, or using invasive BP measures (instead of cuffs) as reference data (26). The Preventicus BP estimation algorithm overestimated SP in low BP ranges (< 130 mmHg), and underestimated SP in medium BP ranges (130–160 mmHg) in pregnant women (25). In a recent study of more than 900 individuals (300 hypertensive) using the Preventicus app, overestimation of SP occurred at lower SP, and underestimation at higher SPs, with decreasing performance at higher pressures. OptiBP, a smartphone app, possessed high accuracy when tested on a range of BPs (hypotensive to hypertensive, 101 subjects total), suggesting it is more likely to be applicable to the general population than Preventicus (22). It is difficult to determine presently the reasons similar techniques (OptiBP vs. Preventicus) are not equally successful since they both follow a similar methodology (finger-pressing). The Preventicus algorithm is a combination of frequency and morphology analysis, and utilizes the knowledge that time difference between the notch and peak represent peripheral resistance and depends on BP (26). The OptiBP algorithm obtains an average waveform over multiple measurements, with less weight attributed to pulses with abnormal morphology. Derivative-based features are extracted from the pulse, and a non-linear model is used to predict BP. The final BP is determined following a calibration procedure (22).

Pulse Transit Time (PTT): Using Signals From Multiple Locations to Estimate Blood Pressure

Calculating Blood Pressure From Pulse Transit Time

Waveform analysis extracts multiple pulse features, inputting them into machine learning algorithms to calculate BP. Prediction accuracy is dependent on selection of correct features based on the waveform shape, which may depend on characteristics of the sensor. Consequently, algorithms developed on one smartphone may not be transferable to a different phone, since each uses a different set of sensors. Algorithms may require further development to automatically detect subtle features in the pulses relatable to pressure, along with acquisition

of large datasets for machine learning. These techniques are often limited to the fingertip (one region), whereas pulse information from multiple regions could increase accuracy of BP estimation algorithms (36).

PTT overcomes some obstacles of waveform analysis, requiring only measurement of relative arrival times of two pulses at different points in the body. PTT is inversely proportional to pulse wave velocity, which increases with BP. It is used as an indirect measure of BP, with a reduced PTT indicating elevated pressure.

PTT and pulse arrival time (PAT) are often used interchangeably. Technically, PTT is the time difference between two points in the PPG waveforms measured at different arterial sites. PAT represents the time difference between the R-peak of the electrocardiogram and a characteristic point in the PPG waveform, such as the foot. Since PTT is not in pressure units, methods using PTT to estimate pressure require calibration to relate the quantities. Accuracy of PTT-based BP measurements therefore depends on calibration quality, possibly requiring recalibration after several months. Calibration for each individual is performed by acquiring multiple pressure and PTT measurements and performing regression analysis to relate the two quantities. Multiple pressure and PTT measurements may be acquired through pressure perturbations such as exercise or changing posture (37).

Inconsistencies impact the accuracy or variability of BP recordings calculated from PTT. The characteristic points used to determine PPG pulse arrival time differ between studies, such as the foot or peak of the waveform, or the peak of the second derivative waveform (38). Either marker may be used, although effects of wave reflection from peripheral arteries are minimized if the foot-to-foot time delay between waveforms is used (37). Conditions for accurate BP calculation from PTT include assumptions of negligible contraction of vasculature *via* smooth muscle and negligible viscous effects, which induce PTT variations without affecting blood pressure; and minimal changes to arterial elasticity in response to disease or aging (37). Due to this final condition, periodic recalibration is required for chronic BP measurements, with calibration period depending on age. In a theoretical study, for a 30-year-old the calibration period to maintain accuracy (<1 mmHg error in BP calculation) is approximately 1-year, decreasing to 6-months for a 70-year-old (39). A study in 14 normotensive subjects (aged 20–36 years) suggest shorter calibration periods, where regression coefficients for calculating BP from PTT in a first test inaccurately predict blood pressure in a repeat test 6-months later (40).

Incorporation of PTT Into Contact-Based Smartphone Techniques

Two measurements are required to incorporate PTT into smartphones: one PPG pulse, and a PPG pulse or indicator of heartbeat. Chandrasekaran et al. (4) used two methods for calculating BP based on recording of heart sound (phonocardiogram, PCG) and finger pulse. The first method required two smartphones. One was pressed against the user's chest to record the PCG, while the other detected the PPG finger pulse. The second method was similar, except a single phone was used. A finger PPG pulse was acquired, and a customized external microphone attached to the smartphone to amplify the acoustic heart signal. Estimated BP values achieved an accuracy of approximately 95–100% when compared to a commercial BP meter. Similar techniques combining PCG and PPG are described in Junior et al. (11, 12) and Nemcova et al. (23). Such techniques have been incorporated into phone apps, such as AuraLife Instant Blood Pressure (IBP) (41). Clinical translation of the AuraLife app has not been successful, where a clinical study (85 participants, 53% with hypertension) of the app found large errors in measured pressure and low sensitivity to detecting hypertension (14).

A potential inaccuracy induced through PCG measurement is reliance of the signal on closing instead of opening of heart valves. This provides incorrect times for PTT determination since valve closure does not indicate when blood is ejected from the heart (20). Therefore, Wang et al. (20) instead investigated the phone's accelerometer to detect vibrations caused by mechanical movement of the heart (seismocardiogram, SCG). The error in DP over all subjects was 5.2 $\pm$ 2.0 mmHg (RMSE $\pm$ std). SP was not calculated since the characteristic PPG point to calculate PTT was the foot of the pulse, which measures arrival time of diastole (20).

Some studies use two cameras and define PTT as the difference in times between two characteristic points in the PPG pulses at different body locations. Liu et al. (9) assembled a prototype device with the front-camera pressed against the temple and the finger contacting the rear-camera. Tabei et al. (24) incorporated two smartphones, defining PTT as the time difference between peak locations for each fingertip PPG. SP and DP were estimated with a regression model, and compared to calculations with a reference device. Estimates demonstrated a mean absolute error for SP and DP of $\sim$2 mmHg.

NON-CONTACT VIDEO CAMERA MEASUREMENT OF BLOOD PRESSURE

A disadvantage of contact techniques is the PPG signal dependence on finger pressing force, in contrast to non-contact video camera techniques. Non-contact techniques can simultaneously measure multiple body parts and regions of the face, which are differentially innervated by sympathetic and parasympathetic neurons. This additional information could increase accuracy of camera prediction methods, compared to more homogeneous data acquisition as per finger-pressing contact techniques (36).

However, non-contact video methods are susceptible to noise and artifacts unrelated to the hemoglobin signal. For example, a dicrotic notch in the blood volume pulse is often absent in video PPG, which could affect calculation of PTT. While some signal reflected from tissue is due to hemoglobin, other light does not pass through tissue and is reflected from the skin surface. This is termed diffuse reflection (42). Other light reflection is due to non-hemoglobin components, such as melanin. In order to predict BP most accurately, and to avoid noise effects being interpreted as part of the signal, mathematical/optical

TABLE 2 | Non-contact methods for smartphone/video blood pressure (BP) measurement.

Publications	Number of subjects and additional experimental details	Accuracy of method	Video processing method
Murakami et al. (44)	10 subjects	Correlation coefficient of PTT with SP: −0.879	Single-Channel Analysis
Sugita et al. (45)	20 subjects	Correlation coefficient with SP: ∼ 0.6 for pulse wave indices from right hand	Single-Channel Analysis
Yoshioka et al. (46)	10 subjects	Correlation coefficient between PTT and SP: −0.879 * same study subjects as Murakami et al. (44) above	Single-Channel Analysis
Jain et al. (47)	45 subjects	SP: 3.90 ± 5.37 (MAE ± std) DP: 3.72 ± 5.08 mmHg (MAE ± std)	Principal Component Analysis
Jeong and Finkelstein (48)	7 subjects	Correlation between SP and PTT: −0.80 - correlation obtained by averaging across values for individual subjects provided in Jeong and Finkelstein (48)	Single-Channel Analysis
Secerbegovic et al. (49)	3 subjects	PTT calculated from ECG and video forehead signal: SP: 9.48 ± 7.13 mmHg (MAE ± std) MP: 4.48 ± 3.29 mmHg (MAE ± std) Correlation between PTT phase delay between forehead and palm video signals and SP:−0.6045	Independent Component Analysis
Huang et al. (50)	13 subjects	SP: 14.02 mmHg (RMSE) DP: 7.38 mmHg (RMSE)	Single-Channel Analysis
Khong et al. (51)	45 subjects	SP: 4.22 ± 3.15 mmHg (MAE ± std) DP: 3.24 ± 2.21 mmHg (MAE ± std)	Single-Channel Analysis
Patil et al. (52)	20 subjects	Morning session SP: 9.62 % (error rate) DP: 11.63 % (error rate) Afternoon session SP: 8.4 % (error rate) DP: 11.18 % (error rate)	Independent Component Analysis
Chen et al. (53)	2 subjects	SP: −2.40 % – 3.43 % (range of error compared to reference) DP: −6.88 % – 5.26 % (range of error compared to reference)	Mathematical/Optical Modeling
Fang et al. (54)	15 subjects	SP: 11.2 mmHg (RMSE) PP: 7.83 mmHg (RMSE)	Mathematical/Optical Modeling
Viejo et al. (55)	15 subjects (70 % training, 15 % validation, 15 % testing)	Correlation coefficient in testing phase between measured BP, heart rate and reference BP, heart rate: 0.71	Single-Channel Analysis
Oiwa et al. (56)	8 subjects	MP: range from 1.50 mmHg – 4.15 mmHg (MAE)	Independent Component Analysis
Shirbani et al. (57)	15 subjects	Slope from plot of PAT measured from video PPG vs. DP: −1.33 ± 1.70 ms/mmHg (mean ± standard error), p = 0.0024	Single-Channel Analysis
Adachi et al. (3)	10 subjects	Without body movement: SP:−1.0 ± 5.6 mmHg (ME ± std) With body movement: SP: −0.1 ± 12.2 mmHg (ME ± std)	Mathematical/Optical Modeling
Luo et al. (36)	1,328 subjects (70 % training, 15 % testing, 15 % validation) (data collected from 2,348 subjects originally)	SP: 0.39 ± 7.30 mmHg (ME ± std) DP: −0.20 ± 6.00 mmHg (ME ± std) PP: 0.52 ± 6.42 mmHg (ME ± std)	Transdermal Optical Imaging
Sugita et al. (45)	17 subjects (20 subjects originally)	Correlation coefficient of right palm index with SP: < −0.5	Single-Channel Analysis
Sugita, Noro, et al. (58)	5 subjects	SP: 25.7 mmHg (RMSE)	Single-Channel Analysis
Fan et al. (59)	6 subjects	SP: 8.42 ± 8.81 mmHg (MAE ± std) DP: 12.34 ± 7.10 mmHg (MAE ± std)	Mathematical/Optical Modeling
Takahashi et al. (60)	4 subjects	Correlation between SP and PTT measured only in face *via* video: −0.4543 (range from −0.7820 to −0.2900) - average value not provided in original publication; averaged here over the four individual values	Mathematical/Optical Modeling
Rong and Li (61)	189 subjects (191 subjects originally; 70 % training; 30 % testing)	SP: 9.97, 2.1 ± 3.35 mmHg (MAE, ME ± std) DP: 7.59, 0.79 ± 2.58 mmHg (MAE, ME ± std) - showing here result obtained with the machine learning method that produces the smallest MAE	Single-Channel Analysis

ME, mean error; MAE, mean absolute error; RMSE, root mean square error; SP, systolic pressure; DP, diastolic pressure; MP, mean pressure; PTT, pulse transit time; PPG, photoplethysmography; PP, pulse pressure; PAT, pulse arrival time.

TABLE 3 | Comparison of contact and non-contact BP measurement techniques.

	Advantages of techniques	Disadvantages of techniques	Comments on specific contact/non-contact techniques
Contact	- higher signal achievable compared to non-contact due to proximity of finger to sensor and LED - reduced sensitivity to subject motion, since the finger is pressed against the camera - less sensitive to external lighting conditions than non-contact methods	- signal may depend on finger pressing force - may require multiple sensors, such as microphone, in addition to camera - limited to certain regions of the body, such as the finger, whereas the face includes pulse information for predicting BP - may depend on height of hand relative to heart	**Oscillometry** - convenient, easy method to learn - may require a special case to sense applied pressure **Waveform Analysis** - prediction accuracy dependent on size of training data, extent to which training data reflects the characteristics of the population, and features extracted for input into machine learning algorithms **Pulse Transit Time** - easy, efficient measure that correlates with BP - data may require calibration every 6-months to 1-year, depending on age and health conditions of subject - multiple sensors required in the case of contact methods **Heart Sounds Only** - may require attachment of a stethoscope to smartphone to amplify heart sounds
Non-contact	- may image multiple regions simultaneously without additional equipment/sensors - signal does not depend on pressing force, yielding more consistency across subjects or between trials - acquisition of blood pressure through 'selfie' or short video	- sensitive to lighting conditions, angle of camera with face, and distance of camera from face - sensitivity to body and surface skin movement - relatively weak signal, since often the camera is held a distance from the face and ambient light is used as the light source	**Single-Channel Analysis** - usually green channel analyzed, followed by application of PTT or waveform analysis - susceptible to skin inhomogeneities, melanin, lighting conditions **Transdermal Optical Imaging** - use of machine learning to extract hemoglobin signal - applied in study of over 1,300 individuals to predict BP **Mathematical/Optical Modeling** - account for light reflectance from skin surface, skin movement, melanin, and lighting. They may therefore be potentially applied to a range of real-life conditions outside the laboratory. **Independent Component Analysis** - assumes each channel contains a hemoglobin component - component with the strongest signal at the heart rate selected as the pulse component **Principal Component Analysis** - may be used to determine which components of the video signal are attributable to hemoglobin pulsation - can reduce data redundancy for optimal performance of machine learning algorithms

models or machine learning techniques are used to specifically extract the hemoglobin signal component. For example, the chrominance model (CHROM) yields a higher blood volume pulse SNR than techniques such as blind source separation, even under conditions of movement due to its ability to remove non-hemoglobin components such as diffuse reflection (42). In exercise (subject movement) conditions, processing data from only a single channel often does not yield a clear blood volume pulse (43).

Consequently, a number of video camera processing techniques have been developed for overcoming deficiencies of single channel analysis, using sophisticated algorithms to minimize noise effects. This section details these techniques. Nevertheless, as will also be described, studies utilizing only a single color channel for analysis are still able to produce accurate estimates of BP or strong correlations between PTT calculations and BP.

These methods are summarized in **Table 2**, while **Table 3** compares the advantages and disadvantages of contact and non-contact techniques.

Single Channel Video Analysis

PPG signal for heart and respiratory rate calculation was detected from facial videos in 2008 by averaging pixel signal over the

green channel (28). Machine learning techniques, such as neural networks, can use features from the filtered, averaged signal to predict BP (55, 61).

An advantage of video techniques is multiple regions may be imaged simultaneously, enabling calculation of PTT using only a single sensor (camera). This was demonstrated by Jeong and Finkelstein, (48) who found a strong correlation between SP and PTT through videos acquired of the face and hand. Their study used high speed video recording (420 fps) to capture 1-min videos between/before/after exercise protocols, with the subject seated and hand on a table. Other studies utilizing PTT to estimate BP from single channel processing include: Yoshioka et al. (46), Huang et al. (50), Khong et al. (51), Murakami et al. (44) and Shirbani et al. (57). As an alternative to PTT, it may be possible to capture videos of hands at different heights, relating the difference in pulse amplitudes between the hands to BP (58). Others use phase difference between PPG waveforms as a surrogate for time delay, which has a higher correlation with SP compared to time delay methods (62). Phase difference, however, is distorted by skin inhomogeneities and may not provide a truly accurate measure of PTT (63).

A disadvantage to requiring two pulse measurement locations, as in Jeong and Finkelstein (48), is PTT varies according to the distance between face and hand or body parts selected for analysis. This will alter BP prediction when a subject adjusts their posture. To overcome potential variability in measuring PTT in different regions, Sugita et al. (45) calculated time difference (T_{BH}) between minimum values in a band-pass filtered waveform from video of the palm, and the raw waveform from the palm. They demonstrated smoothing the PPG waveform causes a phase change indicating heart-rate variability, with T_{BH} indicating the degree of distortion of the PPG waveform. T_{BH} showed a similarly strong correlation with SP as the difference between arrival times of waveforms in the palm and forehead. This suggests T_{BH} has similar accuracy as PTT in calculating BP, but should be more applicable to situations where body movement occurs, such as exercise.

Processing Video Data to Extract Hemoglobin Information

Single-channel (typically green channel) data contains hemoglobin information. The overall channel signal, however, is affected by melanin content (skin tone), lighting, and subject movement. This may be controlled in experimental situations through consistent and bright lighting, and limiting subject movement. In "real world" environments, there are a variety of skin tones, background light, and subject motion. In these situations, analysis of raw or filtered green channel data risks acquisition of a distorted or altered waveform not accounted for in algorithms relating waveform shape to BP. As detailed below, video frames may be processed using machine learning or mathematical algorithms to extract hemoglobin-dependent features of the signal that are then used to estimate BP.

Transdermal Optical Imaging (TOI)

A study by Luo et al. (36) with over 1,300 normo-tensive subjects demonstrated feasibility of non-contact BP measurement with video camera (64, 65). Each 8-bit image from the three color channels contains 8 bitplanes, with each pixel in a bitplane 0 or 1. A machine learning algorithm was trained to select bitplanes corresponding to hemodynamic changes. This technique has demonstrated successful calculation of heart rate and heart rate variability, stress, facial blood flow, BP, and flow responses to stimuli (36, 66, 67)(see **Figure 2**).

To calculate BP, transdermal blood flow data was acquired from 17 facial regions with an Apple iPhone 6. The subject was seated, with back straight and feet on the ground. Data acquisition occurred over 2-min. One hundred and twenty six features were extracted from the videos relating to pulse characteristics, such as shape, amplitude, and heart rate. An additional 29 "meta features" were selected to normalize for variation in imaging conditions across the three channels, and to account for ambient room temperature and demographic characteristics. Principal component analysis (PCA) (68) reduced data dimensions, producing 30 eigenvectors that were input into a multi-layer perceptron to calculate BP. Accuracy was approximately 95%, with an error bias of 0.39 ± 7.30 mmHg (SP), -0.20 ± 6.00 mmHg (DP), and 0.52 ± 6.42 mmHg (pulse pressure).

Mathematical and Optical Modeling

Adachi et al. (3) developed a mathematical model to determine contributions from hemoglobin, melanin, and light shadowing on video signal. Hemoglobin signal (PPG waveform) was extracted through removal of the melanin and lighting effects based on knowledge of the melanin light absorption spectrum and camera spectral sensitivity. Features based on waveform shape and PTT were obtained from the extracted PPG waveforms, and used to predict BP. Following recording of data for 30 s, BP was predicted under conditions of movement vs. no body movement. Without body movement, the mean prediction error from 10 subjects for SP was -1.0 ± 5.6 mmHg; with body movement, the prediction error was -0.1 ± 12.2 mmHg. Prediction error for DP was not included.

Fukunishi et al. (69) applied a model of light travel through the skin to extract the hemoglobin signal and blood volume pulse from an RGB camera. Using a high speed digital camera (capable of 2,000 fps, operated at 500 fps), with each recording set at 8.7 s, the Fukunishi model was applied by Takahashi et al. (60) to calculate PTT between forehead and chin, obtaining correlation coefficients between PTT and SP ranging from ~ -0.3 to -0.8 for four subjects.

The CHROM model reduces the effect of motion and light reflected from the skin that does not possess a pulsatile or blood flow component. Fan et al. (59) and Chen et al. (53) (30 s of video, at 60 fps) (53) adapted CHROM to extract PPG signals from still videos of the face and hand, calculating PTT as the time difference between peaks or the phase difference between waveforms respectively. In both cases, subjects were sitting with facing toward the camera and hand raised. In Fan et al. (59) the authors developed a solution to a problem in video PPG where the dicrotic notch is "buried" in the overall signal, causing a peak shift and mis-estimation of PTT. This was accomplished through adaptive Gaussian modeling, where the signal is modeled as a

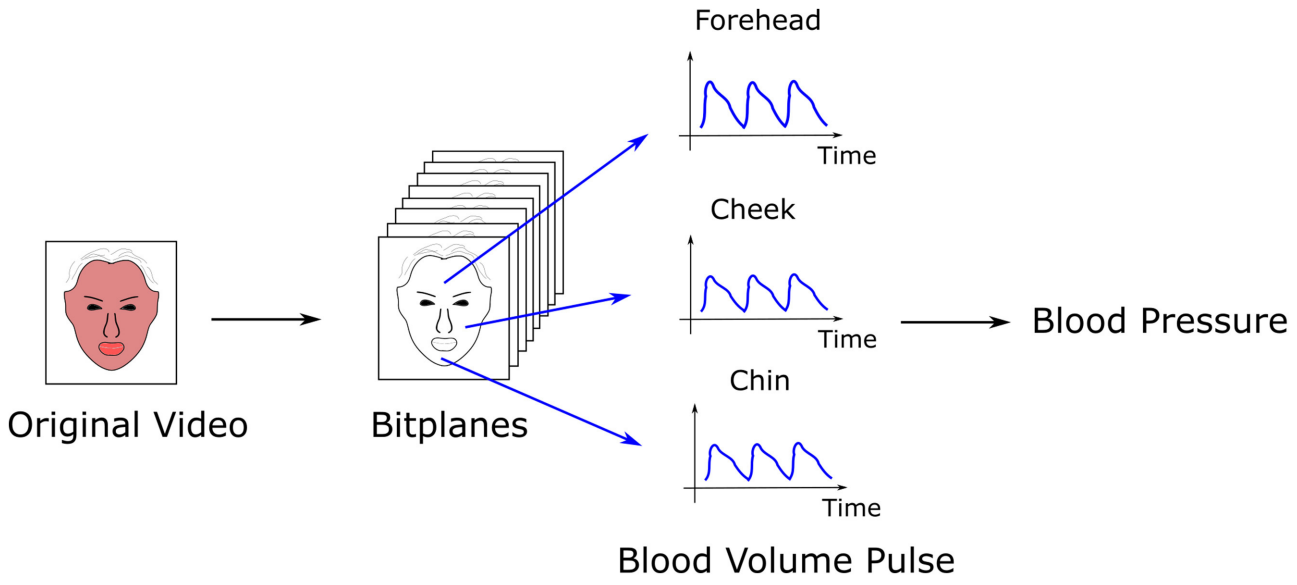


FIGURE 2 | Calculation of blood pressure (BP) by transdermal optical imaging (TOI). A machine learning algorithm selects the bitplanes from a video with the highest signal for hemoglobin, from which the blood volume pulse is extracted from multiple facial regions (17 regions total). Features of the pulse are input into a multilayer perceptron to calculate BP.

sum of two Gaussian curves and the parameters are calculated through least squares minimization. A similar algorithm to CHROM, plane orthogonal to surface (POS), (43) was used by Fang et al. (54) to predict BP. The camera frame-rate was set to 90 fps, 15 subjects were video-imaged, with 10 videos each at 45 s per video. Similar to Chen et al. and Fan et al. above, videos of the face and hand were acquired, with PTT calculated between the cheek and radial artery in palm due to these regions providing the strongest signals.

Independent Component Analysis

Independent component analysis (ICA) assumes a signal is a linear mixture of underlying sources and mathematically extracts them (70). Poh et al. (71) demonstrated data from the color channels could be decomposed into three components, one corresponding to the blood volume pulse. The signal with the largest frequency component corresponding to heart rate is assumed to correspond to the blood volume pulse (72). This analysis assumes each channel contains information on hemoglobin-related fluctuations.

In a study of three healthy individuals, Secerbegovic et al. (49) applied ICA, extracting the ICA component with the largest signal at the heart rate frequency. Using PTT to estimate BP, mean absolute error for SP and mean BP were 9.48 ± 7.13 and 4.48 ± 3.29 mmHg respectively. Frame rate was 25 fps, with subject seated and videos acquired simultaneously of face (forehead) and palm, and video duration 3-min. Patil et al. (52) extracted similar features of the PPG pulse as Adachi et al. These were input into a single hidden layer neural network, obtaining average error rates of 8.4–9.62% (systolic) and 11.18–11.63% (diastolic) between afternoon and evening sessions. Subjects were permitted small head movements to simulate realistic work conditions. Oiwa et al. (56) correlated facial PPG amplitude with reference BP following ICA, obtaining a mean absolute error in the range 1.50–4.15 over eight subjects. Data was acquired over a series of 2-min resting state segments with eyes closed and 1-min cold stimulus state segment, where subjects placed their hand in a cold (14°C) water bath with eyes opened.

Principal Component Analysis

Most of the video signal is not attributed to blood flow fluctuations. PCA (73) calculates the main components that contribute to signal intensity variation in an image. Jain et al. (47) defined the PPG signal for each frame as the difference between the raw video data in the red channel and the main principal components. Twenty features in the time and frequency domain (6, 74) were input into a polynomial regression algorithm (75) to calculate BP. Mean absolute error was 3.72 ± 5.08 mmHg for DP and 3.90 ± 5.37 mmHg for SP. Subjects were seated still with eyes closed. Videos were acquired over 1-min. The initial and final 5-s of the videos were discarded, with the best 10-s of the remaining video processed further for analysis.

DISCUSSION, PERSPECTIVES, AND FUTURE OUTLOOK

This review highlighted smartphone and video camera techniques for measuring BP. Wearables, such as watches or similar devices (76, 77), are outside the scope of the review, although they increasingly play an important role in the field of BP monitoring. Most smartphone methods for predicting BP are PPG-based. Although Peng et al. (10) attached a stethoscope to the smartphone microphone, using heart sounds only to estimate BP, such studies are relatively infrequent.

Smartphone BP monitors are potentially applicable to ambulatory BP monitoring (examining BP continuously during the day). Measurements in a doctor's office are affected by "white coat" syndrome, where patients are recorded as possessing a higher BP recording when measured clinically. Since BP varies throughout the day, improved understanding of how and why these changes occur could assist physicians in prescribing medication. Smartphones may measure pressure easily, with no additional equipment required beyond the phone or case. In contrast, cuffs may be awkward and bulky, inducing arm soreness or rashes after multiple daily uses (78). Although a smartphone may not be used at night when the user is asleep, video cameras with infrared light can monitor vital signs using similar techniques as those developed for the smartphone (79).

Contact methods have higher signal-to-noise-ratio due to the proximity of skin to the sensor and usage of light beyond ambient light (i.e., LED from phone). Additional features in the pulse wave are distinguishable in contact vs. non-contact methods, such as secondary peaks. In using machine learning algorithms to relate waveform shape to BP, these additional features are helpful in increasing the accuracy of the prediction as they may be affected by BP. Contact methods are also less dependent on motion since the finger remains pressed against the camera.

Non-contact methods possess other advantages. Contact methods are influenced by the force with which the finger is pressed against the camera. This could vary between individuals, and between trials conducted in the same individual. For waveform analysis, data from only a single region (fingertip) is acquired. Non-contact methods acquire images of multiple regions simultaneously. This enables simultaneous analysis of waveform shape and PTT. Different regions of the body could be affected differently by the sympathetic and parasympathetic nervous system, which is not accounted for through analysis of a single region. Contact methods that incorporate PTT may require additional sensors such as the microphone which produce a weak signal, or a case containing multiple sensors such as an electrocardiogram and PPG (27, 80). Oscillometry may require a special case to simultaneously measure blood volume changes and applied pressure (16), or the 3D Touch feature on the iPhone X (17), which is not available in all phone types. Non-contact methods do not face this issue since all phones are equipped with a camera.

Several published methods possess a mean error $\pm$ standard deviation within the clinically acceptable 5 ± 8 mmHg. This is not necessarily translatable to the clinic, as indicated in the trial of the AuraLife app (14). Small sample sizes of published studies do not necessarily hold for large populations. It is difficult to compare published techniques based on accuracy measures, or to predict which will be successful when applied to large populations. Studies mostly use normotensive subjects, which may reduce prediction accuracy at high and low BP. In a follow-up study of the Anura TOI-based smartphone app (81), lower BPs tended to be overpredicted, while higher BPs were underpredicted. This was attributed to more limited training data at the extreme ends of BP (81).

Many studies do not meet criteria for validating BP devices. In addition to the AAMI criteria of MAE 5 ± 8 mmHg, several other criteria are listed (82), such as: at least 85 subjects; probability of tolerable errors <10 mmHg is at least 85 %, where a tolerable error is calculated as an average of three measurements against a reference BP; reference BP measurements acquired by two observers; and recording of number of absolute BP differences within 5, 10, and 15 mmHg. The protocol must cover a sufficient time frame to ensure that as the measurement device ages, accuracy is not reduced (83). Many clinical validation protocols are tested on new models, without testing sustained accuracy over time, even though BP devices such as sphygmomanometers decline in accuracy over 18-months (84, 85). Over time, an individual may undergo physical changes in skin (i.e., aging) or changes in size, which may influence PPG extraction and BP estimation. Nevertheless, since non-contact methods should be applicable across camera types and imaging conditions, algorithms trained on data from a variety of subject types (age, sex, still vs. movement, range of skin tones and types) should be accurate for a sustainable time period.

TOI possesses advantages to other techniques for clinical translation. The sample size of Luo et al. (36) is over 1,300; 155 features over 17 ROIs relating to waveform shape, population demographics and PTT predicted BP. This is advantageous over methods with small sample size and those that only measure PTT or analyze a single region. A disadvantage of limiting measurements to PTT is the phase shift used to measure PTT partially depends on skin variability/inhomogeneity, affecting PTT accuracy (63).

Luo et al. (36) acquired images under strict conditions: normotensive, and consistent lighting and camera angle. Future studies may include a wider range of pressures to determine whether TOI may predict hypertension, and a variety of camera angles and lighting conditions.

Video-camera processing techniques other than TOI may be successful if applied to a larger population, or through optimized analysis of PPG waveform features. For example, Adachi et al. (3) only used eight features of the PPG waveform and the time difference between two pulse waves from 10 subjects as input features for learning. Due to the small sample size, it is difficult to extrapolate their success to larger populations. Other algorithms propose first classifying PPG waveforms into one of three categories (hypotensive, normotensive, or hypertensive), calculating BP according to the category to which the PPG pulse was assigned (86). This method is an improvement over traditional techniques which apply a generic algorithm to calculate BP regardless of the subject's BP range. Eulerian video magnification is a video processing technique that enhances blood flow signal (87). It has been applied to calculate PTT in videos of wrist and neck, indicating its applicability to BP measurements (88).

Future experiments may forego traditional image processing techniques. Chen and McDuff (89) developed DeepPhys, a convolutional neural network, and applied it to video frames to recover the blood volume pulse, measuring heart and breath rate. Convolutional neural network techniques may only produce a single, total blood volume pulse. TOI, however, predicts BP from multiple facial regions. This is advantageous since each region is differentially innervated, possibly influencing pressure

prediction. The potential to reduce dependence on feature extraction is exemplified in the study by Baek et al. (21), who applied convolutional neural networks to PPG data without feature extraction to predict BP.

Additional possibilities are outlined in (90). This includes (A) development of techniques and models robust to "real-life" conditions such shadowing or movement; (B) application of infrared light, which acquires videos in dark conditions and may be less sensitive to variable ambient lighting; (C) development of a large publicly available dataset, where different algorithms may be applied and compared; (D) extraction of additional features beyond PTT; (E) development of a model requiring fewer calibrations.

This paper highlighted and compared the variety of methods available for measuring BP with smartphones/video cameras. It emphasizes variations in experimental design and the relevancy of these variations in developing methods for BP measurement that are efficient, easy to use, and non-invasive. This permits regular BP monitoring, contributing to early hypertension or cardiovascular disease risk detection. Possibilities for successful BP monitoring was demonstrated in early studies, such as Lamonaca et al. (5), that used the rear camera of the phone in combination with the LED to extract a strong pulse signal in the finger. Later studies included more features for analysis (18), or used convolutional neural networks to predict BP without waveform feature extraction (21). In parallel, non-contact methods were developed that overcame deficiencies of contact methods, such as BP prediction limited to a small field of view (fingertip). Non-contact video methods initially processed data from single channels, which is affected by motion, lighting, and features not related to hemoglobin such as melanin. Recent techniques, such as TOI or optical models, have extracted the hemoglobin signal specifically and are less sensitive to artifacts from non-hemoglobin sources.

Smartphone and video BP measurements will likely become more common. Compared to cuff-based techniques, they are cost-effective and convenient. Using a single video, BP may be combined with heart rate detection and stress assessment (66), blood oxygen saturation (23), and blood flow. This technology can be developed for improved digital health consultations to assess a number of health conditions. For example, measurement of the multiple parameters described above currently requires a visit to a medical health professional. This is time consuming, and may necessitate meetings with multiple health professionals.

As methods and techniques for processing video images advances, it is foreseeable a video consultation over Zoom can relay to a doctor/nurse a patient's vitals (heart rate, BP, oxygen saturation, respiratory rate, etc.) and relate this information to stroke risk or susceptibility to cardiovascular disease. This information would be provided in real-time and reduce the need for manual measurements by a medical professional. In addition, through continuous and regular daily monitoring of their own vitals privately with a smartphone, patients can be alerted *via* algorithms whether further treatment is necessary. Ambulatory BP monitoring (BP monitoring at regular intervals during day, such as *via* a video camera) may detect abnormal variations in BP not detectable with a single BP measurement session at a doctor's office. Furthermore, ambulatory BP has been demonstrated to correlate more strongly with organ damage caused by hypertension than BP measurements conducted in a clinical setting (91). Data acquired *via* smartphone may be automatically directed to a health care team for further discussions and medical decisions. Such technology can be extended to blood sugar measurements for diabetes patients, and other information related to cholesterol, fats, and hemoglobin (92). Video technology could be on the cusp of a future where a patient's home is transformed into a "smartphone-based doctor's office" where numerous cardiovascular or blood-related metrics are assessed that would previously require expertise and communication across multiple health divisions.

Overall, smartphones and video cameras will provide a more complete and earlier assessment of cardiovascular physiology, helping to prevent stroke and blood vessel-related disorders.

AUTHOR'S NOTE

This review discusses video camera methods for measuring blood pressure. During Covid, there has been an increase in the number of virtual consultations with patients. Thus, there is increased interest in developing technologies that will allow patients to monitor vital signs from home. This prevents unnecessary additional trips to doctors, provides daily information on changes in cardiovascular health, and may help detect signs leading to stroke or disease. In 2019, our lab published a paper in Circulation: Cardiovascular Imaging: "Smartphone-based blood pressure measurement using transdermal optical imaging technology" by Luo et al. In a study of over 1300 subjects, we demonstrated accurate blood pressure prediction *via* video camera. This was achieved through an imaging technology, transdermal optical imaging (TOI), that uses machine learning to extract a cardiovascular pulse signal from facial videos. As discussed in our review, there are numerous additional smartphone technologies beyond TOI that can also be used. Our review compares the different techniques and their technical aspects such as image processing and data analysis. The review concludes with the future of video camera blood pressure measurement, and how it can be combined with measurement of other metrics for a more complete assessment of cardiovascular health.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by Natural Sciences and Engineering Research Council of Canada to Z-PF (RGPIN-2014-0671).

REFERENCES

1. Glazier RH, Green ME, Wu FC, Frymire E, Kopp A, Kiran T. Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada. *CMAJ.* (2021) 193:E200–10. doi: 10.1503/cmaj.202303
2. Kario K. Management of hypertension in the digital era. Hypertension. (2020) 76:640–50. doi: 10.1161/HYPERTENSIONAHA.120.14742
3. Adachi Y, Edo Y, Ogawa R, Tomizawa R, Iwai Y, Okumura T. Noncontact blood pressure monitoring technology using facial photoplethysmograms. *Conference Proceedings: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC). (2019) 2411–5. doi: 10.1109/EMBC.2019.8856439
4. Chandrasekaran V, Dantu R, Jonnada S, Thiyagaraja S, Subbu KP. Cuffless differential blood pressure estimation using smart phones. *IEEE Trans Biomed Eng.* (2013) 60:1080–9. doi: 10.1109/TBME.2012.2211078
5. Lamonaca F, Barbe K, Kurylyak Y, Grimaldi D, Van Moer W, Furfaro A, et al. Application of the Artificial Neural Network for blood pressure evaluation with smartphones. *Conference Proceedings: 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS*). (2013) 408–12. doi: 10.1109/IDAACS.2013.6662717
6. Visvanathan A, Sinha A, Pal A. Estimation of blood pressure levels from reflective Photoplethysmograph using smart phones. *Conference Proceedings: 13th IEEE International Conference on BioInformatics and BioEngineering.* (2013) 1–5. doi: 10.1109/BIBE.2013.6701568
7. Visvanathan A, Banerjee R, Choudhury RD, Sinha A, Kundu S. Smart phone based blood pressure indicator. *Conference Proceedings: MobileHealth '14: Proceedings of the 4th ACM MobiHoc Workshop on Pervasive Wireless Healthcare*. (2014) 19–24. doi: 10.1145/2633651.2633657
8. Banerjee R, Ghose A, Choudhury AD, Sinha A, Pal A. Noise cleaning and Gaussian modeling of smart phone photoplethysmogram to improve blood pressure estimation. *Conference Proceedings: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* (2015) 967–71. doi: 10.1109/ICASSP.2015.7178113
9. Liu H, Ivanov K, Wang Y, Wang L. Toward a smartphone application for estimation of pulse transit time. *Sensors (Basel).* (2015) 15:27303–21. doi: 10.3390/s151027303
10. Peng RC, Yan WR, Zhang NL, Lin WH, Zhou XL, Zhang YT. Cuffless and continuous blood pressure estimation from the heart sound signals. *Sensors (Basel).* (2015) 15:23653–66. doi: 10.3390/s150923653
11. Junior AD, Murali S, Rincon F, Atienza D. Estimation of blood pressure and pulse transit time using your smartphone. *Conference Proceedings: 2015 Euromicro Conference on Digital System Design.* (2015) 173–80. doi: 10.1109/DSD.2015.90
12. Junior AD, Murali S, Rincon F, Atienza D. Methods for reliable estimation of pulse transit time and blood pressure variations using smartphone sensors. *Microprocess Microsyst.* (2016) 46:84–95. doi: 10.1016/j.micpro.2016.06.001
13. Gao SC, Wittek P, Zhao L, Jiang WJ. Data-driven estimation of blood pressure using photoplethysmographic signals. *Conference Proceedings: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2016) 766–9. doi: 10.1109/EMBC.2016.7590814
14. Plante TB, Urrea B, MacFarlane ZT, Blumenthal RS. Miller 3rd ER, Appel LJ, Martin SS. Validation of the instant blood pressure smartphone app. *JAMA Intern Med.* (2016) 176:700–2. doi: 10.1001/jamainternmed.2016.0157
15. Datta S, Choudhury AD, Chowdhury A, Banerjee T, Banerjee R, Bhattacharya S, et al. Novel statistical post processing to improve blood pressure estimation from smartphone photoplethysmogram. *Conference Proceedings: HumanSys'17: Proceedings of the First International Workshop on Human-centered Sensing, Networking, and Systems*. (2017) 31–6. doi: 10.1145/3144730.3144737
16. Chandrasekhar A, Kim CS, Naji M, Natarajan K, Hahn JO, Mukkamala R. Smartphone-based blood pressure monitoring *via* the oscillometric finger-pressing method. *Sci Transl Med.* (2018) 10:eaap8674. doi: 10.1126/scitranslmed.aap8674
17. Chandrasekhar A, Natarajan K, Yavarimanesh M, Mukkamala R. An iPhone application for blood pressure monitoring *via* the oscillometric finger pressing method. *Sci Rep.* (2018) 8:13136. doi: 10.1038/s41598-018-31632-x
18. Dey J, Gaurav A, Tiwari VN. InstaBP: Cuff-less blood pressure monitoring on smartphone using single PPG sensor. *Conference Proceedings: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2018) 5002–5. doi: 10.1109/EMBC.2018.8513189
19. Matsumara K, Rolfe P, Toda S, Yamakoshi T. Cuffless blood pressure estimation using only a smartphone. *Sci Rep.* (2018) 8:7298. doi: 10.1038/s41598-018-25681-5
20. Wang EJ, Zhu J, Jain M, Lee TJ, Saba E, Nachman L, et al. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. *Conference Proceedings: CHI'18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. (2018) 425: 1–9. doi: 10.1145/3173574.3173999
21. Baek S, Jang J, Cho SH, Choi JM, Yoon S. Blood pressure prediction by a smartphone sensor using fully convolutional networks. *Conference Proceedings: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2020) 188–91. doi: 10.1109/EMBC44109.2020.9175902
22. Schoettker P, Degott J, Hofmann G, Proença M, Bonnier G, Lemkaddem A, et al. Blood pressure measurements with the OptiBP smartphone app validated against reference auscultatory measurements. *Sci Rep.* (2020) 10:17827. doi: 10.1038/s41598-020-74955-4
23. Nemcova A, Jordanova I, Varecka M, Smisek R, Marsanova L, Smital L, et al. Monitoring of heart rate, blood oxygen saturation, and blood pressure using a smartphone. *Biomed Signal Process Control.* (2020) 59:101928. doi: 10.1016/j.bspc.2020.101928
24. Tabei F, Gresham JM, Askarian B, Jung K, Chong JW. Cuff-less blood pressure monitoring system using smartphones. *IEEE Access.* (2020) 8:11534–45. doi: 10.1109/ACCESS.2020.2965082
25. Raichle CJ, Eckstein J, Lapaire O, Leonardi L, Brasier N, Vischer AS, et al. Performance of a blood pressure smartphone app in pregnant women: The iPARR trial (iPhone app compared with standard RR measurement). *Hypertension.* (2018) 71:1164–9. doi: 10.1161/HYPERTENSIONAHA.117.10647
26. Dörr M, Weber S, Birkemeyer R, Leonardi L, Winterhalder C, Raichle CJ, et al. iPhone App compared with standard blood pressure measurement –The iPARR trial. *Am Heart J.* (2021) 233:102–8. doi: 10.1016/j.ahj.2020.12.003
27. Elgendi M, Fletcher R, Liang Y, Howard N, Lovell NH, Abbott D, et al. The use of photoplethysmography for assessing hypertension. *NPJ Digit Med.* (2019) 2:60. doi: 10.1038/s41746-019-0136-7
28. Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. *Opt Express.* (2008) 16:21434–45. doi: 10.1364/OE.16.021434
29. Berger A. Oscillatory blood pressure monitoring devices. *BMJ.* (2001) 323:919. doi: 10.1136/bmj.323.7318.919
30. Chandrasekhar A, Yavarimanesh M, Hahn JO, Sung SH, Chen CH, Cheng HM, et al. Formulas to explain popular oscillometric blood pressure estimation algorithms. *Front Physiol.* (2019) 10:1415. doi: 10.3389/fphys.2019.01415
31. Mauck GW, Smith CR, Geddes LA, Bourland JD. The meaning of the point of maximum oscillations in cuff pressure in the indirect measurement of blood pressure–part ii. *J Biomech Eng.* (1980) 102: 28–33. doi: 10.1115/1.3138195
32. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. a freely accessible critical care database. *Sci Data.* (2016) 3:160035. doi: 10.1038/sdata.2016.35
33. Ghamri Y, Proença M, Hofmann G, Renevey P, Bonnier G, Braun F, Axis A, Lemay M, Schoettker P. Automated pulse oximeter waveform analysis to track changes in blood pressure during anesthesia induction: a proof-of-concept study. *Anesth Analg.* (2020) 130:1222–33. doi: 10.1213/ANE.0000000000004678
34. Xing X, Ma Z, Zhang M, Zhou Y, Dong W, Song M. An onobtrusive and calibration-free blood pressure estimation method using photoplethysmography and biometrics. *Sci Rep.* (2019) 9:8611. doi: 10.1038/s41598-019-45175-2
35. Martínez G, Howard N, Abbott D, Lim K, Ward R, Elgendi M. Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure? *J Clin Med.* (2018) 7:316. doi: 10.3390/jcm7100316
36. Luo H, Yang D, Barszczyk A, Vempala N, Wei J, Wu SJ, et al. Smartphone-based blood pressure measurement using transdermal optical imaging technology. *Circ Cardiovasc Imaging.* (2019) 12:e008857. doi: 10.1161/CIRCIMAGING.119.008857

37. Mukkamala R, Hahn JO, Inan OT, Mestha LK, Kim CS, Töreyin H, et al. Toward ubiquitous blood pressure monitoring *via* pulse transit time: theory and practice. *IEEE Trans Biomed Eng.* (2015) 62:1879–901. doi: 10.1109/TBME.2015.2441951
38. Peter L, Noury N. Cerny M. A review of methods for non-invasive and continuous blood pressure monitoring: Pulse transit time method is promising? *IRBM.* (2014) 35:271–82. doi: 10.1016/j.irbm.2014.07.002
39. Mukkamala R, Hahn JO. Toward ubiquitous blood pressure monitoring *via* pulse transit time: Predictions on maximum calibration period and acceptable error limits. *IEEE Trans Biomed Eng.* (2018) 65:1410–20. doi: 10.1109/TBME.2017.2756018
40. Wong MYM, Poon CCY, Zhang YT. An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: a half year study on normotensive subjects. *Cardiovasc Eng.* (2009) 9:32–8. doi: 10.1007/s10558-009-9070-7
41. AuraLife: Instant Blood Pressure. Available online at: www.instantbloodpressure.com
42. de Haan G, Jeanne V. Robust pulse-rate from chrominance-based rPPG. *IEEE Trans Biomed Eng.* (2013) 60:2878–86. doi: 10.1109/TBME.2013.2266196
43. Wang W, den Brinker AC, Stuijk S, de Haan G. Algorithmic principles of remote PPG. *IEEE Trans Biomed Eng.* (2017) 64:1479–1491. doi: 10.1109/TBME.2016.2609282
44. Murakami K, Yoshioka M, Ozawa J. Non-contact pulse transit time measurement using imaging camera, and its relation to blood pressure. *Conference Proceedings: 2015 14th IAPR International Conference on Machine Vision Applications (MVA).* (2015) 414–7. doi: 10.1109/MVA.2015.7153099
45. Sugita N, Yoshizawa M, Abe M, Tanaka A, Homma N, Yambe T. Contactless technique for measuring blood-pressure variability from one region in video plethysmography. *J Med Biol Eng.* (2019) 3:76–85. doi: 10.1007/s40846-018-0388-8
46. Yoshioka M, Murakami K, Ozawa J. Improved human pulse peak estimation using derivative features for noncontact pulse transit time measurements. *Conference Proceedings: 2015 International Joint Conference on Neural Networks (IJCNN).* (2015) 1–6. doi: 10.1109/IJCNN.2015.7280486
47. Jain M, Deb S, Subramanyam AV. Face video based touchless blood pressure and heart rate estimation. *Conference Proceedings: 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP).* (2016) 1–5. doi: 10.1109/MMSP.2016.7813389
48. Jeong IC, Finkelstein J. Introducing contactless blood pressure assessment using a high speed video camera. *J Med Syst.* (2016) 40:77. doi: 10.1007/s10916-016-0439-z
49. Secerbegovic A, Bergsland J, Halvorsen PS, Suljanovic N, Mujcic A, Balasingham I. Blood pressure estimation using video plethysmography. *Conference Proceedings: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI).* (2016) 461–4. doi: 10.1109/ISBI.2016.7493307
50. Huang PW, Lin CH, Chung ML, Lin TM, Wu BF. Image based contactless blood pressure assessment using pulse transit time. *Conference Proceedings: 2017 International Automatic Control Conference (CACS).* (2017) 1–6. doi: 10.1109/CACS.2017.8284275
51. Khong WL, Rao NSVK, Mariappan M. Blood pressure measurements using non-contact video imaging techniques. *Conference Proceedings: 2017 IEEE 2nd International Conference on Automatic Control and Intelligent Systems (I2CACIS).* (2017) 35–40. doi: 10.1109/I2CACIS.2017.8239029
52. Patil OR, Gao Y, Li B, Jin Z. CamBP: a camera-based, non-contact blood pressure monitor. *Conference Proceedings: Ubicomp '17: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers.* (2017) 524–9. doi: 10.1145/3123024.3124428
53. Chen P, Zhou Z, Zhao T, Cao J, Yang H. Non-contact blood pressure measurement based on pulse transit time. *Conference Proceedings: Proc SPIE 10806, Tenth International Conference on Digital Image Processing (ICDIP 2018).* (2018) 108065J. doi: 10.1117/12.2503036
54. Fang YF, Huang PW, Chung ML, Wu BF. A feature selection method for vision-based blood pressure measurement. *Conference Proceedings: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* (2018) 2158–63. doi: 10.1109/SMC.2018.00371
55. Viejo CG, Fuentes S, Torrico DD, Dunshea FR. Non-contact heart rate and blood pressure estimations from video analysis and machine learning modelling applied to food sensory responses: A case study for chocolate. *Sensors (Basel).* (2018) 18:1802. doi: 10.3390/s18061802
56. Oiwa K, Bando S, Nozawa A. Contactless blood pressure sensing using facial visible and thermal images. *Artif Life Robot.* (2018) 23:387–94. doi: 10.1007/s10015-018-0450-1
57. Shirbani F, Blackmore C, Kazzi C, Tan I, Butlin M, Avolio AP. Sensitivity of video-based pulse arrival time to dynamic blood pressure changes. *Conference Proceedings: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2018) 3639–41. doi: 10.1109/EMBC.2018.8513058
58. Sugita N, Noro T, Yoshizawa M, Ichiji K, Yamaki S, Homma N. Estimation of absolute blood pressure using video images captured at different heights from the heart. *Conference Proceedings: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2019) 4458–61. doi: 10.1109/EMBC.2019.8856362
59. Fan X, Ye Q, Yang X, Choudhury SD. Robust blood pressure estimation using an RGB camera. *J Ambient Intell Humaniz Comput.* (2020) 11:4329–36. doi: 10.1007/s12652-018-1026-6
60. Takahashi R, Ogawa-Ochiai K, Tsumura N. Non-contact method of blood pressure estimation using only facial video. *Artif Life Robot.* (2020) 25:343–50. doi: 10.1007/s10015-020-00622-6
61. Rong M, Li K. blood pressure prediction method based on imaging photoplethysmography in combination with machine learning. *Biomed Signal Process Control.* (2021) 64:102328. doi: 10.1016/j.bspc.2020.102328
62. Sugita N, Obara K, Yoshizawa M, Abe M, Tanaka A, Homma N. Techniques for estimating blood pressure variation using video images. *Conference Proceedings: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2015) 4218–21. doi: 10.1109/EMBC.2015.7319325
63. Moço A, Stuijk S, van Gastel M, de Haan G. Impairing factors in remote-PPG pulse transit time measurements on the face. *Conference Proceedings: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* (2018) 1439–14398. doi: 10.1109/CVPRW.2018.00184
64. Mukkamala R. Blood pressure with a click of a camera? *Circ Cardiovasc Imaging.* (2019) 12:e009531. doi: 10.1161/CIRCIMAGING.119.009531
65. Barszczyk A, Lee K. Measuring blood pressure: from cuff to smartphone. *Curr Hypertens Rep.* (2019) 21:84. doi: 10.1007/s11906-019-0990-3
66. Wei J, Luo H, Wu SJ, Zheng PP, Fu G, Lee K. Transdermal optical imaging reveal basal stress *via* heart rate variability analysis: a novel methodology comparable to electrocardiography. *Front Psychol.* (2018) 9:98. doi: 10.3389/fpsyg.2018.00098
67. Liu J, Luo H, Zheng PP, Wu SJ, Lee K. Transdermal optical imaging revealed different spatiotemporal patterns of facial cardiovascular activities. *Sci Rep.* (2018) 8:10588. doi: 10.1038/s41598-018-28804-0
68. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci.* (2016) 374:20150202. doi: 10.1098/rsta.2015.0202
69. Fukunishi M, Kurita K, Yamamoto S, Tsumura N. Non-contact video-based estimation of heart rate variability spectrogram from hemoglobin composition. *Artif Life Robot.* (2017) 22:457–63. doi: 10.1007/s10015-017-0382-1
70. Common P. Independent component analysis a new concept? *Signal Process.* (1994) 36:287–314. doi: 10.1016/0165-1684(94)90029-9
71. Poh MZ, McDuff DJ, Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt Express.* (2010) 18:10762–74. doi: 10.1364/OE.18.010762
72. McDuff D, Blackford E. iPhys: an open non-contact imaging-based physiological measurement toolbox. *Conference Proceedings: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2019) 6521–24. doi: 10.1109/EMBC.2019.8857012
73. Lewandowska M, Nowak J. Measuring pulse rate with a webcam. *J Med Imaging Health Inform.* (2012) 2:87–92. doi: 10.1166/jmihi.2012.1064
74. Kurylyak Y, Lamonaca F, Grimaldi D. A Neural Network-based method for continuous blood pressure estimation from a PPG signal. *Conference Proceedings: 2013 IEEE International*

Instrumentation and Measurement Technology Conference (I2MTC). (2013) 280–83. doi: 10.1109/I2MTC.2013.6555424
75. Goldberg Y, Elhadad M. splitSVM: fast, space-efficient, non-heuristic, polynomial kernel computation for NLP applications. *Conference Proceedings: HLT-Short '08: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers.* (2008) 237–40. doi: 10.3115/1557690.1557758
76. Kario K, Shimbo D, Tomitani N, Kanegae H, Schwartz JE, Williams B. The first study comparing a wearable watch-type blood pressure monitor with a conventional ambulatory blood pressure monitor on in-office and out-of-office settings. *J Clin Hypertens (Greenwich).* (2020) 22:135–41. doi: 10.1111/jch.13799
77. Moon JH, Kang MK, Choi CE, Min J, Lee HY, Lim S. Validation of a wearable cuff-less wristwatch-type blood pressure monitoring device. *Sci Rep.* (2020) 10:19015. doi: 10.1038/s41598-020-75892-y
78. Cleveland Clinic, 24-Hour Ambulatory Blood Pressure Monitoring. Available online at: https://my.clevelandclinic.org/health/diagnostics/16330-24-hour-ambulatory-blood-pressure-monitoring.
79. van Gastel M, Stuijk S, Overeem S, van Dijk JP, van Gilst MM, de Haan G. Camera-based vital signs monitoring during sleep - A proof of concept study. *IEEE J Biomed Health Inform.* (2021) 25:1409–18. doi: 10.1109/JBHI.2020.3045859
80. Patel H. *Mobile Device Casing for Health Monitoring.* United States Patent 9619623B2 (2017).
81. Yang D, Xiao G, Wei J, Luo H. Preliminary assessment of video-based blood pressure measurement according to ANSI/AAMI/ISO81060-2: 2013 guideline accuracy criteria: Anura smartphone app with transdermal optimal imaging technology. *Blood Press Monit.* (2020) 25:295–8. doi: 10.1097/MBP.0000000000000467
82. Stergiou GS, Alpert B, Mieke S, Asmar R, Atkins N, Eckert S, et al. A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) collaborative statement. *Hypertension.* (2018) 71:368–74. doi: 10.1161/HYPERTENSIONAHA.117.10237
83. IEEE standard for wearable cuffless blood pressure measuring devices. *IEEE Std 1708-2014.* (2014) 1–38. doi: 10.1109/IEEESTD.2014.6882122
84. Hodgkinson JA, Lee MM, Milner S, Bradburn P, Stevens R, Hobbs FR, Koshiaris C, Grant S, Mant J, McManus RJ. Accuracy of blood-pressure monitors owned by patients with hypertension (ACCU-RATE study): a cross-sectional, observational study in central England. *Br J Gen Pract.* (2020) 70:e548–54. doi: 10.3399/bjgp20X710381
85. Hodgkinson J, Koshiaris C, Martin U, Mant J, Heneghan C, Hobbs FR, McManus RJ. Accuracy of monitors used for blood pressure checks in English retail pharmacies: a cross-sectional observational study. *Br J Gen Pract.* (2016) 66:e309–14. doi: 10.3399/bjgp16X684769
86. Khalid SG, Liu H, Zia T, Zhang J, Chen F, Zheng D. Cuffless blood pressure estimation using single channel photoplethysmography: a two-step method. *IEEE Access.* (2020) 8:58146–54. doi: 10.1109/ACCESS.2020.2981903
87. Wu HY, Rubinstein M, Shih E, Guttag JV, Durand F, Freeman WT. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans Graph.* (2012) 31:65. doi: 10.1145/2185520.2185561
88. He X, Goubran RA, Liu XP. Using Eulerian video magnification framework to measure pulse transit time. *Conference Proceedings: 2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA).* (2014) 1–4. doi: 10.1109/MeMeA.2014.6860029
89. Chen W, McDuff D. DeepPhys: video-based physiological measurement using convolutional attention networks. *Conference Proceedings: Proceedings of the European Conference on Computer Vision (ECCV).* (2018) 349–65. doi: 10.1007/978-3-030-01216-8_22
90. Lu Y, Wang C, Meng MQH. Video-based contactless blood pressure estimation: a review. *Conference Proceedings: 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR).* (2020) 62–67. doi: 10.1109/RCAR49640.2020.9303040
91. Verdecchia P, Angeli F, Gattobigio R. Clinical usefulness of ambulatory blood pressure monitoring. *J Am Soc Nephrol.* (2004) 15:S30–3. doi: 10.1097/01.ASN.0000093241.62751.95
92. Wang HC, Chang FY, Tsai TM, Chen CH, Chen YY. Design, fabrication, and feasibility analysis of a colorimetric detection system with a smartphone for self-monitoring blood glucose. *J Biomed Opt.* (2019) 24:1–7. doi: 10.1117/1.JBO.24.2.027002

Conflict of Interest: KL reports a patent, system, and method for contactless blood pressure determination (US 10, 376, 192 B2), licensed to Nuralogix Corporation.

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

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

frontiers in Digital Health

PERSPECTIVE
published: 15 December 2021
doi: 10.3389/fdgth.2021.590194

Data Quality and Network Considerations for Mobile Contact Tracing and Health Monitoring

*Riya Dave and Rashmi Gupta**

Cognitive and Behavioural Neuroscience Laboratory, Department of Humanities and Social Sciences, Indian Institute of Technology Bombay, Mumbai, India

Machine Learning (ML) has been a useful tool for scientific advancement during the COVID-19 pandemic. Contact tracing apps are just one area reaping the benefits, as ML can use location and health data from these apps to forecast virus spread, predict "hotspots," and identify vulnerable groups. However, to do so, it is first important to ensure that the dataset these apps yield is accurate, free of biases, and reliable, as any flaw can directly influence ML predictions. Given the lack of criteria to help ensure this, we present two requirements for those exploring using ML to follow. The requirements we presented work to uphold international data quality standards put forth for ML. We then identify where our requirements can be met, as countries have varying contact tracing apps and smartphone usages. Lastly, the advantages, limitations, and ethical considerations of our approach are discussed.

Keywords: digital health, mobile applications, COVID-19, contact tracing, AI

OPEN ACCESS

Edited by:
Richard Ribon Fletcher, Massachusetts Institute of Technology, United States

Reviewed by:
Hemant Ghayvat, Linnaeus University, Sweden
Parisis Gallos, National and Kapodistrian University of Athens, Greece

***Correspondence:**
Rashmi Gupta
rash_cogsci@yahoo.com

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

Received: *31 July 2020*
Accepted: *23 November 2021*
Published: *15 December 2021*

Citation:
Dave R and Gupta R (2021) Data Quality and Network Considerations for Mobile Contact Tracing and Health Monitoring. Front. Digit. Health 3:590194. doi: 10.3389/fdgth.2021.590194

INTRODUCTION

Contact tracing involves identifying infected individuals and those they were in contact with to halt virus transmission (1–3). Contact tracing traditionally involved paper-based methods and have been able to help combat outbreaks such as SARS in 2003 (4, 5), Ebola in Africa in 2014 (6), and smallpox (7). Given the prevalence of smartphones, countries worldwide have rushed to develop contact tracing apps to streamline and enhance the tracking process. These apps use GPS capabilities via Bluetooth on smartphones to collect location data on individuals (3, 8, 9). A risk assessment is conducted if an individual's smartphone is close to an infected individual's smartphone for a long enough time. The individual receives a notification on the next steps they should follow (e.g., self-quarantine or getting tested). Compared to paper-based methods, digital methods reduce the time involved in contacting a set of close contacts. Moreover, a systematic review found that contact tracing apps were less prone to data loss, opening paths for deeper health monitoring (10). Given the promising benefits, various countries have developed proximity-sensing applications to automatically trace contacts, notify users about potential exposures, and invite them to isolate (11).

Prior Work in Digital Contact Tracing: Potential for Health Monitoring

Data collected from contact tracing apps can be evaluated in two primary ways. The first uses data analysis to perform a risk assessment that determines whether an individual should be notified of any exposure to the virus (3, 8, 9), thereby increasing the efficiency and accuracy of contact tracing

methods. As case examples, such methods have been employed to control Ebola in Sierra Leone, tuberculosis in Botswana, and whooping cough (pertussis) in the USA. In addition, models that examined digital contact tracing have replicated disease outbreaks in schools and found that the digital system successfully identified participants' close contacts (10).

The second use of data collected from contact tracing apps involves deeper layers of health monitoring. Governments have taken the aggregated data collected from all users to observe and predict trends. By analyzing location and health data, this type of analysis can monitor the spread of the virus, predict "hotspots," and forecast how resources should be distributed (5). From Ebola to the Zika virus to influenza, predictive analysis has enabled better resource allocation and public health planning (5, 6).

When the COVID-19 pandemic began, this effort was continued. For instance, models used location data in Wuhan to predict where the virus may arrive next in China. Other models strengthened their predictions by combining location data with other datasets, such as social media and credit card transactions, to successfully pinpoint vulnerable groups and predict hospital capacities (6, 7, 12).

Data and Network Consideration

There are unquestionable benefits to applying data analysis for public health planning. However, it is well-known that any model is only as good as the data provided. Biases and inaccuracies, amongst other traits, can easily plague datasets and lead to faulty results. With predictive analysis, this is an ever-greater concern, as the data used for machine learning or model forecasting is not only an input but also a part of training the software itself (13). In other words, with predictive analysis, the overall system performance is assessed, at least primarily, on its dataset, leading any bias or inaccuracy to bring rise to large complications in how a country may distribute resources or declare "hotspots" (13, 14).

For this reason, international data quality standards specific to predictive analyses have been put forth. The standards guide predictive analysis by ensuring the dataset being used for predictive analysis is accurate, reliable, and comprehensive, amongst many other traits. Given the demand for predictive analysis, it is imperative that the datasets used to forecast health planning uphold such data measures. In this paper, we analyze how, in any part, health monitoring via contact tracing upholds data quality standards. We use the international standard of data quality measures to determine which measures are upheld when analyzing data from centralized servers (network topology consideration for data collection) for contact tracing apps. Even in decentralized servers all the decentralized data will be aggregated (data centralization in steps, to serve a large population with variable smartphone capabilities) to be consumed by ML in one way or the other. In addition to data quality standards, we also examine ethical considerations. Our objective is to break down and review data from centralized servers for health monitoring data quality and from an ethical standpoint.

However, apart from data and networks, many other factors might affect data collection and quality. The adoption of the contact tracing app is the most significant factor among them. The following are reasons why contact tracing app intervention failed in many countries, including the USA (15):

(a) Due to outdated laws about privacy, data collection, and intention to use data, contact tracing apps may not be deployed. (16).

(b) App developers may have faced difficulties devising an "acceptable to all solution" due to technological feasibility issues (e.g., type of network topology: centralized, decentralized, or hybrid for communication or to store the data; biased predictive algorithms; or the efficacy of communication channels: Wifi, Bluetooth, Ultrasound, etc.).

(c) People may not have wanted to download/use the application as intended due to distrust in the agency/government that was collecting the data, ethical concerns (misuse of the data, expiry of data), privacy concerns (e.g., surveillance), and over cybersecurity issues (e.g., hacking).

METHODS

Evaluation of Network

Since ML models will be consuming the data generated by contact tracing apps, the centralized network topology is being considered. ML models both consume and generate a lot of data. With the emergency use of contact tracing apps during a pandemic and the current technology available, better prediction can only be achieved through centralized servers.

Evaluation of Data Quality

International standards for data quality were surveyed to determine which quality dimensions to use. Articles put forth for quality dimensions specific to ML were analyzed. A list of potential quality dimensions was developed, and ultimately, the data quality model set forth by Rudraraju and Boyanapally (14) was predominantly used. The model was based on a widely used international data quality model (ISO/IEC 25012) and adapted to the specific needs of ML (14–17). Further details of this data quality model are delved into in the next section.

Criteria for Apps

To determine our criteria for ML application, research papers from PubMed, ScienceDirect, and NCBI were sought out using a combination of terms: contact-tracing, mobile applications, COVID-19, AI, ML, trend-analysis, servers, etc. Reference lists of papers and existing literature reviews were also referred to. Studies published in English on contact tracing apps or applying AI/ML to aggregated data were included (18). The criteria were developed after understanding how to best envelope the data quality dimensions put forth by Rudraraju and Boyanapally (14).

Global Adoption of Criteria

Before smartphones, other types of data (e.g., online news aggregators, expert-curated discussions, and official reports) were also used for epidemiology and predicting the spread of pandemics. For example, a website (HealthMap.org), operated by a team of researchers, epidemiologists, and software developers at Boston Children's Hospital, brings a unified and comprehensive view of the current global state of infectious diseases. This website

uses a continuous automated process (e.g., monitors, organizes, integrates, filters, visualizes, and disseminates) and validated online information data sources (e.g., online news aggregators, expert-curated discussions, and official reports) to predict the current global state of infectious diseases (19).

To determine where our ML/AI application criteria can be met, we reviewed the smartphone penetration rate (percentage of the population actively using smartphones) and the contact tracing app's server type (centralized or decentralized) per country. ITU estimated that at the end of 2019, slightly more than 51 per cent of the global population, or 4 billion people, will be using the Internet; actual results are very close to the predicted one (20). Smartphone penetration rates were taken from Newzoo's Global Mobile Market Report, last updated in September of 2019 (21). The report lists the top 20 countries with the most active smartphone users along with corresponding smartphone penetration rates, which were taken for this review. An active smartphone user is qualified as an individual that uses the device at least once a month. The percent of smartphone users needed to use the contact tracing app to get to the 56% adoption rate (8, 30) was then calculated using the country's population of smartphone users (see **Table 2**). To determine server type, the main contact tracing app put forth by the government of that country was assessed through systematic searches, as some countries have second-party apps (22–27).

Data Quality Dimensions for ML

As noted, to significantly forecast virus spread, the data used to train the ML model must be of quality, that is free of biases, inaccuracies, and inefficiencies, amongst many other traits, before it is analyzed (14). Noting the importance of quality data, international models have been put forth as standards for data scientists to follow. For this paper, we define quality data as data that upholds international and ML-based data standards.

The ISO (International Organization for Standardization) and the IEC (International Electrotechnical Commission) have put forth standardized dataset specifications in their Data Quality Model (ISO/IEC 25012) (17). This model has been widely deployed as a standardized data guideline and is the base of our quality dimension list (14). However, because this model was developed in the context of statistical studies where data is input instead of an architectural component, we also consider quality dimensions put forth for ML specifically (14, 28, 29).

While there are no standardized data quality models for ML, various authors recommended some models (13, 14). For our analysis, we decided to envelop the data quality dimensions for ML set forth by Rudraraju and Boyanapally (14). Their list integrates dimensions from the ISO/IEC 25012 model, interviews with a range of data scientists, and a thorough literature review of data quality attributes. Namely, the quality dimensions we aim to uphold are: Accuracy, Completeness, Credibility, Currentness, Efficiency, Traceability, Understandability, Availability, Reproducibility, Relevancy, Interpretability, Effectiveness, and Satisfaction. Definitions for each were taken from the work of Rudraraju and Boyanapally (14) and are listed in **Table 1**. For a deeper understanding of how these dimensions came to be, we direct the reader to the work of Rudraraju and Boyanapally (14).

TABLE 1 | Data quality dimensions.

Data quality dimension	Description
Understandability	This attribute enables the users to interpret and express the information in appropriate languages and symbols for a specific context of use.
Fairness	The machine is trained with data with the ratio of all races (e.g., Black, white, etc.).
Currentness	This attribute identifies the information that is up to date.
Efficiency	Capability of providing suitable performance according to the number of resources used.
Availability	The degree to which the extracted data can be retrieved by authorized users for that context of use.
Relevance	To retrieve the data based on the requirement of the end-user or targeted customers.
Context Coverage	The level to which the system can be re-trained with the data that matches the end user's requirements.
Reproducibility	The degree to which the data can reproduce the same results and allow others to continue to train new machine learning systems.
Traceability	The extent to which the source of information, including owner and/or author of the information, and any changes made to the information can be verified.
Satisfaction	The extent to which the end-user is satisfied with the trained data.
Effectiveness	The capability to produce the desired output from the extracted data.
Completeness	The ability of data to represent every meaningful state of the represented real-world system.
Accuracy	Data is accurate when data values stored in the database correspond to real-world values or the extent to which data is correct, reliable, and certified.
Interpretability	To extract the data with the right language, units, and symbols with better understandability.
Credibility	The extent to which the information is reputable, objective, and trustable.
Size	Depending on the type of input data, the maximum amount of data that varies is the size of the data.

This table shows the dimensions taken from Rudraraju and Boyanapally (14) that we used in this paper.

ANALYSIS

Ensure That the Contact Tracing App Can Achieve at Least a 56% User Adoption Rate

A group at the University of Oxford came up with an epidemiological simulation model to demonstrate the importance of contact tracing app intervention, indicating that delaying contact tracing by 1 day after the onset of symptoms could affect epidemic control and the resurgence of coronavirus. The model's assumptions and estimation of the key matrices (e.g., vaccination, lockdown, quarantine, other interventions, etc.) were derived from transmission dynamics analysis of early coronavirus outbreaks in China. The group

TABLE 2 | Applying ML to data from Contact-Tracing Apps.

Country	Smartphone penetration rate (%)	Decentralized or centralized?	Meets our standards to apply AI?	Percent of smartphone users needed to meet threshold of 56% adoption rate (%)
United Kingdom	82.9	Decentralized	No	67.6
Germany	79.9	Decentralized	No	70.0
United States	79.1	Decentralized*	No	70.8
France	77.5	Centralized	Yes	72.3
Spain	74.3	Decentralized	No	75.3
South Korea	70.4	Centralized	Yes	79.6
Russia	66.3	Centralized	Yes	84.5
Italy	60.8	Decentralized	No	92.1
China	59.9	Centralized	Yes	93.4
Japan	57.2	Decentralized	No	97.9
Iran	54.8	Centralized	No	102.1
Turkey	54.0	Centralized	No	103.8
Mexico	49.5	——–**	No	112.9
Brazil	45.6	Decentralized	No	122.7
Vietnam	44.9	Decentralized	No	124.8
Philippines	33.6	Decentralized	No	166.8
Indonesia	31.1	Centralized	No	179.9
India	36.7	Centralized	No	152.6
Bangladesh	18.5	Centralized	No	303.7
Pakistan	15.9	Centralized	No	352.5

*The smartphone penetration rate per country and its server type to store contact tracing data. Countries listed above the red line have a smartphone penetration rate of at least 56%. If a country has a centralized server, AI can be feasibly applied to the data (denoted by "yes" and green box). The percentage of smartphone users required to achieve a 56% adoption rate is also listed. *In the United States, it should be noted that while certain states have begun to design official contract tracing apps, there is not a national consensus. **Information on Mexico could not be retrieved.*

measured successful outbreak control as a reduction in daily virus incidence, daily hospitalizations, number of people in or admitted to the hospital and ICU each day, daily deaths, number of people in quarantine each day, and number of tests required each day. This openly available model allows governments to compare and evaluate different contact tracing strategies alongside other real-time interventions (8). Any country may use this simulation model to derive/validate/estimate key matrices and use them for predictions.

Preliminary analysis of the UK National Health Service (NHS) Test and Trace programme at the Isle of Wight, by the same group, showed that contact tracing app intervention has a more significant impact on epidemic control. They concluded that there were significant decreases in incidence and R (basic reproduction number) (30). The group established the 56% adoption rate metric after investigating the effectiveness of contact tracing apps (8). This metric soon became the most cited adoption rate across literature, with the World Health Organization later stating that the adoption rate needed to be 60–70% (4, 9). The authors concluded that combining digital contact tracing with other interventions, such as community testing and continued shielding of vulnerable individuals, can help prevent coronavirus from rapidly re-emerging (8).

To further substantiate this metric, models examining contact tracing apps have shown that an adoption rate lower than 56% does not best represent a region's population, leading to virus resurgence and further lockdowns (8). As a case example, in a contact tracing app study conducted in the Isle of Man, while a 38% app adoption rate did improve aspects of the outbreak, authors noted that it did not effectively shut down virus spread. Although a 56% app adoption rate is far from complete usage by a population, it is clear that this metric brings about a sufficient and broad understanding of virus spread in the population that can be extrapolated from, mitigating the many possible data biases and inaccuracies that can arise.

A significant response to data biases and inaccuracies is necessary when working with data from contact tracing apps. Data riddled with biases can no longer be deemed fair. Further, it can no longer be considered a reliable, relevant, and complete representation of a population, which would generate long-term impacts on the data's effectiveness in stopping virus spread. For example, if ML was applied to understand how to best distribute resources to an incomplete dataset that did not represent the entire population, there would be an imperfect determination of "hotspots" (31). As a result, resource allocation of materials, such as testing kits and personal protective equipment, would be skewed (4, 32). Specifically, disproportionately more resources may be given to wealthier demographics that presumably have better access to smartphones than lower-income levels (33). It is thus imperative to have a majority adoption rate and possibly close to complete cooperation across ages and socioeconomic zones, as ML can not only make inaccurate predictions but also

perpetuate stereotypes, deepening biases across gender, income, and race (30, 33). While an equitable distribution of smartphones may be logistically difficult to achieve, at a minimum, to work toward a credible dataset for ML to effectively act upon, we believe that at least 56% of a region's population must be using the contact tracing app. By achieving an app adoption rate of at least 56%, we believe that the following quality dimensions can be upheld: data completeness, relevancy, credibility, fairness, and effectiveness.

Ensure the Contact Tracing App Has a Centralized Server to Store the Data

There are two dominating server types employed for contact tracing apps: centralized and decentralized systems. Each server has unique qualities and a differentiated ML approach. Before discussing why we believe a centralized system is better suited for upholding quality dimensions for data collected from contact-tracing apps, we will overview centralized and decentralized systems.

Centralized and Decentralized Servers

Amongst contact-tracing apps, there is global differentiation in what a centralized or decentralized server is (22). In a centralized server, data from the user is placed into a central source. In a decentralized system, location and contact data are stored locally on the user's smartphone (7). The process for either approach begins in the same way. When two smartphones come in contact with one another (Bluetooth range), a pseudonym code (anonymous identifier) is sent via Bluetooth to mark that interaction (4, 9). When individuals get tested for the virus, they can choose whether they want to upload the list of unique, anonymous identifiers to a common database in a decentralized model. This is the only information that the database receives, and with it, other phones can compare their unique identifiers with those of infected individuals to see if there is a possibility of exposure (34). However, in a centralized approach, the anonymous identifiers are uploaded along with proximity/interaction data (21, 34).

Deploying ML Application to Centralized Servers

As a result of the differentiated server types, ML is applied distinctly to each type. In centralized ML, the model is applied directly to the aggregated server, which holds the data. In decentralized ML, a model is sent to each smartphone. With the data on that phone, the model is then trained, and only model updates are sent to a central server, preventing any individual data from entering a single port. That updated model is then sent out to the next sequence of phones for training. In both, server communication—whether of data or model updates—is done through Bluetooth, Wi-Fi, or cellular networks. Further, in both, the trained model with the input data is then used to forecast virus spread (35).

Centralized Servers for Contact-Tracing Apps

Our analysis concludes that a centralized server is better suited to uphold data quality dimensions in contact tracing apps. Specifically, we believe that the following dimensions can be upheld: accuracy, efficiency, currentness, size, completeness, traceability, availability, interoperability, understandability, and integrability.

We begin with data completeness, currentness, and size. Contact tracing data must be up to date; any delay in gathering accurate health or location data directly impacts what we know about infected individuals in a community, leading to faulty ML predictions. ML would be applied to location and health data locally in a decentralized system—on each individual's phone. While this offers security benefits, as a centralized system would store the same location and health data in a single server for ML, there is a significant risk that location and health data from some users will not be accounted for (36–38). Suppose a user shuts off their phone, loses battery, or loses network connectivity during the round their phone has been selected for ML. In that case, the model will not take in that user's location and health data for its predictions until the next cycle.

Noting this concern, leading scientists have expressed that those who employ decentralized ML must be open to only a small subset of devices that may be active at each training round (36). While the same connectivity issues could exist with a centralized server (e.g., the phone could be powered down or out of network), a centralized system does not rely on a single point of data exchange for ML training. Updates are sent more continuously, allowing location and health data to be transferred once the phone returns to connectivity.

An additional issue for decentralized architectures is that delays would impact the currentness of the dataset, and any missing information would directly impact the size and completeness of the dataset. This concern is deepened through system heterogeneity–the term used to describe the fact that today's phones come from various manufacturers and that not all ML models can operate similarly on each phone, opening possibilities for data exclusion [for a review of system heterogeneity in health systems, refer to (37)]. We believe that a central port for data collection would help ensure that the location and health data used are up to date and inclusive of nearly all the available users when the ML model is applied.

Next, the dataset must be interoperable, understandable, and integratable. This is especially critical in forecasting virus spread; so as to strengthen predictions of clusters of cases or vulnerable groups, many ML models have combined contact tracing data with other datasets, such as credit card transactions or social media. While a large, aggregated server is more susceptible to complications from a crash, creating an integrated port for data collection makes it easier to combine non-homogeneous data. To elaborate on this point, not all local phones may have the capability to combine outside data readily and feasibly, and doing so would put in question data privacy advantages in a decentralized system. A central port allows data to be cleaned then analyzed together, ridding the architectural restraints of configuring a system locally on each phone.

A central port also helps ensure that the data is traceable and readily available. When contact tracing, it is integral to have a method to trace back to the source of an infection or outbreak. Doing so makes it possible to build upon what contact tracing apps do with other methods—contacting family members in that

household or determining new biomarkers. Through a single aggregated server, it is possible to have a way to trace back data to the point of an issue or a specific user anonymously. This directly contrasts to localized information storage, a strategy that promotes privacy but makes it difficult to locate a datapoint anonymously (36–38).

Lastly, variations in data points between individual phones and communication bottlenecks have led us to deem centralized servers as more efficient for contact tracing apps. With efficiency and currentness, the ML model can be quickly updated to build the best predictions. In decentralized ML, rounds of updates must be sent back to the central model, leading communication efficiency to be a known bottleneck. In addition, variations in data points collected tamper streamlining the ML process as a whole, as the data may not be homogeneously sent or "understandable." Data satisfaction is essential for the app's end-user—the general people—and ensuring that data is communicated well is imperative.

Due to challenges with expensive communication, system heterogeneity, and interoperability, amongst other areas, we believe that a centralized server is better suited for ML application to data collected from contact tracing apps. While a decentralized system offers advantages, mainly when privacy and data currentness are not of central concern, our focus is on maintaining international standards for applying ML to contact tracing datasets. A centralized server will help us do exactly that, consequently better ensuring that a contact tracing model can continuously produce accurate, reliable, and consistent predictions to combat virus spread.

WHERE IS IT POSSIBLE TO MEET THE REQUIREMENTS?

Given the large variability in how countries are developing their apps, it is important to note that not all countries will be using a centralized server. In addition, given the large variability in smartphone penetration rates, not all countries will achieve an app adoption rate of 56%. To reach an adoption rate of 56%, at least 56% of a country's population must have access to a smartphone and use the app through a forced mandate or voluntarily. The data in **Table 2** highlights the global differentiation in server type as well as smartphone usage. Of the countries listed, we see that France, South Korea, Russia, and China meet both the requirements (highlighted in green, **Table 2**). All countries above the red line have a smartphone penetration rate of at least 56%. The United States is marked with an asterisk because while certain states have begun to deploy apps, there is no national consensus on whether the apps should be implemented (24).

LIMITATIONS

Limitations of the Geographical Findings

The Newzoo report only included 20 countries: those with the most active smartphone users. Thus, countries with smaller populations and high smartphone penetration rates were excluded.

First, it is essential to consider how those countries not listed are implementing contact tracing apps to gain a complete global understanding. For example, in Australia, the smartphone penetration rate is estimated to be 80%, and leaders have leaned toward a centralized server for their contact tracing app (39, 40). Secondly, while widely deployed, the use of contact tracing is still a developing effort, as decisions by leaders are amenable to current updates on data privacy and app efficacy. Just as the United Kingdom shifted from a centralized to decentralized approach, Germany and Austria ultimately decided to go with a decentralized approach (22). Lastly, it has been shown that COVID-19 tracking systems do not capture data on immigrants and other marginalized populations (33). Other groups without smartphones, such as the elderly and children under the age of 10, are also not accounted for in any analysis by AI on contact tracing data (8).

Limitations of the Approach

Despite the many advantages, we acknowledge that using a centralized server and ensuring a high user adoption rate is no panacea to ensuring quality data for virus forecasting. The accuracy of the dataset itself is dependent on a myriad of other factors, such as widespread testing, guaranteed app adoption rate, and the app's efficacy (4, 9). In cases where there is not widespread and efficient testing, individuals will struggle to get tested despite the app's recommendations, making it difficult to identify infected individuals and break the chain of transmission (4, 9, 34). In addition, even if there is a smartphone penetration rate of at least 56% in the country, the app's efficacy will only go as far as the number of individuals that agree to download it.

Lastly, despite the proposed benefits of the app itself, there is uncertainty around its true efficacy. The concern is two-fold: (1) There is little, if any, risk assessment or validation done on these apps before launch and (2) There are design limitations (4, 9, 31). For the first point, due to the urgency of the pandemic, countries have rolled out contact tracing apps without a proper assessment of the accuracy and success of the product (4). Will the app accurately track individuals, and should ML be applied if it does not? The second concern surrounds the design limitations of smartphones, and thus, the app itself. GPS/Bluetooth capabilities cannot account for situations in which individuals hold the same geolocation but are spatially distanced (31). For example, two individuals could be separated by a wall or on different floors of a building. Further consideration must be given to the management of false positives, as this would directly impact quality standards, such as data completeness and credibility.

ETHICAL CONSIDERATIONS

In applying AI to data collected from a centralized server, ethical considerations must be addressed (23, 25–27, 39, 41). As mentioned, a decentralized server offers greater privacy because the data is processed locally: on the user's phone. Therefore, a centralized server must encrypt the data and have high-security protocols due to high susceptibility to data breaches (4, 41).

Without security protocols, an individual's right to privacy is violated. As case examples, South Korea and Qatar have recently scrutinized security loopholes found in their apps and the implications on individuals (42). While it has also been shown that decentralized models can also be susceptible to similar data breaches, more research needs to be conducted to gain insight into this field. Furthermore, while the app data is typically anonymized, it has been shown that machine learning can re-identify data, leading to ethical concerns over privacy rights (42, 43).

Next, there is no clear definition of when data should be deleted from the central server. The World Health Organization suggests that data be deleted after the pandemic has ended locally (9). Given the large uncertainty of when that could be, questions surrounding public surveillance and the duration and ease of that surveillance arise (4). How long could governments track individuals, and can individuals ever ask that their data be taken off the server? Is it justifiable to continue to use ML to analyze the data even after the pandemic subsides? What if there is mission creep–analyzing the data outside the defined scope?

Lastly, to achieve greater adoption, countries such as Qatar have decided to mandate the use of the contact tracing app (44). Ethical analyses are necessary to understand whether it is justified to mandate the use of an app despite violations of individual rights. While a mandate would ensure that a majority uses the app, certain individual rights, such as privacy and liberty, would be infringed upon in the process (26).

CONCLUSION

We have presented a proposed method for using ML to analyze data from contact tracing apps consistent with data quality standards. In addition, we have identified the countries in which our methods are most feasible, later discussing ethical considerations, advantages, and limitations of this approach. As the pandemic rages on, it is ever more critical that ML models analyze quality contact tracing data. We hope to shed light on the need for a methodological approach, inspiring further research into this field.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://resources.newzoo.com/hubfs/Reports/2019_Free_Global_Mobile_Market_Report.pdf?utm_campaign=Mobile%20Report%20Launch%202019&utm_medium=email&_hsmi = 76926953&_hsenc=p2ANqtz-9srHwQONqYM2EChvObOaegqly7JvX3KNYKSHWJKFR2j-iOvj6TW4uDROWXR-OJw8oOw_7j9KHakgEYu_NATW8-Q3TW0Q&utm_content=76926953&utm_source=hs_automation.

AUTHOR CONTRIBUTIONS

All authors have written the paper.

REFERENCES

1. Leslie D. *Tackling COVID-19 through Responsible AI Innovation: Five Steps in the Right Direction*. Harvard Data Science Review (2020). Retrieved from: https://hdsr.mitpress.mit.edu/pub/as1p81um
2. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Synd Clin Res Rev*. (2020) 14:337–9. doi: 10.1016/j.dsx.2020.04.012
3. COVID-19 Contact Tracing. *Centers for Disease Control and Prevention*. Available online at: https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/contact-tracing.html (accessed July 31, 2020).
4. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. *Lancet Digital Health*. (2020) 2:E425–34. doi: 10.1016/S2589-7500(20)30137-0
5. Stevens Institute of Technology. *A.I. Tool Provides More Accurate Flu Forecasts: Location Data to Provide Robust Longer-Term Insights on Flu Outbreaks*. (2020). Retrieved from: www.sciencedaily.com/releases/2020/11/201102142250.htm (accessed January 12, 2021).
6. Bansal A, Padappayil RP, Garg, C, Singal A, Gupta M, Klein A. Utility of artificial intelligence amidst the COVID 19 pandemic: a review. *J Med Syst*. (2020) 44:156. doi: 10.1007/s10916-020-01617-3
7. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. (2020) 395:689–97. doi: 10.1016/S0140-6736(20)30260-9
8. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M. et al. *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*. Technical Report, University of Oxford, Oxford, UK (2020).
9. *Ethical Considerations to Guide the Use of Digital Proximity Tracking Technologies for COVID-19 Contact Tracing*. World Health Organization. Available online at: https://www.who.int/publications/i/item/WHO-2019-nCoV-Ethics_Contact_tracing_apps2020.1 (accessed July 31, 2020).
10. Anglemyer A, Moore THM, Parker L, Chambers T, Grady A, Chiu K, et al. Digital contact tracing technologies in epidemics: a rapid review. *Cochrane Database Syst Rev*. (2020) 2020:CD013699. doi: 10.1002/14651858.CD013699
11. Moreno López JAM, ArreguiGarcía B, Bentkowski P, Bioglio L, Pinotti F, Boëlle PY, et al. Anatomy of digital contact tracing: role of age,transmission setting, adoption, case detection. *Sci Adv*. (2021) 7:eabd8750. doi: 10.1126/sciadv.abd8750
12. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *Lancet Digital Health*. (2020) 2:e166. doi: 10.1016/S2589-7500(20)30054-6
13. Picard S, Chapdelaine C, Cappi C, Gardes L, Jenn E, Lefevre B, et al. Ensuring dataset quality for machine learning certification. *2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. (2020). doi: 10.1109/ISSREW51248.2020.00085
14. Rudraraju N, Boyanapally V. *Data quality model for machine learning*. Unpublished master's thesis. Faculty of Computing, Blekinge Institute of Technology, Karlskrona, Sweden (2019).
15. Clark E, Chiao EY, Amirian ES. Why contact tracing efforts have failed to curb coronavirus disease 2019 (COVID-19) transmission in much of the United States. *Clin Infect Dis*. (2021) 72:e415–9. doi: 10.1093/cid/ciaa1155
16. Rich J. *How Our Outdated Privacy Laws Doomed Contact-Tracing Apps*. Brookings (2021). Available online at: https://www.nature.com/articles/d41586- 020-01514-2 (accessed July 31, 2020).
17. Rafique I, Lew P, Abbasi MQ, Li Z. Information quality evaluation framework: extending ISO 25012 data quality model. *Int Scholarly Sci Res Innovat*. (2012) 6:568–73. doi: 10.5281/zenodo.1072956
18. Borgesius FJZ. Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int J Hum Rights*. (2020) 24:10:1572–93. doi: 10.1080/13642987.2020.1743976

19. Health Map. Available online at: https://www.hsdl.org/?abstract&did=758367 (accessed September 16, 2021).
20. Statistics (2021). Available online at: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx (accessed September 16, 2021).
21. Newzoo Global Mobile Market Report - *Light Version*. Retrieved from: https://newzoo.com/insights/trend-reports/newzoo-global-mobile-market-report-2019-lightversion/ (accessed July 31, 2020).
22. Natasha L. *EU Privacy Experts Push a Decentralized Approach to COVID-19 Contacts Tracing*. TechCrunch (2020). Available online at: https://techcrunch.com/2020/04/06/eu-privacy-experts-push-a-decentralized-approach-to-covid19-contacts-tracing/ (accessed July 31, 2020).
23. O'Neill PH, Ryan-Mosley T, Johnson B. *A Flood of Coronavirus Apps Are Tracking Us. Now It's Time to Keep Track of Them*. MIT Technology Review (2020). Available online at: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ (Accessed July 31, 2020).
24. Richtel M. *Contact Tracing With Your Phone: It's Easier but There Are Tradeoffs*. The New York Times (2020). Available online at: https://www.nytimes.com/2020/06/03/health/coronavirus-contact-tracing-apps.html (accessed July 31, 2020).
25. Jennifer V, Singer N, Krolik A. *A Scramble for Virus Apps That Do No Harm*. The New York Times (2020). Available online at: https://www.nytimes.com/2020/04/29/business/coronavirus-cellphone-apps-contact-tracing (accessed July 31, 2020).
26. *A New World for Data Privacy. Mexico | Global Law Firm | Norton Rose Fulbright*. Available online at: https://www.nortonrosefulbright.com/enmx/knowledge/publications/d7a9a296/contact-tracing-apps-a-new-world-for-dataprivacy#Turkey (accessed July 31, 2020).
27. Zak D. *COVID-19's New Reality-These Smartphone Apps Track Infected People Nearby*. Forbes. (2020). Available online at: https://www.forbes.com/sites/zakdoffman/2020/04/07/covid-19s-new-normal-yes-your-phonewill-track-infected-people-nearby/#55f51f507f0d (accessed July 31, 2020).
28. Gudivada V, Apon A, Ding J. Data quality considerations for big data and machine learning: going beyond data cleaning and transformations. *Int J Adv Softw*. (2017) 1:1–20. Available online at: http://www.iariajournals.org/software/
29. Barocas S, Selbst A. *Data Quality and Artificial Intelligence - Mitigating Bias and Error to Protect Fundamental Rights*. (2016). Retrieved from: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf
30. Kendall M, Milsom L, Abeler-Dörner L, Wymant C, Ferretti L, Briers M, et al. Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: a preliminary analysis. *Lancet Digital Health*. (2020) 2:e658–66. doi: 10.1016/S2589-7500(20)30241-7
31. Robert R, Angela B, Alena B. *Ethics of Digital Contact Tracing Apps for the Covid-19 Pandemic Response*. Technical report, Kompetenznetz Public Health (2020).
32. Sera W, Mamas MA, Topol E, Spall HGCV. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digital Health*. (2020) 2:e435–8. doi: 10.1016/S2589-7500(20)30142-4
33. Smith G, Ishita, R, Associate Director at the Center for Equity. *The Problem With COVID-19 Artificial Intelligence Solutions and How to Fix Them (SSIR)*. Stanford Social Innovation Review: Informing and Inspiring Leaders of Social Change. Available online at: https://ssir.org/articles/entry/the_problem_with_covid_19_artificial_intelligence_solutions_and_how_to_fix_them# (accessed July 31, 2020).
34. Mark Z. *Coronavirus Contact-tracing Apps: Can They Slow the Spread of COVID-19?* Nature News (2020). Available online at: https://www.nature.com/articles/d41586-020-01514-2 (accessed July 31, 2020).
35. Vergne J. Decentralized vs. distributed organization: blockchain, machine learning and the future of the digital platform. *Organ Theory*. (2020) 1:263178772097705. doi: 10.1177/2631787720977052
36. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: challenges, methods, future directions. *IEEE Signal Process Mag*. (2020) 37:50–60. doi: 10.1109/MSP.2020.2975749
37. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S. Cardoso MJ. The future of digital health with federated learning. *NPJ Digital Med*. (2020) 3:1–7. doi: 10.1038/s41746-020-00323-1
38. *Federated Learning: Collaborative Machine Learning without Centralized Training Data*. Retrieved from: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html (accessed April 06, 2017).
39. Ariel B. *Three Months On, Does Australia Need a New Coronavirus Contact Tracing App?* ABC News (2020). Available online at: https://www.abc.net.au/news/science/2020-07-27/how-does-covidsafe-compare-contact-tracingapps-apple-google/12488188 (accessed July 31, 2020).
40. Granwal L. *Australia - Smartphone Penetration 2012-2022*. Statista (2020). Available online at: https://www.statista.com/statistics/321477/smartphoneuser-penetration-in-australia/ (accessed July 31, 2020).
41. Jessica M, Cowls J, Taddeo M, Floridi L. *Ethical Guidelines for COVID-19 Tracing Apps*. Nature News (2020). Available online at: https://www.nature.com/articles/d41586-020-01578-0 (accessed July 31, 2020)
42. Choe S, Krolik A, Zhong R, Singer N. *Major Security Flaws Found in South Korea Quarantine App*. The New York Times (2020). Available online at: https://www.nytimes.com/2020/07/21/technology/korea-coronavirus-app-security.html (accessed July 31, 2020).
43. Cohen J. *AI Can Re-identify De-identified Health Data, Study Finds*. Retrieved from: https://www.beckershospitalreview.com/artificial-intelligence/ai-can-re-identifyde-identified-health-data-study-finds.html (accessed July 31, 2020).
44. Dave R, Gupta R. Mandating the use of proximity tracking apps during Covid-19: ethical justifications. *Front Med*. (2020) 7:590265. doi: 10.3389/fmed.2020.590265

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 | Frontiers in Digital Health

REVIEW
published: 17 May 2022
doi: 10.3389/fdgth.2022.773387

Potentials and Challenges of Pervasive Sensing in the Intensive Care Unit

Anis Davoudi[1]*, *Benjamin Shickel*[2], *Patrick James Tighe*[3], *Azra Bihorac*[4] and *Parisa Rashidi*[1]

[1] Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States, [2] Department of Medicine, University of Florida, Gainesville, FL, United States, [3] Department of Anesthesiology, University of Florida, Gainesville, FL, United States, [4] Department of Medicine, University of Florida, Gainesville, FL, United States

Patients in critical care settings often require continuous and multifaceted monitoring. However, current clinical monitoring practices fail to capture important functional and behavioral indices such as mobility or agitation. Recent advances in non-invasive sensing technology, high throughput computing, and deep learning techniques are expected to transform the existing patient monitoring paradigm by enabling and streamlining granular and continuous monitoring of these crucial critical care measures. In this review, we highlight current approaches to pervasive sensing in critical care and identify limitations, future challenges, and opportunities in this emerging field.

Keywords: intensive care unit (ICU), wearable device, computer vision, pervasive sensing, patient monitoring

OPEN ACCESS

Edited by:
Panicos Kyriacou,
City University of London,
United Kingdom

Reviewed by:
Yan Wang,
Guangdong Technion-Israel Institute of Technology (GTIIT), China
Eftim Zdravevski,
Saints Cyril and Methodius University of Skopje, North Macedonia
M. Shamim Kaiser,
Jahangirnagar University, Bangladesh

***Correspondence:**
Anis Davoudi
anisdavoudi@ufl.edu

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

Received: *09 September 2021*
Accepted: *08 April 2022*
Published: *17 May 2022*

Citation:
Davoudi A, Shickel B, Tighe PJ, Bihorac A and Rashidi P (2022) Potentials and Challenges of Pervasive Sensing in the Intensive Care Unit. Front. Digit. Health 4:773387. doi: 10.3389/fdgth.2022.773387

CURRENT CRITICAL MONITORING PARADIGM

Critically ill patients in the intensive care unit (ICU) require constant monitoring. Currently, "continuous" monitoring of ICU patients is limited to automated vital sign measurements such as heart rate, blood pressure, body temperature, oxygen saturation, and respiratory rate. Other monitoring activities are limited by nurse availability for observing and documenting events, e.g., documenting falls or self-extubation events or detecting any exacerbation in important clinical indices such as mobility, agitation, pain, and consciousness. At present, assessment of these indices heavily relies on manual and repetitive examinations by nurses, leading to increased work pressure and the potential for burnout (1, 2). These manual indices also suffer from human error in data entry, observer subjectivity, and limited measurement granularity (3–5).

A more granular and continuous assessment of such critical care indices would enable a more comprehensive view of patient health. For example, granular functional status and behavioral assessment could lead to timely and personalized interventions based on data-driven guidelines. The need for continuous and automated monitoring of ICU patients has led researchers to incorporate non-traditional methodologies such as computer vision, wearable sensing technology, and various analytics algorithms (6). A comprehensive picture of the current state of the research in this domain will help outline the next steps and point out some of the questions that need to be answered. This work evaluates the feasibility of such monitoring approaches that are amenable to pervasive sensing in the ICU.

In the following sections, we detail current applications of pervasive sensing in ICU patient care settings. We then outline the knowledge gap in the literature, discuss current limitations, and highlight potential avenues for future research in augmenting traditional intensive patient care with pervasive sensing technology.

RECENT ADVANCES IN CRITICAL CARE MONITORING

Currently, monitoring ICU patients' functional status and behavioral aspects is limited in both granularity and information-richness. Pervasive sensing of the patient and their surrounding environment can provide a more comprehensive, continuous assessment of patient status. It can aid in quantifying patient health trajectories during the ICU stay. Two of the main avenues of research for pervasive sensing in the ICU involve using wearable accelerometers or computer vision devices, such as thermal or depth cameras, for monitoring the status of the patient and their environment. Wearable accelerometer devices, often resembling wristwatches, are lightweight, non-invasive, and easy to use. They allow for various computational analyses, do not pose any safety or comfort concerns for the patient, and do not impede the care procedures in the ICU. Wearable sensors have been previously used for quantifying human mobility and activity monitoring in various populations, and numerous analytic approaches showcase their potential uses for ICU patients (7–9). Applications of wearable accelerometers in the critical care setting are wide-ranging, including but not limited to physical activity (10–13) and energy expenditure monitoring (14), sleep detection (8, 15), agitation or sedation monitoring (16, 17), physiological signal monitoring (18, 19), fall detection (20, 21), delirium detection and subtyping (9, 22), sepsis subtyping (12), and frailty determination (23).

Another potential avenue for pervasive ICU monitoring involves the application of computer vision techniques. Computer vision provides a means for non-contact sensing of the patient and their surroundings in the ICU, providing rich information on several physical functions and behavioral aspects. Possible applications of computer vision in the critical care setting fall into two categories: (a) healthcare team observation, such as measuring nursing workload (24) or monitoring hand hygiene (25), and (b) patient monitoring. This article focuses on the latter—computer vision applications for assessing patients and their environment. Previous research has demonstrated the potential applications of computer vision for fall detection (26, 27), sleep pose detection (28), agitation detection (29), physical activity monitoring (30, 31), head pose detection (22), physiological signal monitoring (32), and visitation detection (22, 33) in hospital settings (**Figure 1**).

PHYSICAL FUNCTION MONITORING

ICU patients spend most of their time lying in bed, with significantly less time sitting in a chair, standing, or walking (22, 34). Patients' limited physical activity in the ICU has been linked with disruptions in circadian rhythm, a higher risk of delirium, and adverse outcomes in terms of cognitive and functional status at the time of hospital discharge and in the long term. Efforts at introducing physical therapy and early mobilization improve the patients' mobility and clinical outcomes such as delirium days, discharge disposition, and the risk of readmission or death (35–38). However, currently, there is a need for an objective, continuous, and accurate evaluation method to quantify the effect of rehabilitation practices. Additionally, such quantitative measures could be used to evaluate the association between a patient's activity levels and their outcomes during and after their ICU stay.

Existing clinical routine measurements of patients' mobility and physical status consist of limited standardized observational scores such as the ICU Mobility Scale (IMS) (39, 40). These observational evaluations aim to quantify patient mobility (40), but they still lack granularity and objectivity. Additionally, they provide limited information about mobility patterns' complex and dynamic nature throughout the ICU stay.

Computer vision and wearable accelerometer devices can provide more granular, objective, and continuous information on functional activity. Both approaches have been used to study physical activity in clinical settings to examine the association with health outcomes such as delirium or sepsis. However, there is limited research using computer vision or accelerometers in the critical care domain (**Table 1**), despite the high prevalence of sepsis or delirium in ICU patients. For example, delirium prevalence in specialized ICUs can be as high as 87%, and sepsis prevalence can be as high as 39% (46, 47).

Physical Activity

Recent efforts in the field of computer vision have applied deep learning techniques to build models for automated physical activity and posture recognition. Different camera types have been used to detect patient mobility, including Red-Green-Blue (RGB) cameras, depth cameras, and cameras that capture both color and depth images, such as the Microsoft Kinect device. Multi-view settings using multiple cameras installed at different positions also have been used for capturing a more encompassing view of the patient room (28, 31, 41, 42).

Generally, automated detection of patient mobility using computer vision first requires patient identification in the scene. After patient recognition, manually annotated datasets are used to train the model to classify patient pose and mobility. Such systems have been able to accurately classify patients' high-level activities such as *nothing in bed (doing nothing, lying in bed)*, *in-bed activity*, *out-of-bed activity*, and *walking* (30), and postures such as *lying in bed*, *sitting on the bed*, *sitting on the chair*, and *standing* (22) in the ICU. Depth camera-based systems have been able to classify four provider activities: "moving the patient into and out of bed" and "moving the patient into and out of a chair" without incorporating the challenging step of patient recognition (48, 49).

While computer vision techniques can identify a patient's posture, accelerometer devices can quantify the movement intensity. Wearable devices allow for convenient data collection and analysis since they provide continuous and patient-specific data streams. Wearable accelerometers have been used for examining physical activity patterns in different cohorts in various ICU settings, including delirium patients, sepsis patients, and patients with unilaterally motor impairment (10–13, 43, 50).

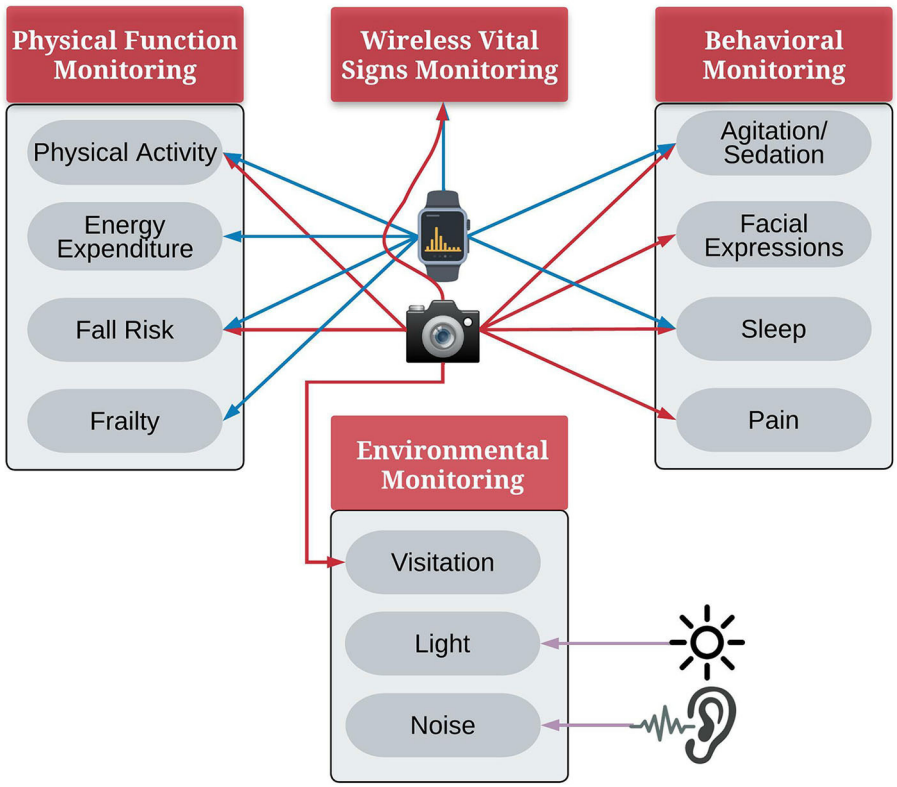


FIGURE 1 | Applications of pervasive sensing for monitoring ICU patients, iconfinder.com.

TABLE 1 | Peer-reviewed publications using wearable devices and computer vision in monitoring patients.

Task	Data modality	References
Physical activity	Vision Wearables	(10–12, 22, 28, 30, 31, 41–43)
Energy expenditure	Wearables	(14)
Fall risk	Vision Accelerometer	(20, 21, 26, 27, 44, 45)
Frailty	Accelerometer	(23)

Energy Expenditure

Building on activity intensity detection, wearable accelerometer devices have been used for determining energy expenditure. Previously energy expenditure estimation of accelerometer devices has been validated in the healthy adult population (14). However, it has been shown that energy expenditure is overestimated in ICU patients by comparing mechanical ventilation with indirect calorimetry. The current methodology of estimation of energy expenditure relies on the detection of physical activity. It does not incorporate physiological conditions such as fever with shivering that may alter the energy expenditure (51). Accurate estimation of energy expenditure in ICU patients enables optimizing enteral feeding details to prevent overfeeding and underfeeding, both of which increase the risk of infection and prolonged weaning from mechanical ventilation (52).

Frailty

Accelerometer devices have also been used to detect frailty (23), a geriatric syndrome defined as "a clinically recognizable state of increased vulnerability, resulting from the aging-associated decline in reserve and function across multiple physiologic systems such that the ability to cope with every day or acute stressors is compromised" (53). There is increasing evidence of frailty being an indicator for decreased reserve and increased vulnerability in critical care patients (54, 55). Frailty has been shown to increase the risk of both adverse events such as death and discharge to skilled nursing homes and prolonged hospitalization and loss of independence after hospital discharge (56, 57).

Falls

ICU patients experience decreased functional status and more muscle atrophy (58) exacerbated by minimal physical activity. They also suffer from impaired consciousness and attention (59), further compounded by disrupted circadian rhythm and sedation. While previous research has demonstrated the importance of physical activity in ICU patients, the increased risk of falls is brought on by confusion and agitation (60). Computer vision can detect falls in hospital settings (26, 27, 44, 45), as well

as accelerometer-based monitoring systems (20, 21). Still, these approaches have not been adequately investigated in the ICU.

Despite the importance of accurate assessment and monitoring of physical activity, energy expenditure, fall risk, and frailty in ICU patients, few studies have investigated the use of pervasive sensing to facilitate assessment and monitoring.

BEHAVIORAL MONITORING

Facial Expressions

Behavioral indices such as pain facial expressions are different from physical activity indices. They do not elucidate gross variations in patients and thus could elude the nurses' observation. Facial expressions can also be influenced by sedative-hypnotics and analgesics commonly used in the ICU. While there is no validated facial expressions score in critical care, a few preliminary studies have examined anxiety-related facial expressions. Anxiety is highly prevalent in critical care patients (61, 62). However, it is rarely screened in routine care settings in the ICU (62). Computer vision approaches have been used for anxiety detection (63, 64), focusing on features such as head movement, mouth, and eye movement, and heart activity as indicators of anxiety. Patient head pose angle and variability have also been studied using computer vision and associated with pain and agitation indicators (22, 65–67).

Pain

Continuous and objective monitoring of patient pain in the ICU, including for non-communicative patient populations, can pave the way for real-time adjustments to analgesics for optimal patient care, patient experience, and better health outcomes.

While wearable accelerometers have previously been used for studying pain (68), no study has investigated the relationship between pain and physical activity in the ICU settings, leaving unanswered the question of the complicated relationship between mobility, mental agitation, stress, and pain. The issue of pain in the ICU has many aspects. In addition to the potential effect of pain on a patient's physical activity, facial expressions and physiological signals may also be affected by pain. Previous work has investigated the feasibility of pain detection using vital signs (69, 70). Although this approach uses data routinely collected in the ICU, it has not shown strong specificity for pain detection. Formalizing facial expression of pain using facial action units (71) and advances in deep learning and computational power available have made it more plausible to move toward automated detection of facial expression of pain in the ICU. Facial expressions are assessed manually using several behavioral pain scales such as Non-verbal Pain Scale (NVP), Behavioral Pain Scale (BPS), and Critical Care Pain Observation Tool (CPOT), particularly for non-communicative patients (72–74). Researchers have used deep learning approaches to detect facial expressions of pain and to recognize individual facial action units associated with pain. Still, robust automated detection of pain in the ICU scene based on facial information requires more research and validation (75–79).

Agitation/Sedation

Agitation is prevalent in the ICU and is a large factor in conditions such as delirium (80). Current methods for assessing delirium rely on transient rather than continuous assessment, which is an important limitation given the waxing and waning characteristics which help define delirium. Over-sedation has been shown to lengthen ICU duration and put a patient at higher risk for delirium (81). In comparison, under-sedation has been linked with increased agitation and a higher risk of self-extubation (82, 83). Optimizing sedation to better control patients' agitation may lower the patient's risk of removing endotracheal tubes (84). Similar to pain, accurate detection of a patient's agitation and sedation levels will improve the administration of sedative interventions to optimize clinical decisions. Previous research has used sensors for monitoring agitation in critical care settings. Agitation detection methods have shown strong performance using accelerometers (16, 17, 85), image-based approaches (29), and pupillometric video devices (86). Using wearable devices to study anxiety, researchers used Google Glass to discover that heart rate, but not spontaneous blink rate, changes in anxious patients (87).

Sleep Detection

Previous studies have shown generally poor sleep quality in critical care settings (88). Sleep disruption in the ICU has been linked to various factors, including but not limited to the type and severity of the underlying medical condition, round the clock health care activities, enteral feeding, medication side effects, lack of natural light exposure and noise levels, and general disruptions to patients' circadian rhythm (88–90).

Sleep disturbance in ICU patients has been studied to determine its effect on patient outcomes and recovery (91) and has been shown to increase the risk of a longer stay in the ICU, worse discharge outcomes, impaired defense mechanism, and sleep disturbances that persist or develop after discharge (90). Determining a patient's sleep quality during their stay in the ICU allows for evaluating the effectiveness of the administered sleep hygiene interventions. Polysomnography, as the gold standard for studying sleep, has previously been investigated in ICU patients. However, polysomnography data require interpretation and might not be feasible for continuous data collection throughout the patient's stay in the ICU since it typically includes several EEG leads, electro-oculography, and chin electromyography (92, 93).

Accelerometer devices have been evaluated in quantifying sleep in healthy populations (15). However, previous studies have shown that they overestimate sleep in the ICU settings, possibly because the implemented sleep detection algorithms rely on a lack of physical activity in determining sleep events. ICU patients typically have low activity throughout the day, resulting in fractured sleep and shallower sleep stages (8). Other researchers have used computer vision to detect sleep pose, but these studies have been investigated based on data from healthy adults, limiting the generalizability of their performance to the ICU setting (28, 94).

ENVIRONMENTAL MONITORING

ICU patients spend most of their ICU stay in one room, leaving the room only for medical procedures. Information about the ICU room environment can enhance our understanding of possible contributing factors to patients' recovery speed. Wearable accelerometer devices are more suited for monitoring patients' physical activity, but they do not capture any information about patients' surroundings. Computer vision techniques offer additional opportunities for capturing and studying the effects of environmental factors on patients' recovery trajectories.

Visitation Detection

To encourage sleep hygiene**,** hospitals generally implement official guidelines for regulating the presence of visitors in the ICU. However, visitations and interactions with the environment may be beneficial in improving patients' experience by reducing their anxiety, leading to a lowered risk of delirium and an overall more positive experience during their ICU stay (95, 96). Accurate detection of the number of visitors and healthcare personnel in the room and environmental factors such as a room's noise and light at all hours allows for quantifying the effects of such disruptions on patients' sleep quality and circadian rhythm integrity. Computer vision has been used to determine the number of people in ICU care rooms (31, 33, 97) to understand the association between visitation and clinical care disruptions to patients with patients' sleep hygiene and outcomes. Such information could assist in developing more accurate evidence-based visitation and sleep quality guidelines for ICU patients.

Light and Noise Monitoring

Light and noise intensity levels are primary contributors to sleep deprivation and fragmentation in the ICU (91). Previous studies have demonstrated the feasibility of using affordable light and noise intensity sensors in ICU rooms (22, 98). Several studies using noise intensity sensors in the ICU have determined that noise level frequently surpasses the levels recommended by the World Health Organization (WHO) (99, 100). Various sleep hygiene interventions in the ICU population incorporate light and noise exposure limits with mixed results in efficacy. Accurate, continuous light and noise intensity measures can enhance our understanding of the effect of environmental factors on patient sleep and the efficacy of sleep hygiene guidelines (90).

NON-CONTACT VITAL SIGN MONITORING

In addition to novel physical activity and behavioral indices, the use of computer vision and wearable devices has enabled non-contact monitoring of vital signs such as heart rate and blood pressure in the ICU. Such devices could remove the need for electrodes and cuff-based devices sensitive to movement artifacts, prone to detachment, and restricting patient movement. Recent research in computer vision and wearable devices has focused on the feasibility of non-contact monitoring of vital signs, including the use of RGB and thermal cameras for contactless estimation of heart rate, blood pressure, blood oxygen saturation, and respiration rate in research, hospital and ICU settings [e.g., (32, 101–106)]. Wearable devices have been used for measuring heart rate, oxygen saturation, and blood pressure in hospital and ICU patients [e.g., (18, 19, 107)].

POST-DISCHARGE MONITORING

Critical care patients often require longitudinal monitoring and follow-up visits after discharge from the hospital. Patients' discharge destination can vary depending on the health status, ranging from home and home care for more stable patients to hospice for those who need more care with a lower chance for recovery. While clinicians have access to many tools to assess patients and determine a prognosis, current assessments do not extend well to post-discharge monitoring, resulting in less quantifiable information about post-discharge recovery (108). Computer vision approaches are not suitable in these scenarios due to technical and especially privacy concerns. However, wearable sensors can be used for physical activity monitoring and patient recovery (109), facilitating more comprehensive evaluations between follow-up visits. There is currently limited research on using this methodology to monitor the improvement in patients' functional status in free-living settings among survivors of critical illness (109–111). With the increasing popularity of accelerometer-equipped smartwatches, it is also possible to monitor physical activity before and after a patient's ICU stay to examine its relation with health recovery.

TECHNICAL CHALLENGES AND FUTURE DIRECTIONS

While the current body of literature shows strong potential for using these novel approaches in critical care for faster, more personalized, and more accurate clinical decisions, several challenges need to be addressed (**Figure 2**).

Ethical and Privacy Concerns

Deep learning methods discussed in many of the computer vision-based studies mentioned in this review require large, labeled datasets for training and validation. The developed models need to be validated in diverse populations in ICU settings and consider age, gender, primary diagnosis, and race. However, ethical and legal reasons rightly prevent the construction of public datasets to protect the privacy of patients, their visitors, and the healthcare team (112). Existing Privacy guidelines typically mandate deidentification and temporary recordings, while the current state of research still requires raw recordings of the ICU room for post-event annotations and analysis. While research in this direction could potentially facilitate the clinical workflow, researchers should be cognizant of the privacy implications and the tradeoff between privacy and technological benefits.

Computer Vision

An essential intermediate step in several computer vision approaches is detecting a patient in the room of each video frame. Previous research (113–115) has shown promising results in

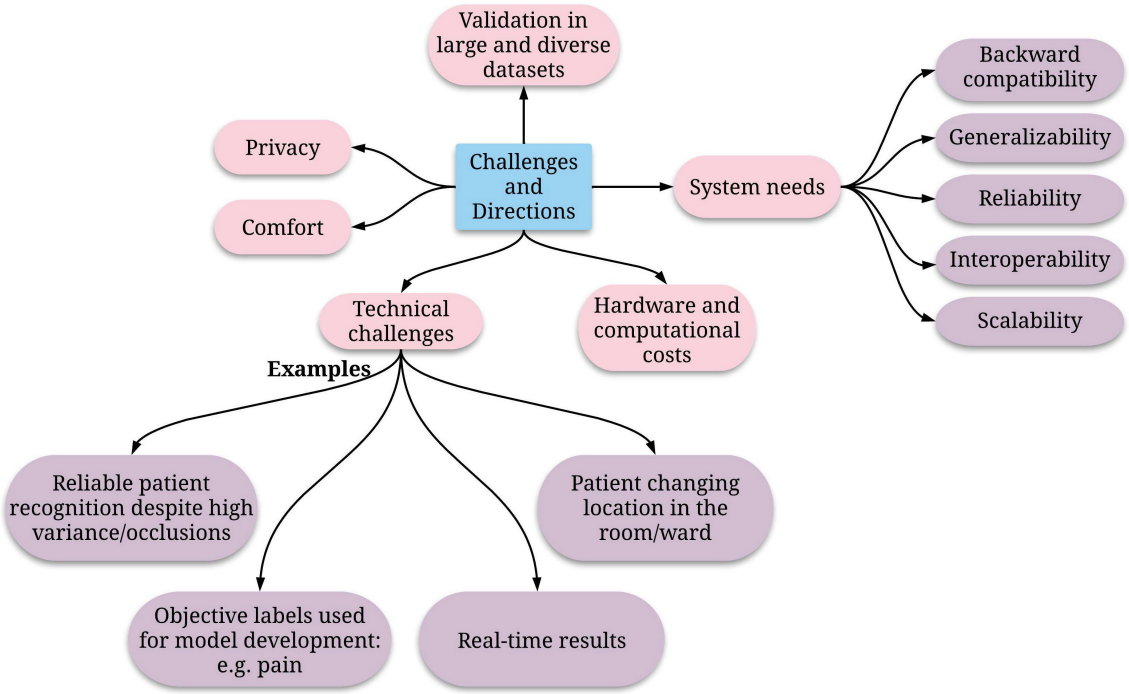


FIGURE 2 | Directions of future research to enable effective and reliable pervasive monitoring in the ICU.

facial recognition in each single frame or tracking the patient face during the recording, but most of these studies are validated on an ideal frontal full-face view of a patient's face or with standard lighting. Real-world ICU rooms may be crowded, with varying degrees of lighting and numerous objects that could partially obscure the patient's face, such as ventilators or oxygen masks. Patient face identification will be even more challenging due to variations in the face angle and obscuring elements such as facial hair or glasses. Improved patient recognition adapted to the ICU setting will make computer vision solutions more robust. Because of the variable location of the bed and patient, the ideal developed methodology should be agnostic to the bed/patient location and position.

Sleep Detection

Sleep detection methods still require further research for reliable use in ICU patients. Future approaches may focus on multimodal models and using wearable sensors to collect information on activity, heart rate, and body temperature. Current wearable activity sensors determine sleep based on a lack of activity and wake periods, but this does not account for the minimal activity of the ICU patients, thus resulting in low specificity.

Pain Assessment

Researchers also need to consider the effect of pain relief medications, nociceptive generators, and interindividual differences in pain processing in studying patients' pain. However, this is a challenging concept since medication effect changes over time depending on pharmacokinetics such as age, sex, weight, body surface area, renal and hepatic function, fluid shifts, medication dosage, time since administration, administration route, and drug-drug interactions. This complexity often increases for continuous infusions of medications. Additionally, model development relies on patient self-report and nurses' observation for non-verbal patients. The uncertainty regarding how to translate pain intensity assessments into objective, rational clinical decisions for analgesic therapies that reduce pain intensity and patient risk further complicates developing generalizable and reliable models. There is also a need to integrate outcomes pertaining to pain, analgesic requirement, and "patient function" to provide a more holistic perspective on recovery trajectories.

Wearable Devices

Another implementation challenge for using wearable devices in the ICU is the requirement of current devices to be tightly secured on the skin to better capture patients' subtle movements. This may prove to be an inconvenience for some patients, as continuous contact with the skin might cause irritation or even could pose risks for infection, tissue ischemia, compartment syndrome, and wound breakdown. Furthermore, for patients with medical equipment on their wrists, wrist-worn accelerometer devices might not be an option unless future medical equipment includes built-in accelerometers.

Forward-Compatibility

The inclusion of multiple data streams for different patient care tasks using pervasive sensing also requires *generalizability*, *interoperability*, *scalability*, and *reliability* of the systems. *Interoperable* and *generalizable* systems can operate together and be incorporated into different monitoring platforms, while *scalable* systems will accommodate collecting data from a larger number of patients. The accompanying analytical algorithms need to be *adaptable* for new hardware choices. Moreover, there must be a consensus on the measured variables better to evaluate the performance of proposed methods and devices. Ultimately, the positive impact of such systems in critical care needs to be rigorously assessed and validated.

Model Validation With Minimal Data Burden

To improve the adoption of pervasive sensing in routine care in the ICU, developed models and devices should be easily manageable by the care team with minimal or no required training. Any developed model should be validated to reduce the false alarm rate in the ICU -as is investigated with vital sign-based alarms (116)-and should optimize the visualization approach to prevent data fatigue. The presentation of new information should be determined by considering the preferences of the physicians and their team and could include facets such as a daily summary, continuous display or separate tabs, or simple alarms for specific, pre-determined events.

Real-Time Models

Ultimately, any developed detection and prediction model should report a patient's status in real-time to allow the ICU team to implement timely interventions, such as incorporating more active physical therapy regimens, improving sleep hygiene routines, and adapting administered medications. The necessary communications infrastructure and reporting medium should be optimized to avoid alarm fatigue to the already overburdened ICU nurses.

CONCLUSION

Patients in the ICU have diverse and heterogeneous health backgrounds, which necessitate more personalized and dynamic treatments and interventions. This calls for developing monitoring methodologies that provide continuous, objective, and quantifiable patient information. Traditional monitoring of vital signs, nursing observations, and self-reported pain scores is essential but does not provide a comprehensive view of the patient's overall health status.

Advances in computation and computer vision fields and the development of accurate measuring devices such as wearable accelerometers have introduced more options for patient monitoring in-home, community, and hospital settings. The acute nature of health events in the ICU can benefit from pervasive, passive sensing methodologies that reduce nurses' workload and replace some of the tasks that require repetition of measurements, such as detection of pain and agitation. Moreover, pervasive sensing technology can enable measuring indices that were not previously recorded, including a patient's physical activity level, facial expressions, and head pose variations. Classification algorithms trained on the data from similar scenarios may allow for more timely prediction of adverse events such as falls and delirium, enabling the healthcare team to prevent their occurrence.

Many of the proposed domains of the ABCDEF bundle,[1] an evidence-based guide for clinicians to improve ICU patients' recovery and outcomes that emphasize pain assessment, prevention, and management (117), require accurate monitoring of the patients during their stay in the ICU. Continuous monitoring of pain levels using computer vision approaches will be helpful for continuous and accurate pain assessment and ultimately for real-time adaptation of pain medications. The choice of analgesia and sedation is another domain in this guideline that can benefit from continuous monitoring of the patients' sedation levels to personalize analgesia choice. Quantifying patients' mobility also improves the evaluation of the effectiveness of the administered mobility and exercise regimens. Moreover, the use of accelerometer and vision sensors to detect delirium can improve delirium assessments, as proposed in the ABCDEF bundle.

AUTHOR CONTRIBUTIONS

AD and PR developed the concept for the study. AD drafted the manuscript and designed and created the figures. PR substantially revised the manuscript and revised the figures. BS, AB, and PT substantially revised the manuscript. All authors were approved the final manuscript.

FUNDING

PR was supported by 5R01AG055337 from National Institute of Aging (NIH/NIA), CAREER award 1750192 from the National Science Foundation, 1R01EB029699 and 1R21EB027344 from the National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB), R01GM-110240 from the National Institute of General Medical Science (NIH/NIGMS), 1R01NS120924 from the National Institute of Neurological Disorders and Stroke (NIH/NINDS), and R01 DK121730 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK). AB was supported R01 GM110240 from the National Institute of General Medical Sciences (NIH/NIGMS), R01 EB029699 and R21 EB027344 from the National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB), R01 NS120924 from the National Institute of Neurological Disorders and Stroke (NIH/NINDS), and by R01 DK121730 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK). PT was supported by 5R01GM114290 and 5R01AG055337.

[1] **A**ssess, Prevent, and Manage Pain, **B**oth Spontaneous Awakening Trials (SAT) and Spontaneous Breathing Trials (SBT), **C**hoice of analgesia and sedation, **D**elirium: Assess, Prevent, and Manage, Early mobility and **E**xercise, and **F**amily engagement and empowerment.

REFERENCES

1. Parry SM, Granger CL, Berney S, Jones J, Beach L, El-Ansary D, et al. Assessment of impairment and activity limitations in the critically ill: a systematic review of measurement instruments and their clinimetric properties. *Intensive Care Med.* (2015) 41:744–62. doi: 10.1007/s00134-015-3672-x
2. Thrush A, Rozek M, Dekerlegand JL. The clinical utility of the functional status score for the intensive care unit (FSS-ICU) at a long-term acute care hospital: a prospective cohort study. *Phys Ther.* (2012) 92:1536–45. doi: 10.2522/ptj.20110412
3. Drews FA. Patient monitors in critical care: Lessons for improvement. In: Kerm H, James BB, Margaret AK, Mary LG, editors. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 3: Performance and Tools)*. Rockville, MD: Agency for Healthcare Research and Quality (US).
4. Eltaybani S, Mohamed N, Abdelwareth M. Nature of nursing errors and their contributing factors in intensive care units. *Nurs Crit Care.* (2019) 24:47–54. doi: 10.1111/nicc.12350
5. Phillips ML, Kuruvilla V, Bailey M. Implementation of the Critical Care Pain Observation Tool increases the frequency of pain assessment for noncommunicative ICU patients. *Austr Crit Care.* (2019) 32:367–72. doi: 10.1016/j.aucc.2018.08.007
6. Lovejoy CA, Buch V, Maruthappu M. Artificial intelligence in the intensive care unit. *Crit Care.* (2019) 23:7. doi: 10.1186/s13054-018-2301-9
7. Schwab KE, To AQ, Chang J, Ronish B, Needham DM, Martin JL, et al. Actigraphy to measure physical activity in the intensive care unit: a systematic review. *J Intensive Care Med.* (2019) 8850666198636 54. doi: 10.1177/0885066619863654
8. Schwab KE, Ronish B, Needham DM, To AQ, Martin JL, Kamdar BB. Actigraphy to evaluate sleep in the intensive care unit. A systematic review. *Ann Am Thorac Soc.* (2018) 15:1075–82. doi: 10.1513/AnnalsATS.201801-004OC
9. Davoudi A, Manini TM, Bihorac A, Rashidi P. Role of wearable accelerometer devices in delirium studies: a systematic review. *Crit Care Expl.* (2019) 1:e0027. doi: 10.1097/CCE.0000000000000027
10. LaBuzetta J, Hermiz J, Gilja V, Karanjia N. Using accelerometers in the neurological ICU to monitor unilaterally motor impaired patients (P3.204). *Neurology*. (2016) 86:P3.204.
11. Winkelman C, Higgins PA, Chen Y-JK. Activity in the chronically critically ill. *Dimens Crit Care Nurs.* (2005) 24:281–90. doi: 10.1097/00003465-200511000-00011
12. Davoudi A, Corbett DB, Ozrazgat-Baslanti T, Bihorac A, Brakenridge SC, Manini TM, et al. *Activity and Circadian Rhythm of Sepsis Patients in the Intensive Care Unit*. Las Vegas, NV (2018). p. 17–20. doi: 10.1109/BHI.2018.8333359
13. Liu R, Fazio SA, Zhang H, Ramli AA, Liu X, Adams JY. Early mobility recognition for intensive care unit patients using accelerometers. *arXiv preprint arXiv:2106.15017*. (2021). doi: 10.48550/arXiv.2106.15017
14. Krüger J, Kraft M, Gründling M, Friesecke S, Gärtner S, Vogt LJ, et al. Evaluation of a non-invasive multisensor accelerometer for calculating energy expenditure in ventilated intensive care patients compared to indirect calorimetry and predictive equations. *J Clin Monit Comput.* (2017) 31:1009–17. doi: 10.1007/s10877-016-9934-5
15. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep.* (2003) 26:342–92. doi: 10.1093/sleep/26.3.342
16. Raj R, Ussavarungsi K, Nugent K. Accelerometer-based devices can be used to monitor sedation/agitation in the intensive care unit. *J Crit Care.* (2014) 29:748–52. doi: 10.1016/j.jcrc.2014.05.014
17. Grap MJ, Borchers CT, Munro CL, Elswick RK, Sessler CN. Actigraphy in the critically ill: correlation with activity, agitation, and sedation. *Am J Crit Care.* (2005) 14:52–60. doi: 10.4037/ajcc2005.14.1.52
18. Murali S, Rincon F, Cassina T, Cook S, Goy J-J. Heart rate and oxygen saturation monitoring with a new wearable wireless device in the intensive care unit: pilot comparison trial. *JMIR Biomed Eng.* (2020) 5:e18158. doi: 10.2196/18158
19. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. *J Med Internet Res.* (2016) 18:e253. doi: 10.2196/jmir.6025
20. Wu F, Zhao H, Zhao Y, Zhong H. Development of a wearable-sensor-based fall detection system. *Int J Telemed Appl.* (2015) 2015:576364–576364. doi: 10.1155/2015/576364
21. Sucerquia A, López JD, Vargas-Bonilla JF. Real-life/real-time elderly fall detection with a triaxial accelerometer. *Sensors.* (2018) 18:1101. doi: 10.3390/s18041101
22. Davoudi A, Malhotra KR, Shickel B, Siegel S, Williams S, Ruppert M, et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. *Sci Rep.* (2019) 9:8020. doi: 10.1038/s41598-019-44004-w
23. Lee H, Joseph B, Enriquez A, Najafi B. Toward using a smartwatch to monitor frailty in a hospital setting: using a single wrist-wearable sensor to assess frailty in bedbound inpatients. *Gerontology.* (2018) 64:389–400. doi: 10.1159/000484241
24. Adomat R, Hicks C. Measuring nursing workload in intensive care: an observational study using closed circuit video cameras. *J Adv Nurs.* (2003) 42:402–12. doi: 10.1046/j.1365-2648.2003.02632.x
25. Haque A, Guo M, Alahi A, Yeung S, Luo Z, Rege A, et al. Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. *arXiv preprint arXiv:1708.00163.* (2017). doi: 10.48550/arXiv.1708.00163
26. Krempel E, Birnstill P, Beyerer J. A privacy-aware fall detection system for hospitals and nursing facilities. *Eur J Secur Res.* (2017) 2:83–95. doi: 10.1007/s41125-017-0016-6
27. Dašić P, Dašić J, Crvenković B. Improving patient safety in hospitals through usage of cloud supported video surveillance. *Open Access Macedonian J Med Sci.* (2017) 5:101–6. doi: 10.3889/oamjms.2017.042
28. Torres C, Hammond SD, Fried JC, Manjunath B. Sleep pose recognition in an icu using multimodal data and environmental feedback. *Springer.* (2015) 56–66. doi: 10.1007/978-3-319-20904-3_6
29. Chase JG, Agogue F, Starfinger C, Lam Z, Shaw GM, Rudge AD, et al. Quantifying agitation in sedated ICU patients using digital imaging. *Comput Methods Programs Biomed.* (2004) 76:131–41. doi: 10.1016/j.cmpb.2004.03.005
30. Ma AJ, Rawat N, Reiter A, Shrock C, Zhan A, Stone A, et al. Measuring patient mobility in the ICU using a novel noninvasive sensor. *Crit Care Med.* (2017) 45:630. doi: 10.1097/CCM.0000000000002265
31. Torres C, Rose K, Fried JC, Manjunath B. Summarization of ICU patient motion from multimodal multiview videos. *arXiv preprint arXiv:1706.09430.* (2017). doi: 10.48550/arXiv.1706.09430
32. Rasche S, Trumpp A, Waldow T, Gaetjen F, Plötze K, Wedekind D, et al. Camera-based photoplethysmography in critical care patients. *Clin Hemorheol Microcirc.* (2016) 64:77–90. doi: 10.3233/CH-162048
33. Malhotra KR, Davoudi A, Siegel S, Bihorac A, Rashidi P. Autonomous detection of disruptions in the intensive care unit using deep mask R-CNN. *IEEE.* (2018) 1944–19442. doi: 10.1109/CVPRW.2018.00241
34. Connolly BA, Mortimore JL, Douiri A, Rose JW, Hart N, Berney SC. Low levels of physical activity during critical illness and weaning: the evidence–reality gap. *J Intensive Care Med.* (2017) 34:818–27. doi: 10.1177/0885066617716377
35. Hunter A, Johnson L, Coustasse A. Reduction of intensive care unit length of stay: the case of early mobilization. *Health Care Manag.* (2014) 33:128–35. doi: 10.1097/HCM.0000000000000006
36. Ota H, Kawai H, Sato M, Ito K, Fujishima S, Suzuki H. Effect of early mobilization on discharge disposition of mechanically ventilated patients. *J Phys Ther Sci.* (2015) 27:859–64. doi: 10.1589/jpts.27.859
37. Mendez-Tellez PA, Nusr R, Feldman D, Needham DM. Early physical rehabilitation in the ICU: a review for the neurohospitalist. *Neurohospitalist.* (2012) 2:96–105. doi: 10.1177/1941874412447631
38. Morris PE, Griffin L, Berry M, Thompson C, Hite RD, Winkelman C, et al. Receiving early mobility during an intensive care unit admission is a predictor of improved outcomes in acute respiratory failure. *Am J Med Sci.* (2011) 341:373–7. doi: 10.1097/MAJ.0b013e31820ab4f6
39. Hodgson C, Needham D, Haines K, Bailey M, Ward A, Harrold M, et al. Feasibility and inter-rater reliability of the ICU Mobility Scale. *Heart Lung J Acute Crit Care.* (2014) 43:19–24. doi: 10.1016/j.hrtlng.2013.11.003

40. Peterson ML, Lukens K, Fulk G. Psychometric properties of physical function measures used in the intensive care unit: a systematic review. *J Acute Care Phys Ther.* (2018) 9:78–90. doi: 10.1097/JAT.0000000000000073
41. Torres C, Fried JC, Rose K, Manjunath BS. A multiview multimodal system for monitoring patient sleep. *IEEE Trans Multimedia.* (2018) 20:3057–68. doi: 10.1109/TMM.2018.2829162
42. Liu B, Guo M, Chou E, Mehra R, Yeung S, Downing NL, et al. *3d Point Cloud-Based Visual Prediction of ICU Mobility Care Activities.* (2018). p. 17–29.
43. Fazio S, Doroy A, Da Marto N, Taylor S, Anderson N, Young HM, et al. Quantifying mobility in the ICU: comparison of electronic health record documentation and accelerometer-based sensors to clinician-annotated video. *Crit Care Expl.* (2020) 2:91. doi: 10.1097/CCE.0000000000000091
44. Mecocci A, Micheli F, Zoppetti C, Baghini A. *Automatic Falls Detection in Hospital-Room Context*. Wroclaw: IEEE (2016). doi: 10.1109/CogInfoCom.2016.7804537
45. Brown Kramer J, Sabalka L, Rush B, Jones K, Nolte T. *Automated Depth Video Monitoring for Fall Reduction: A Case Study*. Seattle, WA: IEEE (2020). p. 294–5. doi: 10.1109/CVPRW50498.2020.00155
46. Sakr Y, Jaschinski U, Wittebole X, Szakmany T, Lipman J, Ñamendys-Silva SA, et al. Sepsis in intensive care unit patients: worldwide data from the intensive care over nations audit. *Open For Infect Dis.* (2018) 5:ofy313. doi: 10.1093/ofid/ofy313
47. Cavallazzi R, Saad M, Marik PE. Delirium in the ICU: an overview. *Ann Intensive Care.* (2012) 2:49–49. doi: 10.1186/2110-5820-2-49
48. Chou E, Tan M, Zou C, Guo M, Haque A, Milstein A, et al. Privacy-preserving action recognition for smart hospitals using low-resolution depth images. *arXiv preprint arXiv:1811.09950*. (2018). doi: 10.48550/arXiv.1811.09950
49. Yeung S, Rinaldo F, Jopling J, Liu B, Mehra R, Downing NL, et al. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. *NPJ Digital Medicine.* (2019) 2:11. doi: 10.1038/s41746-019-0087-z
50. Verceles AC, Hager ER. Use of accelerometry to monitor physical activity in critically ill subjects: a systematic review. *Respir Care.* (2015) 60:1330–6. doi: 10.4187/respcare.03677
51. Moonen HPFX, Beckers KJH, van Zanten ARH. Energy expenditure and indirect calorimetry in critical illness and convalescence: current evidence and practical considerations. *J Intensive Care.* (2021) 9:8. doi: 10.1186/s40560-021-00524-0
52. Reid C. Frequency of under- and overfeeding in mechanically ventilated ICU patients: causes and possible consequences. *J Hum Nutr Diet.* (2006) 19:13–22. doi: 10.1111/j.1365-277X.2006.00661.x
53. Xue Q-L. The frailty syndrome: definition and natural history. *Clin Geriatr Med.* (2011) 27:1. doi: 10.1016/j.cger.2010.08.009
54. Ferrante LE, Pisani MA, Murphy TE, Gahbauer EA, Leo-Summers LS, Gill TM. The association of frailty with post-ICU disability, nursing home admission, and mortality: a longitudinal study. *Chest.* (2018) 153:1378–86. doi: 10.1016/j.chest.2018.03.007
55. Falvey JR, Ferrante LE. Frailty assessment in the ICU: translation to 'real-world' clinical practice. *Anaesthesia.* (2019) 74:700–3. doi: 10.1111/anae.14617
56. Makary MA, Segev DL, Pronovost PJ, Syin D, Bandeen-Roche K, Patel P, et al. Frailty as a predictor of surgical outcomes in older patients. *J Am Coll Surg.* (2010) 210:901–8. doi: 10.1016/j.jamcollsurg.2010.01.028
57. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci.* (2001) 56:M146–156. doi: 10.1093/gerona/56.3.M146
58. Jolley SE, Bunnell AE, Hough CL. ICU-acquired weakness. *Chest.* (2016) 150:1129–40. doi: 10.1016/j.chest.2016.03.045
59. Figueroa-Ramos MI, Arroyo-Novoa CM, Lee KA, Padilla G, Puntillo KA. Sleep and delirium in ICU patients: a review of mechanisms and manifestations. *Intensive Care Med.* (2009) 35:781–95. doi: 10.1007/s00134-009-1397-4
60. Richardson A, Carter R. Falls in critical care: a local review to identify incidence and risk. *Nurs Crit Care.* (2017) 22:270–5. doi: 10.1111/nicc.12151
61. Nelson JE, Meier DE, Oei EJ, Nierman DM, Senzel RS, Manfredi PL, et al. Self-reported symptom experience of critically ill cancer patients receiving intensive care. *Crit Care Med.* (2001) 29:277–82. doi: 10.1097/00003246-200102000-00010
62. Rincon HG, Granados M, Unutzer J, Gomez M, Duran R, Badiel M, et al. Prevalence, detection and treatment of anxiety, depression, and delirium in the adult critical care unit. *Psychosomatics.* (2001) 42:391–6. doi: 10.1176/appi.psy.42.5.391
63. Giannakakis G, Pediaditis M, Manousos D, Kazantzaki E, Chiarugi F, Simos PG, et al. Stress and anxiety detection using facial cues from videos. *Biomed Signal Process Control.* (2017) 31:89–101. doi: 10.1016/j.bspc.2016.06.020
64. Nagasawa T, Takahashi R, Koopipat C, Tsumura N. *Stress Estimation Using Multimodal Biosignal Information From RGB Facial Video*. Seattle, WA: IEEE (2020). p. 292–3. doi: 10.1109/CVPRW50498.2020.00154
65. Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue HC. *Towards Pain Monitoring: Facial Expression, Head Pose, a New Database, an Automatic System and Remaining Challenges*. Bristol: BMVA Press (2013). p. 1–13. doi: 10.5244/C.27.119
66. Werner P, Al-Hamadi A, Limbrecht-Ecklundt K, Walter S, Traue HC. Head movements and postures as pain behavior. *PLoS ONE.* (2018) 13:e0192767. doi: 10.1371/journal.pone.0192767
67. Amos B, Ludwiczuk B, Satyanarayanan M. *OpenFace: A General-Purpose Face Recognition Library With Mobile Applications*. Pittsburgh, PA: CMU School of Computer Science (2016).
68. Perraudin CGM, Illiano VP, Calvo F, O'Hare E, Donnelly SC, Mullan RH, et al. Observational study of a wearable sensor and smartphone application supporting unsupervised exercises to assess pain and stiffness. *Digit Biomark.* (2018) 2:106–25. doi: 10.1159/000493277
69. Arbour C, Choinière M, Topolovec-Vranic J, Loiselle CG, Gélinas C. Can fluctuations in vital signs be used for pain assessment in critically ill patients with a traumatic brain injury? *Pain Res Treat.* (2014) 2014:175794–175794. doi: 10.1155/2014/175794
70. Erden S, Demir Kalkan NA, Ugras G, Arslan U, Arslan S. *Vital Signs: Valid Indicators to Assess Pain in Intensive Care Unit Patients? An Observational, Descriptive Study*. New York, NY (2018). doi: 10.1111/nhs.12543
71. Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I. *Painful Data: The UNBC-McMaster Shoulder Pain Expression Archive Database*. Santa Barbara, CA: IEEE (2011). p. 57–64. doi: 10.1109/FG.2011.5771462
72. Arif-Rahu M, Grap MJ. Facial expression and pain in the critically ill non-communicative patient: state of science review. *Intens Crit Care Nurs.* (2010) 26:343–52. doi: 10.1016/j.iccn.2010.08.007
73. McGuire DB, Kaiser KS, Haisfield-Wolfe ME, Iyamu F. Pain assessment in noncommunicative adult palliative care patients. *Nurs Clin North Am.* (2016) 51:397–431. doi: 10.1016/j.cnur.2016.05.009
74. Odhner M, Wegman D, Freeland N, Steinmetz A, Ingersoll GL. Assessing pain control in nonverbal critically ill adults. *Dimension Crit Care Nurs.* (2003) 22:260–7. doi: 10.1097/00003465-200311000-00010
75. Rahu MA, Grap MJ, Cohn JF, Munro CL, Lyon DE, Sessler CN. Facial expression as an indicator of pain in critically ill intubated adults during endotracheal suctioning. *Am J Crit Care.* (2013) 22:412–22. doi: 10.4037/ajcc2013705
76. Lucey P, Cohn JF, Matthews I, Lucey S, Sridharan S, Howlett J, et al. Automatically detecting pain in video through facial action units. *IEEE Trans Syst Man Cybernet B.* (2011) 41:664–74. doi: 10.1109/TSMCB.2010.2082525
77. Chen Z, Ansari R, Wilkie D. Automated pain detection from facial expressions using FACS: a review. *arXiv e-prints.* (2018). doi: 10.48550/arXiv.1811.07988
78. Soar J, Bargshady G, Zhou X, Whittaker F. *Deep Learning Model for Detection of Pain Intensity from Facial Expression*. Cham: Springer International Publishing (2018). p. 249–54. doi: 10.1007/978-3-319-94523-1_22
79. Nerella S, Bihorac A, Tighe P, Rashidi P. Facial action unit detection on icu data for pain assessment. *arXiv preprint arXiv:2005.02121*. (2020). doi: 10.48550/arXiv.2005.02121
80. Chevrolet J-C, Jolliet P. Clinical review: agitation and delirium in the critically ill–significance and management. *Crit Care.* (2007) 11:214–214. doi: 10.1186/cc5787
81. Hughes CG, McGrane S, Pandharipande PP. Sedation in the intensive care setting. *Clin Pharmacol.* (2012) 4:53. doi: 10.2147/CPAA.S26582

82. Jackson DL, Proudfoot CW, Cann KF, Walsh TS. The incidence of sub-optimal sedation in the ICU: a systematic review. *Crit Care.* (2009) 13:R204. doi: 10.1186/cc8212
83. Selvan K, Edriss H, Sigler M, Tseng J. Self-extubation in ICU patients. *Southw Respirat Crit Care Chronicles.* (2014) 2:31–4. doi: 10.12746/swrccc.v2i8.169
84. Singh PM, Rewari V, Chandralekha A, Arora MK, Trikha A. A retrospective analysis of determinants of self-extubation in a tertiary care intensive care unit. *J Emerg Trauma Shock.* (2013) 6:241–5. doi: 10.4103/0974-2700.120363
85. Casale P, Pujol O, Radeva P. *Classifying Agitation in Sedated ICU Patients*. Girona: Citeseer (2010).
86. Rouche O, Wolak-Thierry A, Destoop Q, Milloncourt L, Floch T, Raclot P, et al. Evaluation of the depth of sedation in an intensive care unit based on the photo motor reflex variations measured by video pupillometry. *Ann Intensive Care.* (2013) 3:5–5. doi: 10.1186/2110-5820-3-5
87. Miranda D, Calderón M, Favela J. Anxiety detection using wearable monitoring. In: *Proceedings of the 5th Mexican Conference on Human-Computer Interaction*. New York, NY: ACM (2014). doi: 10.1145/2676690.2676694
88. Medrzycka-Dabrowska W, Lewandowska K, Kwiecień-Jaguś K, Czyz-Szypenbajl K. Sleep deprivation in intensive care unit - systematic review. *Open Med.* (2018) 13:384–93. doi: 10.1515/med-2018-0057
89. Ding Q, Redeker NS, Pisani MA, Yaggi HK, Knauert MP. Factors influencing patients' sleep in the intensive care unit: perceptions of patients and clinical staff. *Am J Crit Care.* (2017) 26:278–86. doi: 10.4037/ajcc2017333
90. Pisani MA, Friese RS, Gehlbach BK, Schwab RJ, Weinhouse GL, Jones SF. Sleep in the intensive care unit. *Am J Respir Crit Care Med.* (2015) 191:731–8. doi: 10.1164/rccm.201411-2099CI
91. Delaney LJ, Van Haren F, Lopez V. Sleeping on a problem: the impact of sleep disturbance on intensive care patients - a clinical review. *Ann Intensive Care.* (2015) 5:3–3. doi: 10.1186/s13613-015-0043-2
92. Andersen J, Boesen H, Skovgaard OK. Sleep in the Intensive Care Unit measured by polysomnography. *Minerva Anestesiol.* (2013) 79:804–15.
93. Elliott R, McKinley S, Cistulli P, Fien M. Characterisation of sleep in intensive care using 24-hour polysomnography: an observational study. *Crit Care.* (2013) 17:R46–R46. doi: 10.1186/cc12565
94. Lee J, Hong M, Ryu S. Sleep monitoring system using kinect sensor. *Int J Distribut Sensor Netw.* (2015) 11:875371. doi: 10.1155/2015/875371
95. Rosa RG, Tonietto TF, da Silva DB, Gutierres FA, Ascoli AM, Madeira LC, et al. Effectiveness and safety of an extended ICU visitation model for delirium prevention: a before and after study. *Crit Care Med.* (2017) 45:1660–7. doi: 10.1097/CCM.0000000000002588
96. Teixeira C, Rosa RG. The rationale of flexible ICU visiting hours for delirium prevention. *J Emerg Crit Care Med.* (2018) 2:7. doi: 10.21037/jeccm.2018.10.07
97. Torres C, Fried JC, Manjunath BS. Healthcare event and activity logging. *IEEE J Transl Eng Health Med.* (2018) 6:1600212. doi: 10.1109/JTEHM.2018.2863386
98. Voigt LP, Reynolds K, Mehryar M, Chan WS, Kostelecky N, Pastores SM, et al. Monitoring sound and light continuously in an intensive care unit patient room: a pilot study. *J Crit Care.* (2017) 39:36–9. doi: 10.1016/j.jcrc.2016.12.020
99. Berglund B, Lindvall T, Schwela DH, Organization WH. *Guidelines for Community Noise*. London: World Health Organization (1999).
100. Tainter CR, Levine AR, Quraishi SA, Butterly AD, Stahl DL, Eikermann M, et al. Noise levels in surgical ICUs are consistently above recommended standards. *Crit Care Med.* (2016) 44:147–52. doi: 10.1097/CCM.0000000000001378
101. Ben-Ari J, Zimlichman E, Adi N, Sorkine P. Contactless respiratory and heart rate monitoring: validation of an innovative tool. *J Med Eng Technol.* (2010) 34:393–8. doi: 10.3109/03091902.2010.503308
102. Condrea F, Ivan V-A, Leordeanu M. *In Search of Life: Learning from Synthetic Data to Detect Vital Signs in Videos*. (2020). p. 298–9. doi: 10.1109/CVPRW50498.2020.00157
103. van Gastel M, Stuijk S, de Haan G. New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring. *Sci Rep.* (2016) 6:38609. doi: 10.1038/srep38609
104. Guazzi AR, Villarroel M, Jorge J, Daly J, Frise MC, Robbins PA, et al. Non-contact measurement of oxygen saturation with an RGB camera. *Biomed Opt Express.* (2015) 6:3320–38. doi: 10.1364/BOE.6.003320
105. Lyra S, Mayer L, Ou L, Chen D, Timms P, Tay A, et al. A deep learning-based camera approach for vital sign monitoring using thermography images for ICU patients. *Sensors.* (2021) 21:1495. doi: 10.3390/s21041495
106. L'Her E, Nazir S, Pateau V, Visvikis D. Accuracy of noncontact surface imaging for tidal volume and respiratory rate measurements in the ICU. *J Clin Monitor Comput.* (2021) 21:708. doi: 10.1007/s10877-021-00708-x
107. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. *JMIR Mhealth Uhealth.* (2017) 5:e91. doi: 10.2196/mhealth.7208
108. Mebazaa A, Casadio MC, Azoulay E, Guidet B, Jaber S, Levy B, et al. Post-ICU discharge and outcome: rationale and methods of the The French and euROpean Outcome reGistry in Intensive Care Units (FROG-ICU) observational study. *BMC Anesthesiol.* (2015) 15:143. doi: 10.1186/s12871-015-0129-2
109. Gluck S, Chapple L-AS, Chapman MJ, Iwashyna TJ, Deane AM. A scoping review of use of wearable devices to evaluate outcomes in survivors of critical illness. *Crit Care Resuscitat.* (2017) 19:197.
110. Anderson JL, Yoward LS, Green AJ. A study investigating the validity of an accelerometer in quantification of step count in adult hospital inpatients recovering from critical illness. *Clin Rehabil.* (2019) 33:936–42. doi: 10.1177/0269215519829893
111. Gandotra S, Bakhru R, Shields K, Berry M, Files D. Monitoring activity levels in ICU survivors with accelerometers. *Chest.* (2017) 152:A394. doi: 10.1016/j.chest.2017.08.420
112. Gerke S, Yeung S, Cohen IG. Ethical and legal aspects of ambient intelligence in hospitals. *J Am Med Assoc.* (2020) 323:601–2. doi: 10.1001/jama.2019.21699
113. Parkhi OM, Vedaldi A, Zisserman A. *Deep Face Recognition*. Swansea: BMVA Press (2015). p. 6. doi: 10.5244/C.29.41
114. Schroff F, Kalenichenko D, Philbin J. *Facenet: A Unified Embedding for Face Recognition and Clustering*. (2015). p. 815–23. doi: 10.1109/CVPR.2015.7298682
115. Lahasan B, Lutfi SL, San-Segundo R. A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. *Artif Intellig Rev.* (2019) 52:949–79. doi: 10.1007/s10462-017-9578-y
116. Au-Yeung W-TM, Sahani AK, Isselbacher EM, Armoundas AA. Reduction of false alarms in the intensive care unit using an optimized machine learning based approach. *NPJ Digit Med.* (2019) 2:1–5. doi: 10.1038/s41746-019-0160-7
117. Marra A, Ely EW, Pandharipande PP, Patel MB. The ABCDEF bundle in critical care. *Crit Care Clin.* (2017) 33:225–43. doi: 10.1016/j.ccc.2016.12.005

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 | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 08 August 2022
DOI 10.3389/fdgth.2022.877234

OPEN ACCESS

EDITED BY
Mohamed Elgendi,
ETH Zürich, Switzerland

REVIEWED BY
Yongbo Liang,
Guilin University of Electronic Technology, China
Colin K. Drummond,
Case Western Reserve University, United States

*CORRESPONDENCE
Masamitsu Kamon
Masamitsu.mk@gmail.com

SPECIALTY SECTION
This article was submitted to Health Informatics, a section of the Journal Frontiers in Digital Health

RECEIVED 16 February 2022
ACCEPTED 19 July 2022
PUBLISHED 08 August 2022

CITATION
Kamon M, Okada S, Furuta M and Yoshida K (2022) Development of a non-contact sleep monitoring system for children.
Front. Digit. Health 4:877234.
doi: 10.3389/fdgth.2022.877234

Development of a non-contact sleep monitoring system for children

Masamitsu Kamon[1]*, Shima Okada[1], Masafumi Furuta[2] and Koki Yoshida[2]

[1]Department of Robotics, Ritsumeikan University, Shiga, Japan, [2]Technology Research Laboratory, Shimadzu Corporation, Kyoto, Japan

Daily monitoring is important, even for healthy children, because sleep plays a critical role in their development and growth. Polysomnography is necessary for sleep monitoring. However, measuring sleep requires specialized equipment and knowledge and is difficult to do at home. In recent years, smartwatches and other devices have been developed to easily measure sleep. However, they cannot measure children's sleep, and contact devices may disturb their sleep.

A non-contact method of measuring sleep is the use of video during sleep. This is most suitable for the daily monitoring of children's sleep, as it is simple and inexpensive. However, the algorithms have been developed only based on adult sleep, whereas children's sleep is known to differ considerably from that of adults.

For this reason, we conducted a non-contact estimation of sleep stages for children using video. The participants were children between the ages of 0–6 years old. We estimated the four stages of sleep using the body movement information calculated from the videos recorded. Six parameters were calculated from body movement information. As children's sleep is known to change significantly as they grow, estimation was divided into two groups (0–2 and 3–6 years).

The results show average estimation accuracies of 46.7 ± 6.6 and $49.0 \pm 4.8\%$ and kappa coefficients of 0.24 ± 0.11 and 0.28 ± 0.06 in the age groups of 0–2 and 3–6 years, respectively. This performance is comparable to or better than that reported in previous adult studies.

KEYWORDS

sleep stage, sleep monitoring, children, video monitoring, video image processing, machine leaning

Introduction

Sleep is an important part of daily life as it reduces stress and aids recovery. Accordingly, monitoring sleep at home can contribute to health management (1, 2). In addition, sleep plays an important role in the development of children (3) and is also related to their parents' health (4). Sleep quality in children can contribute to health management. Therefore, it is important to monitor sleep quality in children on a daily basis to detect sleep problems. Polysomnography (PSG) is commonly used to assess sleep quality in clinical practice, but it requires measuring various biological signals,

such as electroencephalograms (EEGs), electromyograms, and electrooculograms and requires specialized equipment and knowledge (5–7). Therefore, PSG scoring is not a realistic method for assessing the sleep stage in the home environment. In addition, the installation of various devices may diminish the quality of sleep in children. Thus, measurement methods that reduce the burden on equipment installation are needed.

In recent years, electrocardiograms (ECGs), pulse-rate laser Doppler sensors (8, 9), and cameras (10) have been used to quantify sleep quality in homes (11). In particular, camera-based methods are considered the most suitable for monitoring children's sleep because they are entirely noncontact and easy to install. This method uses the relationship between body movements and sleep stages, and it has been reported that there are significant differences in the frequencies of body movements among sleep stages (12). Nochino et al. (13) used an infrared camera to measure sleep quality in four stages. However, this method is intended for adults and is unsuitable for assessing sleep in children.

Long et al. (14) quantified sleep features and patterns in children by using an infrared camera. In this study, children's sleep was estimated during both waking and sleeping stages. However, it is believed that the secretion of growth hormone responsible for children's development occurs during deep sleep (15). In addition, de Goederen et al. (16) used a radar system to estimate the four stages of sleep in children, and achieved an accuracy of 58.0%. However, the radar system is difficult to install and expensive, thus making it difficult to use at home.

There are various sleep measurement devices, and methods based on body movements (17). Two commonly used devices include the smartwatch and smartphone. Chinoy et al. (18) evaluated the smartwatch Fitbit Alta HR. The device provided an estimation accuracy of 90% for the distinction between sleep and wakefulness for adults. Patel et al. (19) used a dedicated application to assess sleep on smartphones. Their results confirmed the lack of a correlation relationship between the sleep stage determined by the application and that determined by PSG.

Therefore, it is necessary to develop a simple method for monitoring children's sleep. We used an infrared camera, which is easy to install and is inexpensive, to estimate the four sleep stages of children aged 0–6 years. As it is known that sleep in children changes considerably as they grow, we estimated the sleep stages of children aged 0–2 years and 3–6 years. Previous studies (13) have utilized support vector machines (SVMs) (20). In this study, the extremely randomized trees (Extra Trees) ensemble machine learning algorithm was used.

Methods

In this study, we extracted body movement during sleep from video recordings. **Figure 1** shows the measurement and analysis procedure. The measurements were performed in the subjects' homes. Eight subjects participated who were aged 0–6 years, with a mean age of 2.3 ± 2.1 years. The measurements were conducted for 1–4 nights per person for a total of 14 nights. The variability in the number of nights per person is attributed to the fact that the measurements were conducted at home, and the measurements were not always accurate.

Informed consent for measurements was obtained from the parent of all subjects. The experiment was approved by the Ethical Review Committee for Medical Research Involving Human Subjects, Ritsumeikan University (BKC-Human Medicine-2020-053). A simple PSG was used as the gold standard for learning and evaluating the classifier. This reduced the burden on the child because the sleep stages were assessed with far fewer electrodes than those used in PSG in clinical practice. A small sensor was attached to the body for recordings with the simplified PSG. The recording device was connected wirelessly to reduce the burden on the child during sleep. A simple PSG was used offline by a technician to score the sleep stages. A video of the child sleeping was recorded at the same time as the simple PSG recording. **Figure 2** shows the flow of sleep stage estimation from video recordings. Body movements were extracted from the videos using video processing. Six machine learning parameters were calculated from the extracted body movements. Sleep stages were estimated from six parameters using Extra Trees machine learning with the simple PSG sleep stage as the answer response sleep stages.

PSG recording and sleep stage scoring

Simple PSG scoring was performed using a ZA-X EEG sensor (Proassist, Japan). This device has four electrodes and

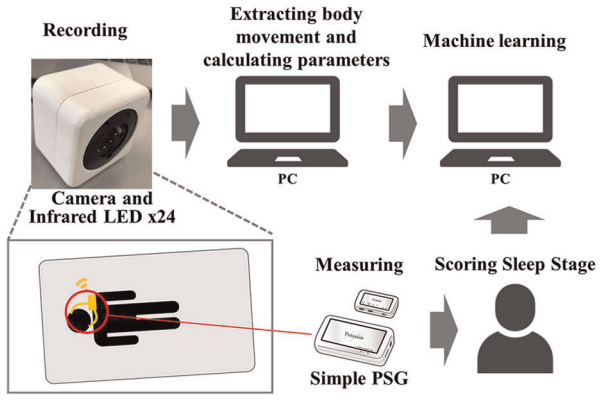


FIGURE 1
Measurement and analysis methods. A sleeping child was recorded by an infrared camera, and body movements were extracted from the video data. Some features were calculated by the body movements as the training parameter for machine learning. And machine learning was performed using the calculated futures and sleep stage obtained by a simple PSG as the correct data.

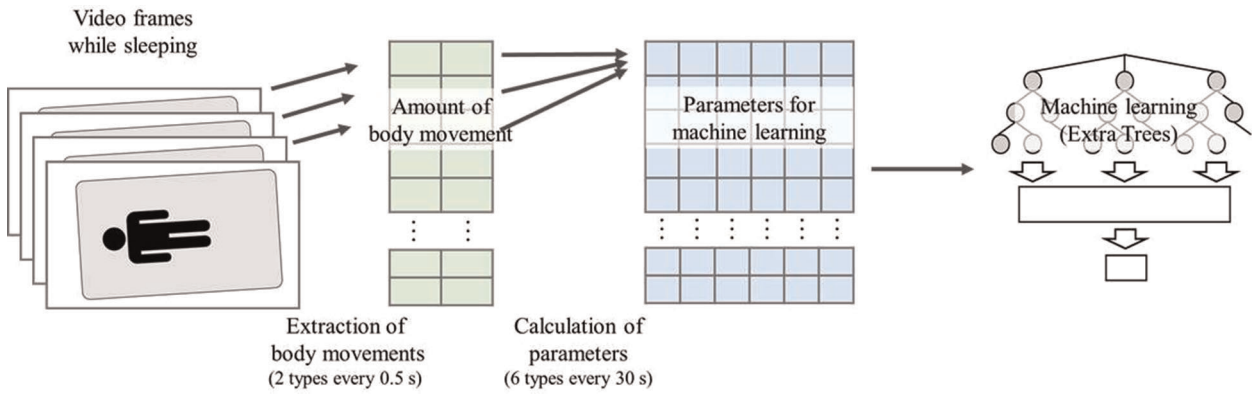


FIGURE 2
Flow of sleep stage estimation. Two types of body movements were calculated every 0.5 s using difference processing, and six features were calculated every 30 s. Then, machine learning was apply these feature values to sleep stage estimation.

can measure data to estimate the sleep stage. The results of this simple PSG instrument have been reported to have a concordance rate of 85.5% compared with PSG scoring (21). The measured data were used to estimate the sleep stages automatically every 30 s based on the American Academy of Sleep Medicine (8). Subsequently, a specialist technician corrected the results. Stage W is defined as Wake, stage R as rapid eye movement (REM), stages N1 and N2 as Light, and stage N3 as Deep.

Region-of-interest selection and body movement extraction

An ELP-USBFHD05MT-DL36-J camera (Shenzhen Ailipu Technology Co., Ltd., China) equipped with infrared light-emitting diodes (LEDs) was used. The camera's resolution was $1{,}920 \times 1{,}080$ pixels, and measurements were performed at 2 frames per second. The camera was controlled by a Jetson Nano (NVIDIA Corporation, United States of America). The camera was placed around the bedding, so that the entire bedding area could be observed. The camera was placed at a height of about 1–2 m from the bed. For the experimental design, it is better to install the cameras at the same height under the same conditions. However, since this study was conducted under the conditions that each household could install the camera, it was not possible to standardize the height conditions.

Figure 3 shows the body movement extraction procedure. The bedding was extracted as a region-of-interest (ROI) in the video for body motion extraction. The ROI extraction eliminated body movements (other than those related to the subject), such as the unintended appearance of parents. The ROI was set manually by the researcher. The extracted ROIs were transformed by trapezoidal correction according to their size. For the extraction of body motion, we used differences and binarization. The difference calculation method is shown in Equation 1. I_d is the difference image, I_n is the n th frame, x is the vertical pixel coordinate, and y is the horizontal pixel coordinate.

$$I_d(x, y) = |I_n(x, y) - I_{n-1}(x, y)| \tag{1}$$

In this process, the difference in pixel-by-pixel luminance values between the nth and the $(n-1)$th frames of the grayscale video was calculated. In the binarization process, pixels above the threshold were converted to white pixels (luminance value: 255), and the pixels below the threshold to black (luminance value: 0). Finally, the number of white pixels was counted as the amount of body movement. Equation 2 shows the equation for the amount of body motion. $dif\ f_n$ is the amount of body motion in the nth frame, I_{bn} is the image after binarization in the nth frame x is the vertical pixel coordinate, and y is the horizontal pixel coordinate.

$$dif\ f_n = \frac{\sum_{i=1}^{x} \sum_{j=1}^{y} I_{bn}(x, y)}{255} \tag{2}$$

Motion extraction was performed twice, every 0.5 s, and two motion values were extracted. The first amount of body movement was calculated between the nth and $(n-1)$th frames, corresponding to a very small temporal difference (0.5 s), allowing us to detect fast body movements. The second amount of body movement was computed between the nth and $(n-6)$th frames, which corresponds to a large temporal difference (3.0 s), therefore allowing the detection of slow body movements.

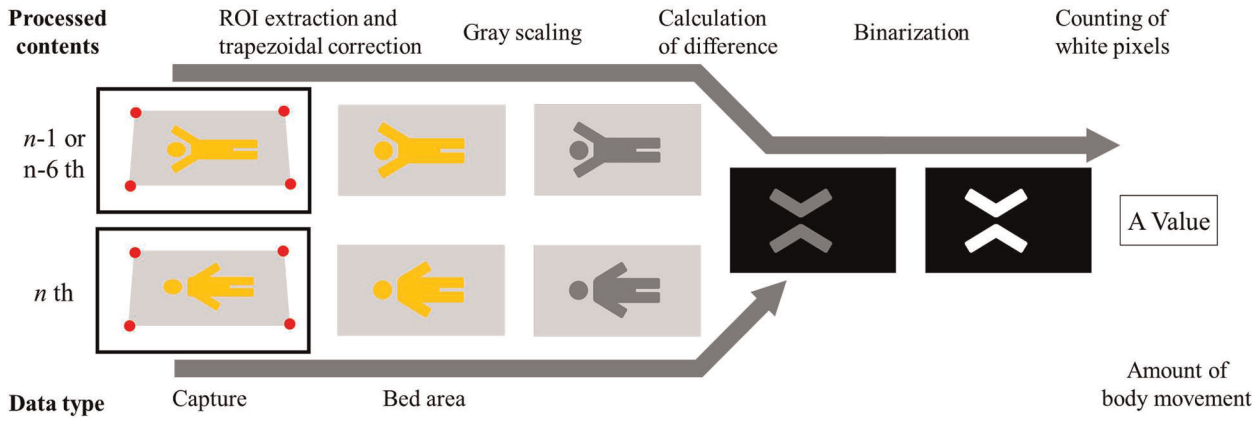


FIGURE 3
Procedure of the body movement extraction. Only the bed area was extracted from the video using trapezoidal correction, and converted to a black-and-white image using grayscaling, differencing, and binarization processes. The number of white color pixels, It means that a pixel has had a change in luminance value since the previous frame, was counted as the amount of body movement.

Sleep stage estimation

Extra Trees was used to estimate the sleep stages from body movements (22). To use this classification method, six parameters were calculated from two types of extracted body movements. These parameters were calculated once every 30 s. This process is the same timing as the determination of sleep stage based on simple PSG scoring. Extra Trees was then trained using six parameters and the sleep stages from simple PSG scoring. Sleep stages were estimated sequentially by using the six parameters. The amount of body movement varied with sleep stages (12). The amount of physical activity was related to sleep, even in children (23, 24). In Wake, body movements were large and frequent; in Light and REM, body movements tended to be more frequent, but the number of body movements was greater in Light than in REM. Conversely, in the case of Deep, body movements were almost absent. These results indicate that the average amount of body movements, frequency of body movements, and elapsed time are necessary to identify sleep stages. Therefore, six parameters were calculated using the following equation to represent the characteristics necessary for estimation. Note that $dif\,f_{i,0.5}$ is the fast body movement of the i th frame, and $dif\,f_{i,3.0}$ is the slow body movement of the i th frame.

Parameter 1: The first parameter was estimated using Equation 3 ($n = 1,\ 61,\ 121,\ \cdots$). This parameter is a moving average of body movements during 30 s and represents the magnitude of body movements. This parameter was used for Wake classification.

$$SumMean_{30} = \frac{\sum_{i=n}^{n+59} dif\,f_{i,0.5} + dif\,f_{i,3.0}}{60} \tag{3}$$

Parameter 2: The second parameter was estimated based on Equation 4 ($n = 1,\ 61,\ 121,\ \cdots$). This parameter is the logarithm of the moving average over a period of 300 s. If the result is less than one, then it is replaced by zero. If it is greater than one, then the logarithm is calculated. This parameter represents the frequency of body movements over 5 min and is used to discriminate between REM and Light.

$$PropMean_{300} = \log_{10}\left(\frac{\sum_{i=n}^{n+599} dif\,f_{i,0.5} \times dif\,f_{i,3.0}}{600}\right) \tag{4}$$

Parameter 3: The third parameter is given by Equation 5 ($n = 1,\ 61,\ 121,\ \cdots$). This parameter is a moving average over a period of 300 s and represents the magnitude of body movement over a long period of time. This parameter was used to discriminate the long-time Wake and long-time Deep with little movement.

$$SumMean_{300} = \frac{\sum_{i=n}^{n+599} (dif\,f_{i,0.5} + dif\,f_{i,3.0})}{600} \tag{5}$$

Parameter 4: The fourth parameter is given by Equation 6. This parameter is the duration of frames during which body motion is below a certain level, thus representing the duration of the stationary state. Similar to parameter 3, this parameter was used to discriminate between Deep and little movements.

$$\begin{cases} NonBM = 0 & dif\,f_{i,0.5} + dif\,f_{i,3.0} > k \\ NonBM = NonBM + 1 & dif\,f_{i,0.5} + dif\,f_{i,3.0} \leq k \end{cases} \tag{6}$$

Parameter 5: The fifth parameter is given by Equation 7 ($n = 1,\ 61,\ 121,\ \cdots$). This is the time elapsed since the start of the measurements. This parameter represents the number of epochs. There was a tendency for deeper sleep in the first half and more REM in the second half. Therefore, this parameter was used to determine the difference between Deep and REM.

$$Epoch = \frac{n + 59}{60} \tag{7}$$

Parameter 6: The sixth parameter is given by Equation 8 ($n = 1,\ 61,\ 121,\ \cdots$). This parameter denotes the dispersion of body movements during a period of 30 s and represents the frequency of body movements. It was used to classify Deep, which contains few body movements, and REM, which contains many body movements.

$$Variance_{30} = \left(\sum_{i=n}^{n+59} (dif\,f_{i,0.5} + dif\,f_{i,3.0})^2 - \overline{(dif\,f_{i,0.5} + dif\,f_{i,3.0})}^2 \right) \tag{8}$$

The size of a child's body varies considerably at different stages of growth. In this study, the amount of body movement used varied considerably depending on the size of the child's body. Sleep duration also varied between the subjects and days. This may lead to inaccurate estimation of the sleep stages. Therefore, all parameters were standardized to have a mean of zero and a variance of one. Standardization was implemented based on Equation 9, where i denotes the number of parameters, $Param_i$ is the i th parameter, $\overline{Param_i}$ is the average of the i th parameter, and S_{Param_i} denotes the variance of the i th parameter.

$$Standard\_Param_i = \frac{Param_i - \overline{Param_i}}{S_{Param_i}} \tag{9}$$

These parameters constitute discrete data. The parameters were not expected to differ considerably for each sleep stage. Therefore, the classifier needs to be trained by dividing the data into smaller pieces. For this reason, we used Extra Trees based on decision trees as the classifier used in this study.

Examples of the results for the six calculated parameters are presented in Figure 4. Because the tendency of the sleep stages varied considerably depending on the child's development, machine learning with the Extra Trees classifier was used to estimate and evaluate the two groups of children (0–2 years old, and 3–6 years old).

Leave-one-out cross-validation was performed to evaluate the classification performance, and accuracy, sensitivity, and specificity were calculated for each of the four sleep stages to evaluate the classification performance. The overall estimated accuracy and Cohen's kappa coefficient (25) were calculated.

Results and discussion

Results of simple PSG scoring

In total, 50 nights were evaluated: 21 nights for children with ages in the range of 0–2 years, and 29 nights for those

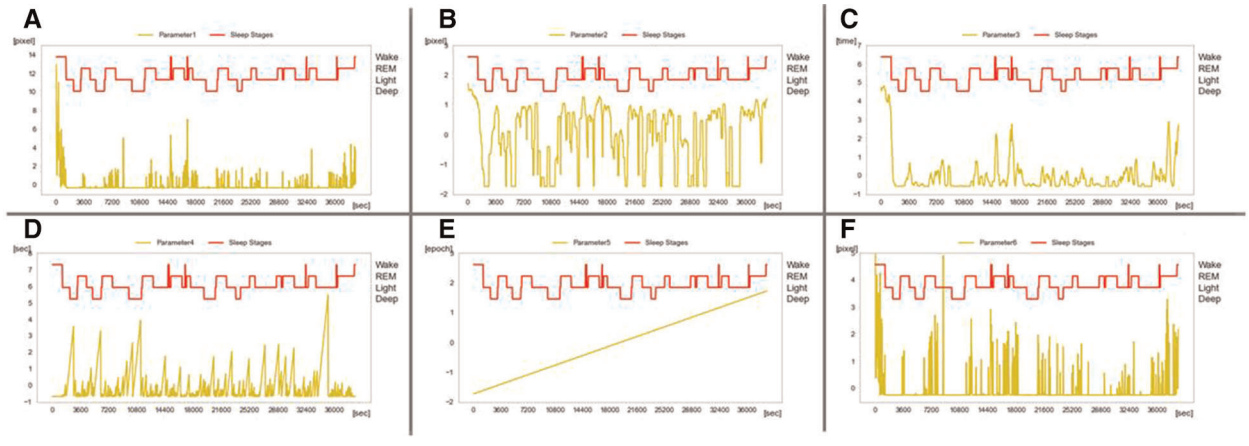


FIGURE 4
Sleep stage and calculated parameters. (A) Moving average over 30 s and represents the magnitude of the body movement over a short period of time. (B) Logarithm of the moving average over 300 s and representation of the frequency of body movements over a long period of time. (C) Moving average over 300 s and representation of the magnitude of the body movement over a long period of time. (D) Duration of body movements below a certain level and representation of the duration of stillness. (E) Number of frames converted to the number of epochs and representation of the elapsed time. (F) Variance of body movements over a period of 30 s, and representation of the frequency of body movements over a short period of time.

in the 3–6 year age range. However, in many cases, scoring with the simple PSG failed, and the data available for analysis were limited to six nights for children aged 0–2 years old, and eight nights for children aged 3–6 years old (a total of 14 nights). Most of the simple PSG scoring failures were caused by children disliking the electrodes and removing them. Another reason was the difficulty in operating the equipment. Although PSG scoring is necessary for research, PSG scoring is not a suitable method for daily monitoring of sleep stages in children.

The percentage of each sleep stage obtained by PSG as correct data is shown in **Figure 5**. The epochs are the total number of sleep stages determined once every 30 s from PSG. From this, REM sleep is more common in children aged 0–2 years old, while REM is less common in children aged 3–6 years old due to changes in the sleep cycles. However, in clinical PSG examinations, the depth was approximately 30% in normal subjects (26). However, the simple PSG scoring results were low: 18% in children who were 0–2 years old, and 22% in children who were 3–8 years old. This may be because the device used was a simple PSG scoring device, the subjects were children, and their sleep cycles were not stable.

Sleep stage estimation

Examples of estimated sleep stages are shown in **Figure 6**. **Table 1** presents the estimated results for each sleep stage. **Table 2** shows the estimation accuracy and kappa coefficient of each dataset. The estimation accuracy of Wake was 90.3 ± 5.1% and was the highest in the age group of 0–2 years, and 97.5 ± 1.3% in the age group of 3–6 years. This was owing to frequent body movements during wakefulness, which was well characterized by parameter 1. However, the specificity of the 4th, 5th, and 6th datasets of children with ages in the range of 0–2 years was low. This may be owing to the short time between bedtime and sleep onset and the low frequency of mid-waking. The same tendency was observed in the second children data with ages in the range of 3–6 years. The estimation accuracy of Light was the lowest in both grooves, and its sensitivity and specificity were also low. This was due to the fact that Light accounts for approximately half of the sleep period and includes the transition period to other sleep stages. Therefore, it is inevitable that misclassification will increase in machine learning. The accuracy of REM was 70.6 ± 7.2% in children with ages in the range of 0–2 years, and 66.8 ± 4.7% in the 3–6 age group, thus indicating that the accuracy of REM estimation was higher in the 0–2 age group compared with that for children the 3–6 age group. This was because REM accounts for a larger proportion of total sleep in children with ages in the range of 0–2 years. In addition, children's sleep showed the same performance as that of adults in the previous study, despite a higher percentage of REM sleep. This is attributed to the effects of parameters 2, 5, and 6. The accuracy of Deep sleep estimation was 75.9 ± 3.7% in children with ages in the range of 0–2 years, and 79.5 ± 3.3% in children with ages in the range of 3–6 years, probably because parameters 3–6 could be used to determine Deep sleep. However, the low specificity of Deep sleep in children with ages in the range of 0–2 years may be owing to the low number of Deep sleep patterns in the sleep stages as judged by simple PSG. The accuracy of the estimation of each sleep stage is comparable to that reported in a previous study on adults (13).

Table 2 shows that the mean accuracy of the estimates for the age group 0–2 years old is 46.7 ± 6.6%, and the kappa coefficient is 0.24 ± 0.11. In the age group of 3–6 years old, the mean accuracy of the estimation is 49.0 ± 4.8%, and the kappa coefficient is 0.28 ± 0.06. In a previous study (13), the results for adults yielded an average estimation accuracy of

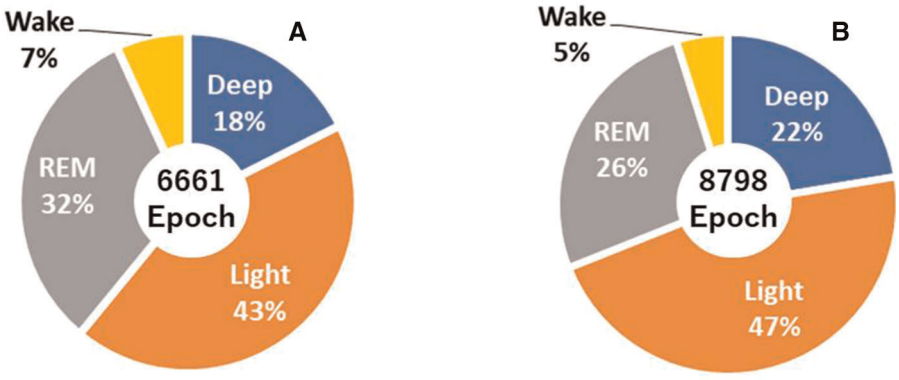


FIGURE 5
(A) Percentages of sleep stages determined by simple polysomnography (PSG) at ages in the range of 0–2 years. (B) Percentage of sleep stages determined by simple PSG at ages in the range of 3–6 years.

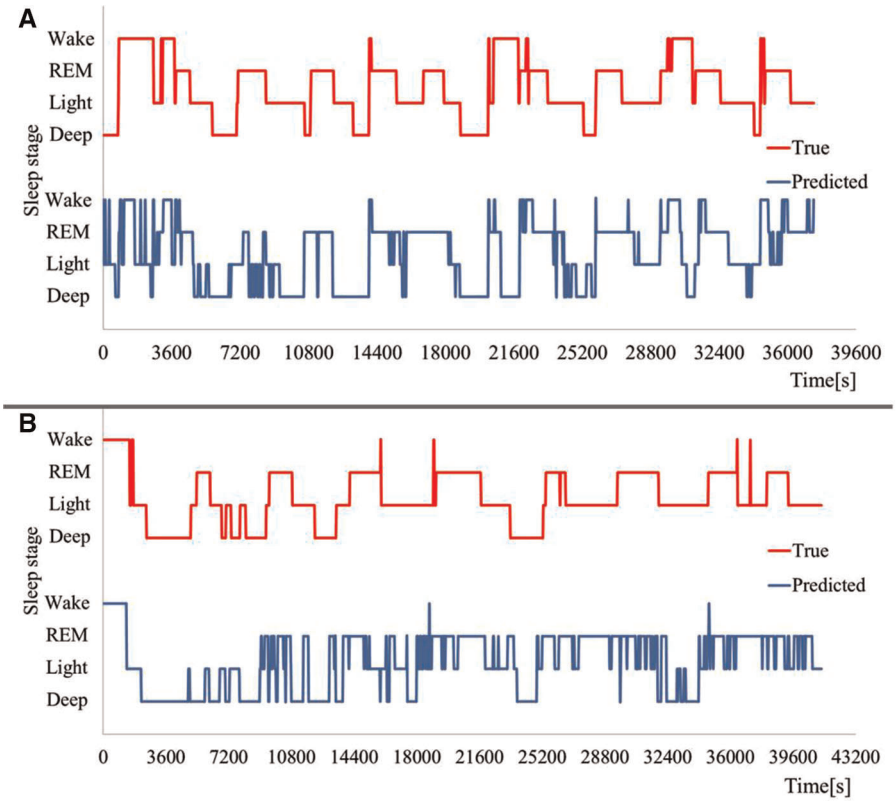


FIGURE 6
Example of sleep stage estimation results. (A) Children with ages in the ranges of (A) 0–2 years, and (B) 3–6 years.

TABLE 1 Estimation of the different sleep stages for each studied subject.

Age	Stages		Data number								Average
			1	2	3	4	5	6	7	8	
0–2	Deep	Accuracy	73.7	78.1	78.9	73.0	71.5	80.4	–	–	75.9 ± 3.7
		Sensitivity	31.8	66.5	16.3	23.8	58.3	66.4	–	–	43.8 ± 22.5
		Specificity	19.9	37.3	21.5	35.6	37.5	50.4	–	–	33.7 ± 11.4
	Light	Accuracy	60.8	62.9	47.9	49.8	54.8	62.9	–	–	56.5 ± 6.7
		Sensitivity	36.6	34.2	37.3	38.3	26.1	34.4	–	–	34.5 ± 4.4
		Specificity	60.4	53.9	44.0	37.5	51.2	63.1	–	–	51.7 ± 9.7
	REM	Accuracy	79.8	75.6	60.1	70.6	64.9	72.8	–	–	70.6 ± 7.2
		Sensitivity	85.0	70.6	50.4	47.5	61.1	65.4	–	–	63.3 ± 13.8
		Specificity	61.7	59.1	42.2	54.8	43.6	61.9	–	–	53.8 ± 8.9
	Wake	Accuracy	93.3	86.1	95.4	83.2	95.2	88.5	–	–	90.3 ± 5.1
		Sensitivity	65.2	40.8	96.5	44.6	28.6	83.3	–	–	59.8 ± 26.4
		Specificity	75.3	54.0	49.1	14.5	36.4	3.9	–	–	38.9 ± 26.4
3–6	Deep	Accuracy	80.8	82.2	82.6	79.0	73.8	78.4	76.1	84.4	79.5 ± 3.3
		Sensitivity	69.7	71.4	67.6	70.4	54.3	65.5	51.4	75.2	66.5 ± 6.3
		Specificity	52.8	57.4	54.4	45.5	58.7	46.3	58.4	68.8	52.5 ± 5.5
	Light	Accuracy	58.5	57.5	57.2	55.0	51.0	49.4	58.7	50.5	54.8 ± 3.8
		Sensitivity	33.5	45.6	29.2	29.1	34.4	20.5	51.9	16.7	32.0 ± 8.3
		Specificity	62.2	66.7	60.3	52.4	33.1	43.6	45.6	44.3	53.1 ± 12.7
	REM	Accuracy	72.1	71.5	65.1	68.5	63.5	60.2	68.7	58.5	66.8 ± 4.7
		Sensitivity	71.8	53.4	73.9	67.7	27.4	58.9	39.4	67.1	58.8 ± 17.3
		Specificity	45.5	39.9	44.5	49.5	30.0	34.0	41.6	29.1	40.6 ± 7.4
	Wake	Accuracy	97.3	95.1	98.6	98.8	96.5	97.4	99.1	97.2	97.3 ± 1.4
		Sensitivity	89.7	75.0	85.1	74.6	97.8	71.4	91.2	66.7	82.3 ± 10.3
		Specificity	76.1	9.2	78.4	95.7	62.9	89.6	96.3	73.9	68.6 ± 31.2

TABLE 2 Total accuracy and kappa coefficient outcomes.

Age		Data number								
		1	2	3	4	5	6	7	8	Average
0–2	Total accuracy [%]	53.8	51.4	41.2	38.3	43.2	52.3	–	–	46.7 ± 6.6
	Kappa	0.35	0.33	0.1	0.12	0.19	0.32	–	–	0.24 ± 0.11
3–6	Total accuracy [%]	54.36	53.15	51.83	50.66	42.42	42.71	51.3	45.3	49.0 ± 4.8
	Kappa	0.37	0.3	0.31	0.29	0.18	0.21	0.3	0.25	0.28 ± 0.06

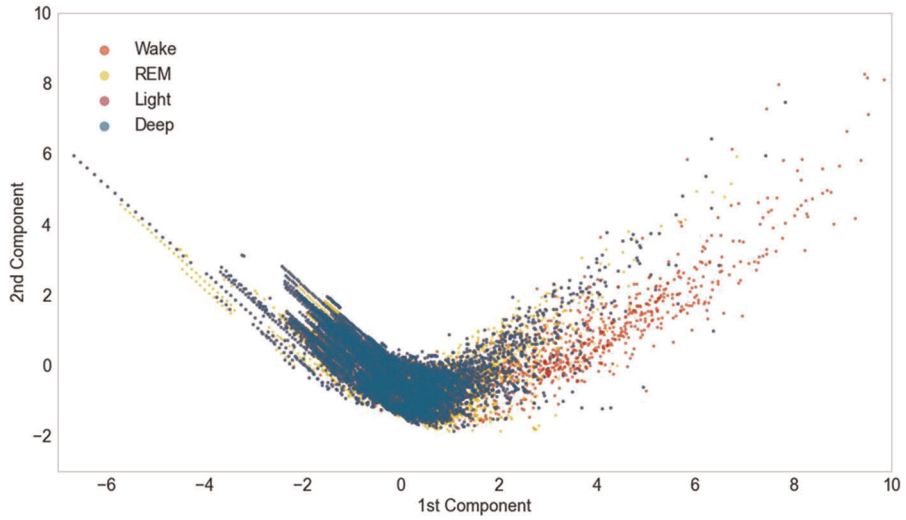


FIGURE 7
Principal component analysis (PCA) results for six parameters, namely, magnitude of short-time body movement, frequency of long-time body movement, magnitude of long-time body movement, rest time, elapsed time, and frequency of short-time body movement. The horizontal axis is the first principal component, and the vertical axis is the second principal component.

40.5 ± 2.2% and a kappa coefficient of 0.19 ± 0.04, which are considered to reflect an equivalent or a better performance. We believe that this is caused by the normalization of the parameters. This may have reduced the effects of differences in body size and total sleep time.

Another reason for the improved accuracy is attributed to the fact that we changed the classifier from an SVM to Extra Trees. **Figure 7** shows the results of the principal component analysis (PCA) for the six parameters used for visualization (27). The repeatability of the PCA was 66.8%. The results show that Extra Trees is more suitable than SVM (which uses boundaries) because the difference between the features of each stage is very small in the sleep stage.

Liang et al. (28) proposed a method to estimate sleep stages using Fitbit. In this study, they reported that the estimation accuracy of the four sleep stages for adults was 73.1 ± 11.9%. Although the accuracy of our method was lower than that of the study by Liang et al. (28), it is sufficient to be able to conduct complete measurements without contact, and to be applicable to children. In this study, a simple PSG was used as the reference for actual sleep stages. However, even the clinically used PSG tests are not completely consistent among specialist technicians (29, 30).

The estimated results for the age group of 0–2 years were lower than those for the age group of 3–6 years. This is because the sleep cycle of children with ages in the range of 0–2 years is less stable than that of children with ages in the range of 3–6 years. This is also confirmed by the variance, which shows that the estimation accuracy and variance of the kappa coefficient are greater for children with ages in the range of 0–2 years.

Although our system is less accurate than contact devices, it is useful because it can be used for children with ages in the range of 0–6 and because it is noncontact. However, our system also has limitations. The system does not measure autonomic nervous system activity. Thus, these indicators are also needed if high-estimation accuracy is desired. Therefore, it is possible to reduce the misclassification of Deep and REM to Light.

Extra Trees was used in this study. However, as there are many types of classifiers, it is necessary to compare the performance with other classifiers.

Conclusions

The estimation of sleep stages in children has not been studied as comprehensively as studied in adults owing to difficulties in measurement and PSG scoring. In this study, we developed a sleep monitoring system for children with the use of an infrared camera. The developed system extracted body movements from a video and calculated six parameters from body movements. The system then used Extra Trees to estimate the four stages of sleep from the parameters. The accuracy of the developed system was approximately 45% on average compared with that of the simple PSG. However, the average estimation accuracy for Deep, where growth hormone is said to be secreted, was more than 75%. The performance of this system was comparable to or better than those reported in previous studies. For these reasons, it is suggested that this system can be used as a noncontact sleep monitoring system for children at home.

Our system can estimate four sleep stages, but not the REM/NREM sleep cycle. It is known that children's growth can be confirmed by the REM/NREM sleep cycle [26]. Therefore, developing a method to estimate the REM/NREM sleep cycle is necessary.

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

The studies involving human participants were reviewed and approved by Ethical Review Committee for Medical Research Involving Human Subjects. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

MK, SO, MF, and KY contributed to conception and design of the study. MK has organized and parsed the database. MK wrote the manuscript. MK wrote the manuscript and all authors contributed to the current revision. All authors have read and approved the submitted version. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to Dr. Masako Taniike of Osaka University for helpful discussions and comments on the study.

Conflict of interest

Masafumi and Koki works for Shimadzu Corporation, but there is no conflict of interest.

Publisher's note

References

1. Van Reeth O, Weibel L, Spiegel K, Leproult R, Dugovic C, Maccari S. Physiology of sleep (review)–interactions between stress and sleep: from basic research to clinical situations. *Sleep Med Rev*. (2000) 4(2):201–19. doi: 10.1053/smrv.1999.0097

2. Brandolim Becker N, Neves de Jesus S, Nuno Viseu J. Sleep quality and stress: a literature review. Medical Education and Motivation in Family Medicine Residency View Project Female sexual dysfunctions View project [Internet]. (2015) 53–61.

3. Graven SN, Browne JV. Sleep and brain development. The critical role of sleep in fetal and early neonatal brain development. *Newborn Infant Nurs Rev*. (2008) 8(4):173–9. doi: 10.1053/j.nainr.2008.10.008

4. Bayer JK, Hiscock H, Hampton A, Wake M. Sleep problems in young infants and maternal mental and physical health. *J Paediatr Child Health*. (2007) 43(1–2):66–73. doi: 10.1111/j.1440-1754.2007.01005.x

5. Rechtschaffen A. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. *Brain Inf Serv*. (1968) 20 (2):246–7. doi: 10.1001/archpsyc.1969.01740140118016

6. Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, et al. The visual scoring of sleep in adults. *J Clin Sleep Med*. (2007) 3(02):121–31. doi: 10.5664/jcsm.26814

7. Berry RB, Brooks R, Gamaldo C, Harding SM, Lloyd RM, Quan SF, et al. AASM Scoring manual updates for 2017 (version 2.4). *J Clin Sleep Med*. (2017) 13(5):665–6. doi: 10.5664/jcsm.6576

8. Yoon H, Hwang SH, Choi JW, Lee YJ, Jeong DU, Park KS. Slow-wave sleep estimation for healthy subjects and OSA patients using R-R intervals. *IEEE J Biomed Heal Inf*. (2018) 22(1):119–28. doi: 10.1109/JBHI.2017.2712861

9. Kagawa M, Sasaki N, Suzumura K, Matsui T. Sleep stage classification by body movement index and respiratory interval indices using multiple radar sensors. *IEEE Eng Med Biol Soc Annu Int Conf*. (2015) 7606–9. doi: 10.1109/EMBC.2015.7320153

10. Hong H, Zhang L, Gu C, Li Y, Zhou G, Zhu X. Noncontact sleep stage estimation using a CW Doppler radar. *IEEE Trans Emerg Sel Top Circuits Syst*. (2017) 3357(c):1–11. doi: 10.1109/JETCAS.2017.2789278

11. Aaronson ST, Rashed S, Biber MP, Hobson JA. Brain state and body position. *Arch Gen Psychiatry*. (1982) 39(3):330–5. doi: 10.1001/archpsyc.1982.04290030062011

12. Wilde-Frenz J, Schulz H. Rate and distribution of body movements and during sleep in humans. *Percept Mot Ski*. (1983) 56(1):275–83. doi: 10.2466/pms.1983.56.1.275

13. Nochino T, Ohno Y, Kato T, Taniike M, Okada S. Sleep stage estimation method using a camera for home use. *Biomed Eng Lett*. (2019) 9(2):257–65. doi: 10.1007/s13534-019-00108-w

14. Long X, Otte R, van der Sanden E, Werth J, Tan T. Video-based actigraphy for monitoring wake and sleep in healthy infants: a laboratory study. *Sensors (Switzerland)*. (2019) 19(5):1075. doi: 10.3390/s19051075

15. Vermes I, Dohanics J, Toth G, Pongracz J. Maturation of the circadian rhythm of the adrenocortical functions in human neonates and infants. *Horm Res*. (1980) 12(5):237–44. doi: 10.1159/000179126

16. de Goederen R, Pu S, Silos Viu M, Doan D, Overeem S, Serdijn WA, et al. Radar-based sleep stage classification in children undergoing polysomnography: a pilot-study. *Sleep Med*. (2021) 82:1–8. doi: 10.1016/j.sleep.2021.03.022

17. Lujan MR, Perez-Pozuelo I, Grandner MA. Past, present, and future of multisensory wearable technology to monitor sleep and circadian rhythms. *Front Digit Heal*. (2021) 3:721919. doi: 10.3389/fdgth.2021.721919

18. Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*. (2021) 44(5):zsaa291. doi: 10.1093/sleep/zsaa291

19. Patel P, Kim JY, Brooks LJ. Accuracy of a smartphone application in estimating sleep in children. *Sleep Breath*. (2017) 21(2):505–11. doi: 10.1007/s11325-016-1425-x

20. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. (1995) 20 (3):273–97. doi: 10.1007/BF00994018

21. Nonoue S, Mashita M, Haraki S, Mikami A, Adachi H, Yatani H, et al. Inter-scorer reliability of sleep assessment using EEG and EOG recording system in comparison to polysomnography. *Sleep Biol Rhythms*. (2017) 15(1):39–48. doi: 10.1007/s41105-016-0078-2

22. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. (2006) 63(1):3–42. doi: 10.1007/s10994-006-6226-1

23. So K, Michael Adamson T, Horne RSC. The use of actigraphy for assessment of the development of sleep/wake patterns in infants during the first 12 months of life. *J Sleep Res*. (2007) 16(2):181–7. doi: 10.1111/j.1365-2869.2007.00582.x

24. Sadeh A, Acebo C, Seifer R, Aytur S, Carskadon MA. Activity-based assessment of sleep-wake patterns during the 1st year of life. *Infant Behav Dev*. (1995) 18(3):329–37. doi: 10.1016/0163-6383(95)90021-7

25. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. (1960) 20(1):37–46. doi: 10.1177/001316446002000104

26. Davis KF, Parker KP, Montgomery GL. Sleep in infants and young children: part one: normal sleep. *J Pediatr Heal Care*. (2004) 18(2):65–71. doi: 10.1016/S0891-5245(03)00149-4

27. Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol*. (1933) 24(6):417–41. doi: 10.1037/h0071325

28. Liang Z, Chapa-Martell MA. A multi-level classification approach for sleep stage prediction with processed data derived from consumer wearable activity trackers. *Front Digit Heal*. (2021) 3:665946. doi: 10.3389/fdgth.2021.665946

29. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: respiratory events. *J Clin Sleep Med*. (2014) 10 (4):447–54. doi: 10.5664/jcsm.3630

30. Younes M, Raneri J, Hanly P. Staging sleep in polysomnograms: analysis of inter-scorer variability. *J Clin Sleep Med*. (2016) 12(6):885–94. doi: 10.5664/jcsm.5894

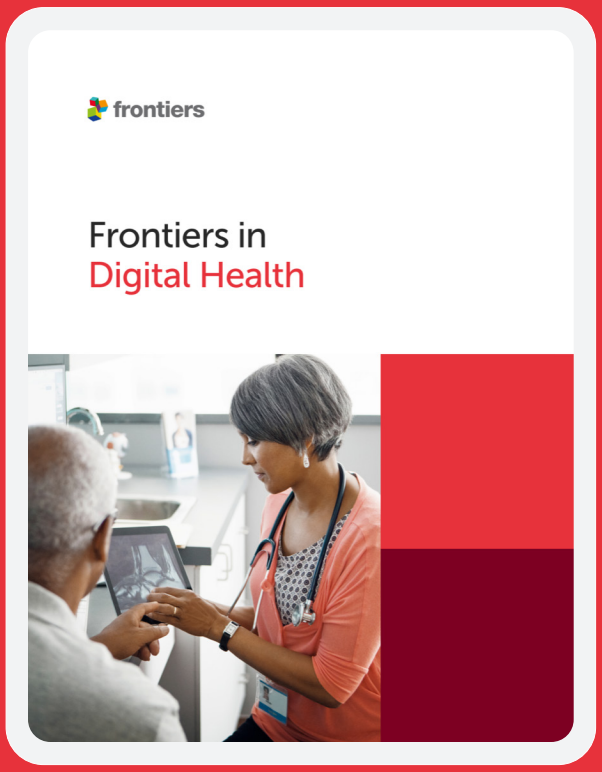
Frontiers in
Digital Health

Explores digital innovation to transform modern healthcare

A multidisciplinary journal that focuses on how we can transform healthcare with innovative digital tools. It provides a forum for an era of health service marked by increased prediction and prevention.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact