JMIR Biomedical Engineering

Engineering for health technologies, medical devices, and innovative medical treatments and procedures
Volume 8 (2023) ISSN 2561-3278 Editor in Chief: Syed A. A. Rizvi, MD, PhD, MBA, MPH, BSN

Contents

Review

Original Papers

Review

Mixed Reality Platforms in Telehealth Delivery: Scoping Review

Hemendra Worlikar[1], BEng, MSc; Sean Coleman[1,2], MB BCh BAO; Jack Kelly[1,2], MB BCh BAO; Sadhbh O'Connor[1,2], MB BCh BAO; Aoife Murray[1], MB BCh BAO; Terri McVeigh[3], MB BCh BAO, PhD; Jennifer Doran[1], BA, LLB, MB BCh BAO; Ian McCabe[1], BSc, MSc, PhD; Derek O'Keeffe[2,4,5], BEng, MB BCh BAO, MEng, MBA, PhD, MD

[1]Health Innovation Via Engineering Laboratory, Cúram Science Foundation Ireland Research Centre for Medical Devices, University of Galway, Galway, Ireland

[2]Department of Medicine, University Hospital Galway, Galway, Ireland

[3]Cancer Genetics Unit, The Royal Marsden National Health Service Foundation Trust, London, United Kingdom

[4]School of Medicine, College of Medicine Nursing and Health Sciences, University of Galway, Galway, Ireland

[5]Lero, Science Foundation Ireland Centre for Software Research, University of Limerick, Limerick, Ireland

Corresponding Author:
Hemendra Worlikar, BEng, MSc
Health Innovation Via Engineering Laboratory
Cúram Science Foundation Ireland Research Centre for Medical Devices
University of Galway
Lambe Institute
Galway, H91 TK33
Ireland
Phone: 353 899413431
Email: worlikarhemendra09@gmail.com

Abstract

Background: The distinctive features of the digital reality platforms, namely augmented reality (AR), virtual reality (VR), and mixed reality (MR) have extended to medical education, training, simulation, and patient care. Furthermore, this digital reality technology seamlessly merges with information and communication technology creating an enriched telehealth ecosystem. This review provides a composite overview of the prospects of telehealth delivered using the MR platform in clinical settings.

Objective: This review identifies various clinical applications of high-fidelity digital display technology, namely AR, VR, and MR, delivered using telehealth capabilities. Next, the review focuses on the technical characteristics, hardware, and software technologies used in the composition of AR, VR, and MR in telehealth.

Methods: We conducted a scoping review using the methodological framework and reporting design using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Full-length articles in English were obtained from the Embase, PubMed, and Web of Science databases. The search protocol was based on the following keywords and Medical Subject Headings to obtain relevant results: "augmented reality," "virtual reality," "mixed-reality," "telemedicine," "telehealth," and "digital health." A predefined inclusion-exclusion criterion was developed in filtering the obtained results and the final selection of the articles, followed by data extraction and construction of the review.

Results: We identified 4407 articles, of which 320 were eligible for full-text screening. A total of 134 full-text articles were included in the review. Telerehabilitation, telementoring, teleconsultation, telemonitoring, telepsychiatry, telesurgery, and telediagnosis were the segments of the telehealth division that explored the use of AR, VR, and MR platforms. Telerehabilitation using VR was the most commonly recurring segment in the included studies. AR and MR has been mainly used for telementoring and teleconsultation. The most important technical features of digital reality technology to emerge with telehealth were virtual environment, exergaming, 3D avatars, telepresence, anchoring annotations, and first-person viewpoint. Different arrangements of technology—3D modeling and viewing tools, communication and streaming platforms, file transfer and sharing platforms, sensors, high-fidelity displays, and controllers—formed the basis of most systems.

Conclusions: This review constitutes a recent overview of the evolving digital AR and VR in various clinical applications using the telehealth setup. This combination of telehealth with AR, VR, and MR allows for remote facilitation of clinical expertise and further development of home-based treatment. This review explores the rapidly growing suite of technologies available to users within the digital health sector and examines the opportunities and challenges they present.

(JMIR Biomed Eng 2023;8:e42709) doi:10.2196/42709

KEYWORDS

augmented reality; virtual reality; mixed realities; telemedicine; eHealth; mobile health; mHealth

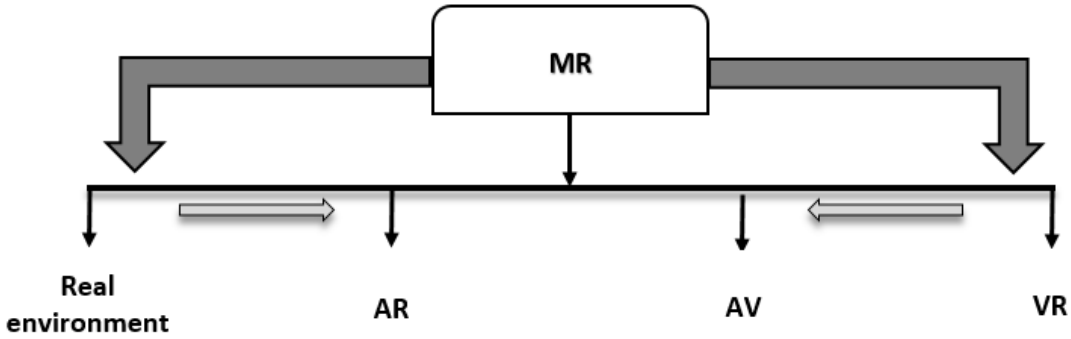
Introduction

Background

The term telemedicine refers to the provision of clinical health care services over a distance through information and communication technology (ICT) channels. Telemedicine overcomes geographical barriers in facilitating remote medical services. Building on this, the concept of telehealth extends to include continuing health education, research, and evaluation by medical professionals, all while promoting the health outcomes of individuals and communities [1]. Telehealth broadly encompasses the delivery of remote health-related services, including nonclinical services such as medical provider training; medical education; public health education; administrative meetings; and electronic exchange of clinical data enabling diagnosis, evaluation, consultation, treatment, and care management. The term telehealth has evolved as available technologies have improved, such that the term "digital health" is now often used as a more inclusive term reflecting the application of various different types of technologies and telecommunications systems in health care delivery. Digital health platforms can be either provider-to-provider or direct-to-consumer systems supported by the ICT infrastructure [2,3]. The telehealth sector has seen an effective increase in the past few years and has grown exponentially because of COVID-19 pandemic restrictions. According to the report published by Fortune Business Insights, the global telehealth market size was estimated at around US $144.38 billion in 2020 and is likely to reach US $636.38 billion by 2028 [4].

From the reality-virtuality continuum model, according to Milgram et al [5] (as seen in Figure 1), the real environment is that which is viewed without any overlay of the computer-generated entity, while at the opposite end of this continuum, immersive virtual reality (VR) is observed as completely enhanced computer-generated environments viewed through a head-mounted display unit. In the augmented reality (AR)–based display, digital information or entities are overlaid in the real environment, such that different aspects of reality are observed between the real and virtual environment. These augmentation-based realities can be discovered by optical see-through head-mounted displays (HMDs), mobile phones, tablets, or computer monitors [5].

Figure 1. Representation of reality-virtuality continuum by Milgram et al [5]. AR: augmented reality; AV: augmented virtuality; MR: mixed reality; VR: virtual reality.



In AR technology, the digitally created data directly coincide with the user's real-world environment, where the user can see the computer-generated 2D or 3D entities such as holograms. The virtual entities superimposed or mapped onto the real-world space are typically rendered using optical see-through display such as HMDs or mobile-based devices, also allowing for stereoscopic visualization. The next most advanced form of reality platform, the mixed reality (MR), follows the footstep of AR and allows interaction with these virtual entities by using hand gesture inputs, gaze recognition, or controllers. The VR platform is a completely enhanced digital representation featuring a 2D or 3D virtual environment or objects that can replicate real-life surroundings. VR provides engaging sensory perceptions for both visual and acoustic stimulation. Immersive VR relies on headsets or stand-alone VR devices, whereas nonimmersive VR relies on the monitor display [6].

The introduction of VR and AR technologies in medicine has been focused on clinical-related research. The key areas incorporating this digital reality are surgery, psychology, neurological condition, rehabilitation, and medical educational [7]. The 3D picturing capabilities of the VR- and AR-based platforms have been sought for applications in the visualization of scientific experimental imaging data, tools for surgical planning and studying anatomy, and other collaborative interfaces for education and telehealth [8]. Surgical simulation has distinctively used digital reality, while VR is principally used for visual and haptic rendering, whereas AR and MR were predominantly positioned for the tracking system and graphical rendering, with the latter being used in a real surgical setting [9]. The usefulness of VR education and training using simulation methods for nursing students was comparable with the standard models of education and training on the outcomes of skills, confidence, satisfaction, and performance time [10]. The current prospects of AR software applications in medical criteria are treatment and training based [11]. Surgical development using an MR platform has been linked as a predominant utilization tool for training and simulation

technology, advanced imaging and navigation, and broadening the extent of clinical application. Recently, MR has been adapted to neurosurgery, otolaryngology, ophthalmology, urology, and dentistry [12]. Digital reality technology has been incorporated into the preoperative surgical planning for several cranial-based applications for the neurosurgical subspecialty [13]. VR-based exposure therapy is used for various psychiatric disorders such as anxiety, trauma and stress, neurocognitive disorders, and several mental disorders. The effects of VR have been studied to have long-lasting positive outcomes for the treatment [14]. VR-based training has been effective in the improvement of executive limb function and cognitive function in patients with stroke [15,16].

Objectives

Many published studies have reviewed the use of AR and VR capabilities in medical research and practice and have not detailed its implication in telehealth, thus addressing this research gap. This systematic scoping review provides an overview of the prospects of AR and VR applications delivered using telehealth platforms in clinical settings. This review offers end users and providers an update of the current use of AR, VR, and MR effectively in telehealth delivery and highlights the prospects of such technologies in the future. This review aims to explore the following research questions:

- What clinical specialties have incorporated digital reality platforms such as AR, VR, or MR exclusively with telehealth?
- What are the different hardware and software technology formats used in AR, VR, or MR within telehealth?
- Which important technical features of AR and VR have been used in telehealth?

Methods

Overview

This scoping review used the framework of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [17]. Included studies from the database were solely concerned with the application of high-fidelity simulation technology such as AR, VR, or MR exclusively delivered via the telehealth platform. The study has no written or published protocol.

Database and Search Strategy

Articles from Embase, PubMed, and Web of Science were explored to obtain relative pieces of evidence. An exploded search strategy string was developed with the support of a university librarian. The search string included appropriate keywords and Medical Subject Headings terms—“augmented reality,” “virtual reality,” “mixed reality,” “extended reality,” “telemedicine,” “telehealth,” “m-Health,” “e-Health,” and “digital health.” The search strategy was initially developed on the Embase database and replicated across the other databases using predefined filtering techniques. The entire search strategy can be seen in Multimedia Appendix 1.

Eligibility Criteria

The studies included must satisfy the active component use of AR, VR, or MR delivered via telehealth approaches and should have been published between the years 2016 and 2021, since such devices with this technology format became commercially available, marked in reference to the release date for the first-generation Microsoft HoloLens [18]. The collaboration aspect of AR and MR technology into social or digital communication avenues could be observed during the same period [19]. Telemedicine or telehealth includes a broad spectrum of health care delivery, including education prospects; however, this review will focus on clinical aspects, including simulation. Only full-length text articles available on the web in the English language were included. Full-length text from peer-reviewed articles such as randomized controlled trials, feasibility studies, exploratory studies, narrative reviews, systematic reviews, case and cohort studies, book sections, and technical reports was considered eligible for inclusion. Any studies highlighting the mentioned technology for gaming, entertainment, or medical education were excluded. Correspondence papers, letters, conference abstracts (no full texts), editorial, commentary, poster presentations, and gray literature were also excluded from this review.

Study Selection and Data Extraction

The papers obtained from the applied search strategy from the information databases were imported to the reference manager EndNote 20 library, and duplicates were discarded [20]. Three researchers (HW, SC, and JK) performed initial screenings based on titles, abstracts, and keyword searches. Author HW conducted eligibility criteria and full-text screening. The selected studies were then reviewed based on the article type, study design, clinical condition addressed in the study, mode of telehealth communication, acceptance criteria, and the hardware and software used in the studies for the guidance for data synthesis. Finally, the relevant information from the studies was tabulated into an Excel (Microsoft Corp) spreadsheet, and a descriptive synthesis of the data was generated. In our review, we summarized and grouped the various telehealth branches using digital reality platforms for the various clinical condition based on descriptive statistical findings for the included studies. Different facets of the digital reality technology were detailed for its application in clinical research.

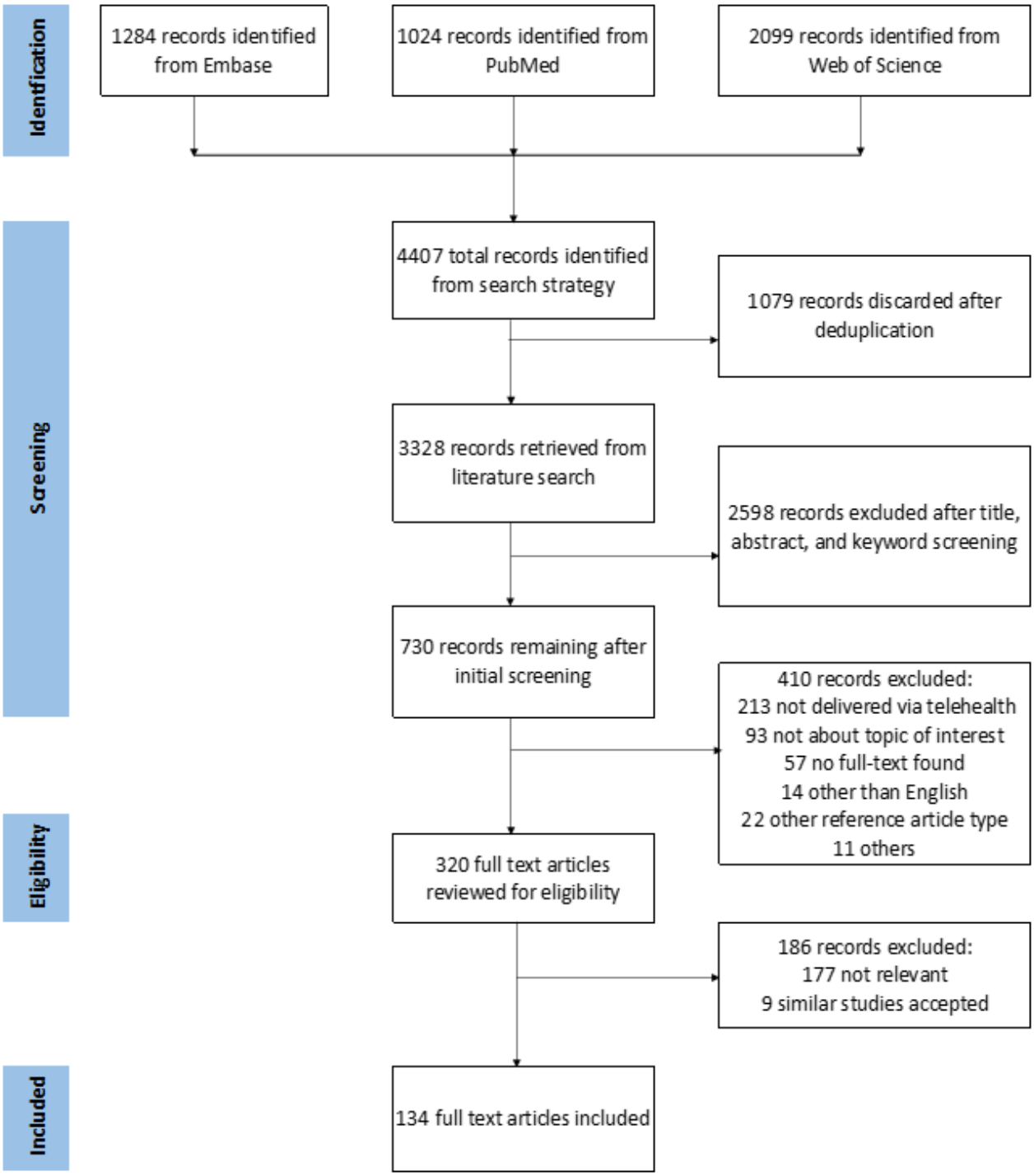
Results

Overview

Of the 4407 abstracts identified from the search protocol, 134 full-text articles fulfilled the inclusion criteria. A total of 1079 duplicate records were removed, 2598 records were discarded after title, abstract, and keyword search, and 410 records were deemed not fit after the initial screening as these articles were not about topic of interest having objectives that did not align with the outcomes of this review and did not satisfy the inclusion criteria. Of the 320 articles that were subjected to full-text review, 177 articles were deemed not relevant because they either included the digital reality technology or telehealth strategies but not delivered jointly, and 9 were excluded after

recognizing multiple papers published on the same topic by the HW, SC, and JK (Figure 2).

Figure 2. Flowchart for the structured literature search and selection.



Digital Reality Platform via Telehealth

As demonstrated in Figure 3, VR and AR cover most of the listed telehealth domains for the eligible studies. The most studied and researched area is telerehabilitation accomplished using VR. The other subareas involving VR use include telepsychiatry for evaluation and treatment, telediagnosis, and teleconsultation. In addition, AR and MR are prevalent modes of the reality technology platform for telementoring and teleconsultation. Finally, telesurgery and telemonitoring are the 2 subfields of telehealth where AR technology have seen an upward trend.

Clinically based digital health applications were considered for the review as various specific branches of the telehealth spectrum (Figure 4). *Telerehabilitation* is a postclinical care service delivered at home or remotely for recovery purposes and constitutes most of the included telehealth group from the included studies [21,22]. Evident from the included studies, stroke rehabilitation emerges as the leading medical condition

that has seen an uptake of these services. Different aspects of rehabilitation, such as functional motor training, including upper-extremity training and fine motor skills, cognitive functional training, visuomotor tracking training, and balance and gait training are primarily used for treating poststroke survivors [23-41]. In turn, the patient groups who have used telehealth for the purposes of rehabilitation have reported improvements in their quality of life, increased daily activities, and improved levels of motivation [42]. From the multiple studies included, telerehabilitation has been experimented as a home-based treatment for various neurological and cognitive disorders or diseases such as Parkinson disease, acquired brain injuries, multiple sclerosis, cerebral palsy, mild cognitive impairment, Alzheimer disease, and dementia [43-58].

Conventional therapy programs in the form of physical therapy and behavioral therapy are the nonpharmacological treatments that have used this remote delivery platform. In a small number of studies, the home-based rehabilitation in the form of novel telerehabilitation have been used for patients undergoing surgical procedures, such as total hip replacement, total knee arthroplasty, and total knee replacement, as a postrecovery treatment measure [59-62]. Mirror therapy for patients with phantom limb pain and physiotherapy treatment for patients with chronic body pain have incorporated this model of remote teletherapy [63-68]. This field has also been applied in physical rehabilitation for musculoskeletal disorders, provision of vestibular rehabilitation therapy in patients with a balance disorder, and kinesiotherapy for older adults at risk of falls [2,69,70]. Physical therapy in the pediatric group and musical therapy in patients with spinal cord injury have explored this stream of technology [71,72]. *Pulmonary rehabilitation therapy* for respiratory disorders such as chronic pulmonary respiratory disorder, pulmonary fibrosis, and myocardial infarction; *low vision rehabilitation* in providing functional visual assistance; and the *COVID-19 pandemic* have been an influential factor in accelerating remote rehabilitation therapy [22,73-77].

Telementoring is a subset of telemedicine that reflects remote expert guidance such as training or telenavigation to medical and nonmedical personnel in performance of life-sustaining procedures [78]. The impact and usability of the telementoring technique in provision of cardiopulmonary resuscitation in treating cardiac arrest has been demonstrated by different authors in simulated environments, with the assistance of a remote mentor using an HMD or Google Glass [79,80]. Other authors have explored the use of telementoring guidance, in intraoperative telenavigation, and preoperative planning in simulated battlefield and emergency trauma. The telementoring approach for preoperative planning and telenavigation during the intraoperative process has been demonstrated in complex emergency hand reconstruction surgery [81,82]. Forward damage control procedure performed on a patient-simulator model depicting a right-sided femoral gunshot wound and simulated trauma injuries such as airway obstruction by conducting cricothyroidotomy have been carried out using remote instruction—as have, lung decompression, tracheostomy, or REBOA (resuscitative endovascular balloon occlusion of the aorta) catheter deployment to deal with specific trauma injuries with the aid of a remote medical expert [82-86]. The feasibility of telementoring applicability in the performance of chest thoracotomy, skin grafting, and fasciotomy has been evaluated using ex vivo animal models [78,87,88]. Telementoring has been used to great effect in different stages of surgical planning in various orthopedic, craniofacial, spinal cord, vascular, and cardiothoracic surgeries [6,89-100].

Figure 3. Collaboration between digital reality technology and telehealth for the included studies.

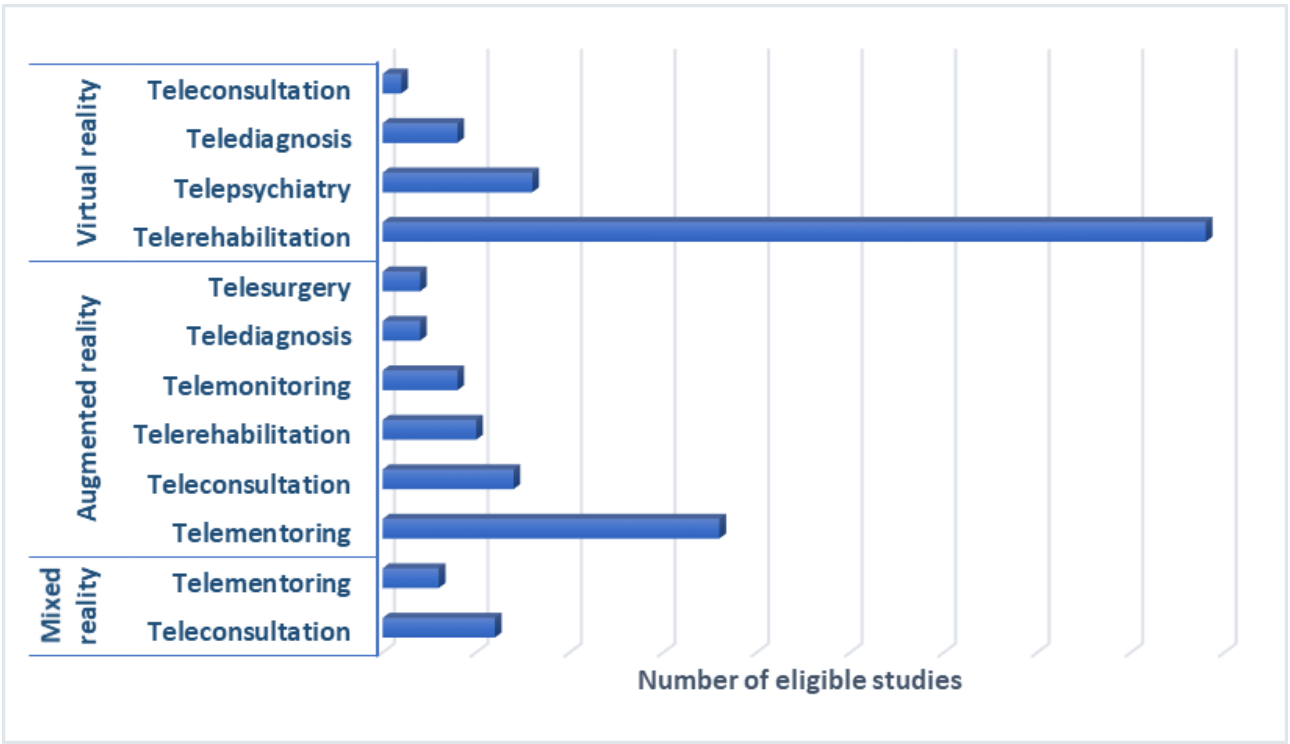
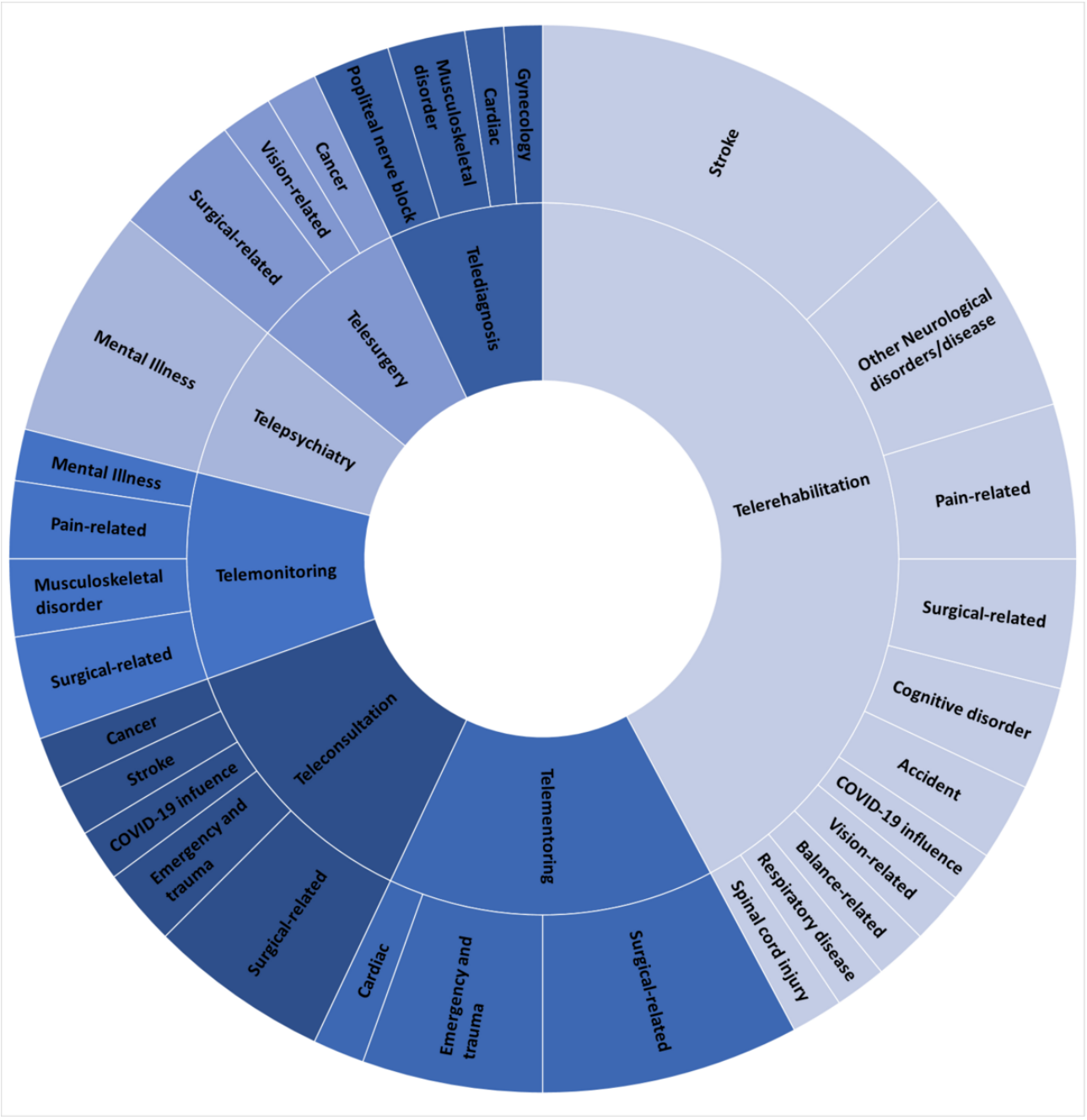


Figure 4. An overview of different clinical conditions and groups categorized within specific telehealth domains.



Teleconsultation is a primary segment of telehealth services, broadly consisting of remote consultation services using ICT. This remote consultation can be synchronous or asynchronous and between clinicians (provider-to-provider) for shared decision-making or between clinicians and their patients (provider-patient) [101]. This approach has been applied in patient assessment using the National Institute of Health Stroke Scale for patients with acute stroke and during remote clinical rounds in isolation wards for patients with COVID-19, thereby reducing direct exposure of the staff [102,103]. This technique has also been evaluated in trauma and emergency-related scenarios, such as remote consultation in reading and interpreting electrocardiogram reports related to drug intoxication or poisoning [104,105]. The effectiveness of provider-to-provider teleconsultation has been demonstrated in provision of support for ambulatory staff and first-responders in triage during simulation of major trauma [106]. The applications of teleconsultation in provision of surgical care are broad, allowing collaborative, contextual, and presurgical planning and visualization and intraoperative surgical navigation through high-fidelity immersive reality platforms and devices, as well as facilitating remote delivery of complex information to patients [107-115]. Teleconsultation via the reality platforms has been used to explore the feasibility of telepathology in carrying out an autopsy, image scanning, and transfer of serially sectioned cancer tissue from a mouse [116].

Telemonitoring is an advanced form of clinical care service that provides patient-centered care. This method allows health care providers to collect and track patient information and deliver remote care assistance [117]. This branch of telehealth has been evaluated in pediatric cohorts dealing with hospital-induced stress as a shared experience on a mobile-based AR game for

play therapy. This aspect allows managing pediatric patient profiles, data collection, and further analysis for effective treatment [118]. Telemonitoring via holographic conversational agents; that is, a computer-generated character to deliver physiotherapy home exercises to patients with musculoskeletal disorders and chronic pain has been demonstrated to increase their treatment adherence [119]. Supervised AR-based home training has been used for patients with phantom limb pain by providing mirror therapy, thereby promoting visuomotor integration by reengaging the neural circuits related to lost limbs [120]. Telemonitoring has been used in postoperative care and wound assessment in orthopedic and neurosurgical cases and has also been applied for *teleproctoring* or *remote monitoring* in pilot simulation as training for fundamentals of laparoscopic surgery examination [121-123].

Telepsychiatry uses ICT to offer a range of clinical and nonclinical services such as psychiatric evaluation, therapy (individual or group-based), patient education, and management remotely [124]. Studies using this element of telehealth and computer-generated virtual environments have evaluated the feasibility of remote therapy such as Virtual Reality Exposure Therapy for patients with acrophobia and evaluation of the technical system in delivering specific phobia treatment for arachnophobia [125,126]. Remotely delivered psychological treatment by the mental health professional include behavioral intervention therapy, cognitive behavioral therapy, mindfulness therapy, and acceptance and commitment therapy for patients facing stress, anxiety, public speaking anxiety, and social anxiety disorder, among others [127-130]. In a simulation study, cognitive and affective assessment of astronauts has been carried out to characterize social isolation from space [131]. Evaluation of telepsychiatry using the reality platform such as VR versus the traditional videoconferencing platform, and the development of newer platforms such as social VR for older adults in urban areas has demonstrated such techniques could lead to improved quality of life by reducing social isolation [132]. Telepsychiatry assessment via VR as a home-based treatment delivered by mental health professionals, such as a psychiatrist, psychologist, licensed social worker, or a mental health counselor, has been demonstrated to mitigate clinician burnout [133,134].

Another exciting subsection of the telehealth sphere, *telesurgery*, enables teleoperation in an operating field executed over a distance. Telesurgery involves using various disciplines such as communication technology, imaging techniques, motor control systems, robotics, reality platforms, and digital signal processing [135,136]. For example, in an experimental setup, a VR-based teleoperative system consisting of a robotic catheter operating system can be used to imitate vascular interventional surgery for arterial aneurysms or other vascular diseases. This method allowed unskilled surgeons to train in essential catheter guidance skills and enabled experienced physicians to conduct surgeries cooperatively [137]. In addition, a telesurgical experiment was conducted with a tendon-driven continuum robot via telenavigation for endoscopic and minimally invasive surgical procedures by tracking coordinate trajectory registration [138]. Finally, in another simulation case, a magnetically driven endoscopic capsule enabled the teleoperator or user to receive visual feedback in VR to conduct capsule endoscopy for colorectal cancer [139].

Moreover, the reality platform is streamed as a functional stereoscopic display and navigates space during telesurgery. This aspect of telesurgery has been experimented with as a visualization opportunity using smartphone-delivered vision and VR headsets to perform microsurgery for cataract and phacoemulsification [140]. In addition, Stereoscopic AR Predictive Display using the da Vinci R Surgical System to perform laparoscopic surgery and AR-assisted robotic surgery for kidney transplant procedures are some of the current practical applications of telesurgery [141-143]. *Telediagnosis* refers to the detection or evaluation of a disease or condition using telematics technology. It is achieved remotely while the patient is at a local site with remote diagnostic tools and devices [144,145]. For instance, in experimental analysis, locating and evaluating tumor-bearing hysteromyoma coordinates during a 3D navigated gynecological operation facilitates telediagnosis when visualized on a 3D user interface of the medical record [146]. Another study proposes a framework based on bidirectional haptic feedback and tele immersion in the evaluation of range of motion and maximum isometric strength using the 10 arm movements method in the diagnosis of musculoskeletal disorders, poststroke rehabilitation, or postshoulder surgery [147]. Ultrasonography (USG) is a field in which telediagnosis using the high-fidelity visualization system has been used to great effect. Evaluation of 3D VR telenavigation in cardiac USG has been undertaken in simulated settings. The added benefits of AR enable real-time teleguidance on procedural performance and image registration for point-of-focus ultrasonography (POCUS) and foveated imaging pipeline in extending VR-based telediagnosis [148-150]. Another study mentions AR video communication projected by mobile-based AR guidance to conduct POCUS on popliteal nerve block and a subsequent diagnosis based on the availed health information [151].

Overview of the Hardware and Software Units for the Included Studies

To experience MR, high-simulation visualization hardware devices and some of the commercial ones included in the selected studies are listed in Multimedia Appendix 2. These include high-end AR and VR devices, smart glasses, mobile devices, standard LED (light emitting diode) and LCD (liquid crystal display) television or display screen, 3D television, and 3D projectors. The commonly included immersive reality-capable devices are mostly wearable technology such as smart glasses, VR or AR HMDs, and nonimmersive standard display units. However, these high-fidelity simulation display technologies form the final part of any system and are primarily used in combination with optical capturing and tracking devices and input devices. The optical capturing and tracking systems or devices incorporate 3D depth and color-sensing camera sensors. The input devices such as controllers, trackers, or customized input modules help navigate the immediate VR or any MR environment. Various studies have included the VR gaming element in their rehabilitation programs, with some having their own developed VR rehabilitation system. Most included studies have used biometric devices for specific

medical parameter evaluation to draw analysis and simulation models to conduct various training. Other relative hardware devices and systems that have been used are listed in the Multimedia Appendix 3.

The graphical representation for any software harbors a visualization platform, and more specifically, the MR system incorporates contextual 3D figures and scenes. The various software applications and source platforms that were featured and used in the included studies are listed as a table in the Multimedia Appendix 4. These applications are grouped as 3D modeling and visualization software, communication and streaming software, file-sharing and transfer applications, and other specific and personalized software applications. The 3D composite images and environment for the MR technology are created using the computer graphic designing software and gaming engine platforms. Processing and accessing the 3D computer-generated environment or images needs specific and compatible visualization file applications supported by the device. The featured 3D modeling and viewing application allows for creating and editing static and interactive multidimensional models and VR scenes, animations, and games, conversion of the produced scanned images to computer-aided design models, and stereoscopic 3D display content. The telehealth domain explores the ICT for effective remote clinical services while using the streaming facilities offered by various low-bandwidth platforms. This domain allows offline or real-time interactive communication and collaboration for any dedicated clinical services. Many communication and streaming applications allow for remote one-to-one or group video calls and messaging, screen sharing, file sharing, hosting channels, and video broadcasting. Some of these platforms allow for direct AR and VR integration and acceptance. The file-sharing and other specific applications are synchronously and explicitly used as a sequential fragment of the entire system. The developed software from the studies mentioned in the table encapsulates the combination of AR and VR cooperatively with the remote telehealth applications.

Virtual Environment

A virtual environment (VE) recreates a coordinated appearance of sensory information representing that of a physical environment that can be unreal, interactive, or wholly imagined environment perceived when the user wears an appropriate gadget [152]. In addition, the term *virtualworlds* has been interchangeably used with a VE. Developing this state-of-the-art perceived environment is created using a subset of tools arising from computer game technology, specifically through commercial game engines. The scene can be a 2D or 3D illustration, which is a complex and time-consuming process for its creation [153]. This element of VE has been recreated in almost every aspect using the VR platform. For example, the study by Levy et al [125] demonstrated the use of virtual worlds such as a subway station and a 24-story high-rise building as background scenes to overcome acrophobia as a VR exposure therapy. Similarly, Cikajlo et al [127] developed a program called ReCoVR (Realizing Collaborative Virtual Reality for Well-being and Self-Healing). The participant attends a remote guided mindfulness program as part of a group. This mindfulness program was organized as 360° video scenes where they carried out different tasks and exercises. Initially, all the participants that joined were seated in the virtual fireplace room; upon the program's progression, they were switched to other 3D VEs, such as the Dooney Rock, River Bonnet, or the mountain-view. Shao and Lee [132] have addressed a social VR platform that uses the 3D scenes in the VE for real-time face-to-face communication in different distant locations to learn about its value and urgency in the urban older adult population. Tamplin et al [72] developed a web-based music therapy telehealth platform using social VR, vTIME (vTime Limited), allowing group music therapy sessions in VE, such as singing around a campfire in a forest.

Gaming-Based VE

Moreover, many studies used VE in interactive game–based settings for rehabilitative exercise programs. In a program described by Meca-Lallana et al [53], patients were required to carry out specific tasks to accomplish a mission in 2 different scenes: a medieval fantasy world and a deserted island. Yet again, in another exercise setup, VR exercises depict a wooden church in Hrabova Roztoka. The patient explores this particular place using a VR headset, thereby facilitating lower-limb rehabilitation [36]. Telerehab VR, a custom-built application program that runs on either a mobile-based tablet or PC, was developed using the game engine Unity (Unity Technologies Inc). This system provides upper-limb rehabilitation for patients with multiple sclerosis. They perform various activities of daily living tasks happening in the VE in a realistic home setting. A leap motion controller (Ultraleap) was used to track and control the hand motion executed while performing the gaming tasks [48].

Telepresence

Telepresence describes the characteristic of directly interacting with the actual physical state, experienced from the first-person viewpoint of the user located remotely [154]. Tian et al [147] used the H-TIME (Haptic Enable Tele-Immersion Musculoskeletal Examination) set up at both the patient and doctor ends to conduct a remote diagnosis of musculoskeletal examination. At both sites, the doctor and patient could feel each other's movements because of the bidirectional force feedback mechanism. They could view and communicate with each other in the VE, bringing them to the same examination room virtually. In another instance, in treating phobia, in particular, fear of spiders, the patients were allowed to interact in the VE, where the therapist gradually added the feared creature to the scene. This treatment is performed remotely via the tactile internet with VR headsets or standard computer screens using a hand-tracking and haptic device such as a glove [126].

Teleoperation refers to performing designated highly skilled manual tasks remotely, similar to a telerobotic medical system in minimally invasive surgery [155]. In a simulated study, an endoscopist performs a teleoperation process using a haptic device that controls the position of an external permanent magnet positioned at the end of a robotic arm. The user is wearing a VR headset and receives the corresponding visual information from the camera of the endoscopic capsule and then proceeds with the navigation process inside the colon [139].

Another simulated study used teleoperated ultrasonography that builds on the VE developed as a 3D representation of a real USG probe and a mannequin imitating a patient's body highlighted with a geometric mesh for the purpose of following the examination. The user wears an Oculus Rift DK2 HMD (Oculus) to perform this simulation of tele-USG [148]. Syawaludin et al [150] introduced the use of 360° foveated pipeline imaging viewed via an HMD. The image or video capture is facilitated by the use of an omnidirectional pantilt-zoom camera module, and the remote physician can remotely diagnose the wound by zooming in and inspecting it in a 360° view over the HMD.

Exergaming and Serious Gaming via VR

In the context of virtual telerehabilitation, *exergaming* and *seriousgames* are the 2 most popular applications that emerge. Simply put, exergaming is an activity connected with playing video games that involve physical exercise [70]. In contrast, serious games follow the objective of games, implicitly focusing on increasing skills and abilities and gaining experience and knowledge [156]. The TELEKIN system, a beta edition, uses the interface of the serious game to rehabilitate cognitive and musculoskeletal disorders using a web-based framework [43]. The training sessions are conducted and played in a 3D VE that includes a number of games. Two of them involve physical actions—flexion and extension of wrist, hip, or shoulder as the doctor chooses to control the ball and platform in executing the game. Another game focuses on cognitive rehabilitation by featuring random words that the user must arrange them to form a sentence, which can be achieved using gesture-based controls [43]. Gandolfi et al [44] used the Tele Wii Lab platform as a home-based balance training, and Sheehy et al [28] and Allegue et al [30] used the Jintronix system (Jintronix, Inc) in upper-extremity rehabilitation of chronic poststroke patients, which used the exergaming platform. The study by Triandafilou et al [24] that developed a networked multiuser gaming format, Virtual Environment for Rehabilitative Gaming Exercise (VERGE), conducted a feasibility trial to determine the effectiveness of this developed system with other potential home treatments. The VERGE system features a set of 3 exercises, namely Ball Bump, where the users pass the ball back and forth across the table; Food Fight, where the users in multiplayer can pick up the food on the table and throw it at each other; and the Trajectory Trace game, where 1 player draws a trajectory path in the space while another player retraces the trajectory to erase it. Burdea et al [32] included a commercial rehabilitation system with a novel therapeutic game controller, BrightBrainer (Bright Cloud International Corp). This system offers a multitude of interactive games (Breakout 3D and Card Island Towers of Hanoi, among others), training motor, cognitive, and executive functions for chronic poststroke patients. Qiu et al [35] demonstrated the feasibility of a home-based VR system that features 12 developed games focusing on the elbow-shoulder, hand, wrist, and entire arm for upper-extremity rehabilitation in poststroke patients (finger games: car, bowling, and piano; hand games: piano and fruit picking; wrist games: Wakamole and wrist flying; and finally, the shoulder-elbow games: the Maze, Arm Flying, Brick Break, and soccer goalie)

Avatar Representations or Virtual Agents

The term *avatar* is a distinguishable digital characterization of a human form (either specific or random) [128]. Moreover, these avatars can be either in 2D or 3D illustration, representing a specific part of the body, usually arms or an entire body structure with particular facial expressions. 3D avatars have been a central representation in the scope of VR and AR. The study by Anton et al [59] implements the Kinect-based Telerehabilitation (ie, KiRES) interface, providing two 3D avatars to guide the patient during their physical therapy session. One of the avatars represents the remote therapist and represents the local user or the patient, colored red and blue, respectively, so the patient can follow and perform the exercises executed by the 3D remote therapist avatar (in red). The patient can see their movements reflected by their blue avatar changing their positions as per the scenes from the therapy. In the study trial conducted by Jung et al [74], in a telerehabilitation program—Pulmonary Rehabilitation in Virtual Reality (PR in VR) program—each patient was provided with a VR headset, pico G2 4k (Pico Immersive Pte Ltd), preloaded with the PR in VR application. This application contains education and rehabilitation modules, and the chronic obstructive pulmonary disorder rehabilitation module comprises several physical exercises directed by a virtual instructor in 3D avatar embodiment.

The REWIRE autonomous telerehabilitation platform offers home-based intensive rehabilitation as offline remote monitoring by hospital clinicians. This system features a virtual therapist with artificial intelligence implanted and provides real-time feedback to maintain correct posture. In addition, the exercises performed by the patients are showcased as a 3D avatar on the screen. This intelligent system highlights each body segment of the exercise in a different color, intense green for the proper posture and red color for the incorrect posture [23]. The VERGE system enables the use of avatars to control and manipulate objects in the virtual gaming environment, allowing the capability to include multiple avatars and different users to manipulate the same object [24]. In the social VR app, vTIME (vTime Limited), an avatar persona is used for self-characterization to communicate in an immersive VE [128]. Afyouni et al [65] describe the use of RehaBot, a virtual assistant that illustrates to the user how to perform the exercises correctly (both the therapist and patient can replay the session in a 3D avatar). The RehaBot embeds real-time pattern and gesture recognition together with a dynamic correction module that considers the game difficulty level and reading from the virtual assistant to produce a tailored set of exercises that are rather fitting to the patient's native abilities.

Telestration and Annotations via AR and MR

Telestration enables the drawing of freehand representations, also known as annotations (such as lines, circles, or any other symbols or sketches) over any image or video feed. With the latest AR and MR technology, this telestration can be achieved in 2D and 3D and superimpose this annotation in the live video streaming during the video call [157]. The Virtual Interactive Presence and AR tool is a mobile or tablet-based augmented reality platform running on an iPad device (Apple). It

incorporates the telestration feature, allowing the remote expert surgeon to freeze the screen and then draw an image using a 2D pen tool. This composite video feed, viewed on both the local and remote stations, enables intraoperative telecollaboration in real-time [90]. In the feasibility study by Wang et al [100], POCUS using the HoloLens was conducted by the trainee in a simulated teleconference session. The MR capture video from the trainee was broadcasted. Live guidance provided by the expert mentor facilitates the trainee to complete a right quadrant Focussed Assessment using Sonography in Trauma examination. The broadcasting was achieved using VSee, a proprietary low-bandwidth, group videoconferencing and screen-sharing application. To perform complex hand reconstruction of a patient after a bomb-blast injury, a telementoring network was established between an expert surgeon in Lebanon and a local surgeon in Gaza. This session was hosted using a cloud-based AR platform Proximie (Proximie Limited), allowing the remote surgeon to superimpose their own hands or range of annotations and drawing tools into the virtual surgical field [82].

Mitsuno et al [93] demonstrated telementoring in a simulation study to perform craniofacial surgery by using a teleconferencing setup, Skype (Microsoft Corp) for HoloLens, enabling the telestrated features and images overlaid on the receivers' visual field. A POCUS examination was performed using a novel smartphone app Vuforia Chalk (PTC Inc), an AR video platform for remote AR assistance, anchoring the AR annotations in each other's supposed visual environment [151]. In the study, Ritcher et al [141] proposed the first predictive display with AR registration and rendering using stereoscopic displays designed for teleoperated surgical robots known as Stereoscopic AR Predictive Display. The simulation study measured the effectiveness of Stereoscopic AR Predictive Display conducted on the da Vinci R Surgical System (Intuitive Surgical) to complete the peg transfer task. The System for Telementoring with AR (STAR) platform now combines optical see-through display, HoloLens AR HMD. Similarly, this system allows for telementoring guidance by overlaying 3D graphical annotations onto the mentee's view of the surgical field, which remains anchored in the same place even after the mentee moves their head position [95].

The experiment by Zhang et al [110] aimed to enhance teleconsultation by using the AR technology ARkit (Apple) to create an immersive replica of the consultant. Using a Kinect sensor (Microsoft Corp) to capture the skeletal feature points of the consultant, the patient views a 3D dynamic virtual avatar doctor appearing in the patient's telepresence environment on their iPad device. A qualitative study was conducted to gain the experience and perception of AR Glasses in patients with pulmonary disorders for home-based telerehabilitation. The web-based telerehabilitation system Optimov (Optimov) enabled via an AR Glasses device Laster WAV headset (Laster Technologies) provides exercise coaching using a 3D virtual agent [73]. A holographic virtual therapist was deployed in the HOLOBALANCE, a novel health care platform for providing vestibular rehabilitation therapy for patients with balance disorders [158]. In the design and evaluation user study by Kowatsch et al [119], a hybrid ubiquitous coaching model relying on mobile and holographic conversation agents was introduced. The 3D virtual conversation agent demonstrated the squat exercise, engaging in real-time audio feedback for counting the repetition or providing automatic error detection for incorrectly or incompletely following the exercise. An innovative 3D point tracking module and unique AR system integrated with the HoloLens was used for surgical applications using telementoring. This module allowed for real-time 3D position tracking of the virtual scalpel handled by an experienced surgeon remotely. The inexperienced trainee wearing the HoloLens can see the surgical annotation superimposed with the actual surgical scene; the virtual path coregistered on the phantom arm model [78]. Next-generation mobile-based AR games for pediatric health care allow shared experiences with multiple other AR-supported devices to detect and interact using the same local area network. Several games were developed using the Unity game engine and ARCore Unity, a software development kit for Android operating software. Jungle Adventure, Map explorer, and Wakamole implemented AR interaction, whereas Map explorer and Wakamole particularly enabled the inclusion of a 2-player for a shared collaborative experience [118].

First-Person View for AR Capture Video Feed

Noorian et al [102] demonstrated smart reality glasses to conduct remote consultation using the National Institutes of Health Stroke Scale scores for stroke assessment. The onsite doctor wears the reality glass, Google glass. This Google glass is embedded with the Xpert Eye platform (AMA XpertEye), capable of assisted reality, allowing the person wearing this device to share their field of view in a 2-way real-time videoconferencing. Similarly, Nikouline et al [123] presented a feasibility study using the Google glass live video stream coming from the onsite proctor and the participant tasks related to fundamentals of Laparoscopy for scoring and evaluation done by the remote proctor. In their experiment, Lin et al [84] implement projective video texture-mapping that supplements a robust high-level stabilization video feed obtained from the mentee's first-person view. This effective format provides the remote expert with an effective workspace visualization, allowing seamless integration of annotations in an effective AR surgical telementoring. The prospective observational study by Martin et al [103] uses HoloLens 2 MR device to conduct remote clinical consultation in a COVID-19 ward. A senior staff member would enter the COVID-19 ward to undertake clinical rounds, and the other staff members of the staff team would join virtually, thereby minimizing exposure and infection transmission. Dynamic 365 Remote Assist (Microsoft Corp) software allowed for bidirectional audio and video functionality through which the remote staff team could see the first-person view from the HoloLens 2 device. In addition, this platform allowed to place relevant imaging and electronic health record data in the user field of view, improving situational awareness and better clinical decision-making. Finally, it significantly reduced the risk of direct viral transmission.

Web- and Cloud-Based Telehealth Delivery Modes

As digital communication network and services evolve, these are rapidly being adopted in health care delivery. The web- and

cloud-based applications have become prevalent in telehealth. Telehealth relies on the backbone of internet infrastructure supported using various broadband connections such as digital subscriber lines, fiber broadband, and wireless connection, including fixed wireless broadband, cellular network or mobile broadband, and satellite communication. Thus, ICT has become central to offering an array of digital health solutions such as real-time audio and videoconferencing, remote patient monitoring, store and forward technologies, and mobile health, among others [159,160]. A web-based application principally operates on the webserver. It is accessed through a web browser over an internet connection, whereas cloud-based applications operate similarly to web applications, operating on either or both the client and server sides [161]. The custom-developed systems KiRES [59] and STAR [85] rely on the WebRTC framework, an open-source application programing interface allowing for real-time audio-video and multimedia connection. In addition, the study by da Silva et al [55], included a web-based gaming application MoveHero, to evaluate the feasibility of home-based nonimmersive serious games in patients with cerebral palsy.

Interestingly, any virtual web-based application feeds on the information; thus, data storage and hosting become integral to all online services. The study by Kato et al [42], adopted the cloud-based storage and file hosting service Dropbox (Dropbox Inc) for collecting the spatial coordinate data for each joint using the 3D optical camera during the VR telerehabilitation. The proof-of-concept study by Sirilak et al [107], implemented an e-consultation system based on the AR and MR systems using the HoloLens device for remote consultancy services in the intensive care environment. This e-consultation platform depended on a cloud-based data center that performed as an information exchange and provided services for the end devices. It also consisted of body area network technology to integrate the vital physiological information from different client devices to the data center. Prvu Bettger et al [62], used a virtual telehealth system—virtual exercise rehabilitation assistant or virtual exercise rehabilitation assistant (Reflexion Health, Inc)—for posthospital care for total knee arthroplasty, Tsiouris et al [2] included a custom-developed platform HOLOBALANCE system in managing balance disorders, both using the technology-forward cloud-based platform.

Discussion

Principal Findings

This scoping review explores state-of-the-art extended reality platforms and telehealth solutions used in the clinical context. This review highlights the reported evidence-based practical and probable applications of the extended and MR platform with telehealth used in different clinical specialties. This review also addresses the technical characteristics of the AR and VR features used in telehealth services, including various hardware and software arrangements.

Stroke is the leading clinical condition incorporating telerehabilitation, a segment of the telehealth service and digital VR [23-41]. Approximately half of the included studies from the search strategy feature the use of telerehabilitation (Figure 4). Other clinical conditions such as neurological or cognitive disorders, musculoskeletal conditions, and postsurgical recovery have also adopted telerehabilitation facilities in the home or remote settings to continue treatment. Telerehabilitation used technical attributes of exergaming and serious gaming in improving the motor and cognitive functional skills [43-58].

Other divisions of telehealth, namely telementoring, teleconsultation, and telemonitoring, have been more frequently used for surgical-related procedures, emergencies and trauma, and in several disaster simulation for disaster response and preparedness [162]. The AR and MR technologies are more prevalent with telementoring, teleconsultation, and telesurgery (Figure 3). Exposure therapy under telepsychiatry has used VR to give the patient a photorealistic experience of overcoming their pathological response to their fear [126].

Telestrated AR features through anchoring of annotations in real-time and space, performed remotely via various communication channels: a useful aid in telesurgery [81]. The technical features from the digital reality technology of VE, digital avatars, telestration or the 3D rendering of annotations, and first-person viewpoint have demonstrated telemedical capabilities. The web- and cloud-based applications have various potential uses across the web-based clinical sphere [110]. Most of the included studies relied on existing commercial high-fidelity simulation technological hardware devices such as head-mounted AR and VR displays. The study and software designs for most of the included studies were codeveloped by the respective research teams by using multiple supportive platforms as a direct requirement for the project objectives.

These novel reality technologies of AR, VR, and MR enable 3D visualization, thereby creating a visual sense and experience of high ecological validity [163]. This technology has been extended as a remote, home-based solution for patients, thereby enabling patient empowerment [164]. This technology is highly engaging and motivating from the patient responses to telerehabilitation, consequently necessitating initial patient training needs may become an arduous task to the facilitator [42,165]. Network connectivity, internet and server security concerns, and technological constraints are some of the most common pitfalls across several studies included in this review [99,166]. The lack of interoperability between the hardware and software platforms poses a significant challenge in realizing the potential of this technology [2]. The need for improved network infrastructure and scalability poses a challenge and target for telehealth services; however, there is a risk of complete network failure, which can affect the use of such systems in critical care applications [83]. Patient confidentiality is integral at any stage during electronic exchange of health-related data; thus, network and data security protection are crucial factors for accessing telehealth services and should be robustly adhered to the governing regulations [66,149,167].

The exploded search strategy captures a broad array of important clinical applications of this high-fidelity reality technology and telehealth facilities. This review presents a current road map and the prospects of digital reality technology and telehealth in the clinical space. The determining factors presented allow the readers and researchers to evaluate the relevance of this

technology and its subsequent uptake in the clinical health sector. The study protocol was not registered, the included studies were not classified for risk of bias assessment, and the general characterization for the included studies were not presented. In addition, the review only included studies available in the English language and no relevant additional pieces of information was considered from the gray literature. From the broad array of literature-based evidence, most of the included experimental studies were pilot feasibility studies with small sample sizes, leading to reporting bias.

Conclusions

This review uniquely details the current and potential applications of digital reality technologies such as VR and AR and telehealth solutions. The feasible and practical application of AR and VR in the digital clinical space has been explored, as well as the challenges this multiparty technology endures in effective implementation and adoption. This suite of technologies offers a collaborative experience among health care professionals and their patient community. The telehealth component with the high-fidelity digital reality allows for an immersive and integrative means for teleconsultation, telesurgical procedures, and telementoring among the medical peer-to-peer group allowing for effective decision-making and treatment approaches. The uptake of VR and exergaming in various telerehabilitation programs has opened new avenues to posttreatment measures. This essential application of telehealth enhances the traditional health care delivery approach by enabling remotely delivered clinical care and services and developing home-based treatment programs. Further validated studies are needed to evaluate the overall assessment of this trending technology, thereby leading to commercial pathways. A robust and secure communication infrastructure will improve the accessibility of telehealth capabilities and extend the interoperability of the digital reality platform allowing for a diverse digital health care ecosystem.

Acknowledgments

This review was supported by the Science Foundation Ireland under grants (13/RC/2073) and (13/RC/2094). The authors would like to thank the university librarian, Rosie Dunne, for providing feedback on the search strategy and helping in finding the appropriate databases to conduct the search in.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 29 KB - biomedeng_v8i1e42709_app1.docx]

Multimedia Appendix 2
Augmented, virtual, and mixed reality hardware devices.
[DOCX File , 22 KB - biomedeng_v8i1e42709_app2.docx]

Multimedia Appendix 3
Other relative hardware devices.
[DOCX File , 22 KB - biomedeng_v8i1e42709_app3.docx]

Multimedia Appendix 4
List of software and source platforms.
[DOCX File , 27 KB - biomedeng_v8i1e42709_app4.docx]

References

1. Anthony Jr B. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
2. Tsiouris KM, Gatsios D, Tsakanikas V, Pardalis AA, Kouris I, Androutsou T, et al. Designing interoperable telehealth platforms: bridging IoT devices with cloud infrastructures. Enterp Inf Syst 2020 Apr 30;14(8):1194-1218. [doi: 10.1080/17517575.2020.1759146]
3. Telehealth and Telemedicine. American Academy of Family Physicians. URL: https://www.aafp.org/about/policies/all/telehealth-telemedicine.html [accessed 2021-12-01]
4. Telehealth market revenue, growth. Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/industry-reports/toc/telehealth-market-101065# [accessed 2022-07-25]

5. Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: a class of displays on the reality-virtuality continuum. In: Proceedings of the 1994 SPIE Conference on Telemanipulator and Telepresence Technologies. 1994 Presented at: SPIE '94; October 31-November 4, 1994; Boston, MA, USA. [doi: 10.1117/12.197321]
6. Verhey JT, Haglin JM, Verhey EM, Hartigan DE. Virtual, augmented, and mixed reality applications in orthopedic surgery. Int J Med Robot 2020 Apr;16(2):e2067. [doi: 10.1002/rcs.2067] [Medline: 31867864]
7. Yeung AW, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]
8. Venkatesan M, Mohan H, Ryan JR, Schürch CM, Nolan GP, Frakes DH, et al. Virtual and augmented reality for biomedical applications. Cell Rep Med 2021 Jul 20;2(7):100348 [FREE Full text] [doi: 10.1016/j.xcrm.2021.100348] [Medline: 34337564]
9. Lungu AJ, Swinkels W, Claesen L, Tu P, Egger J, Chen X. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. Expert Rev Med Devices 2021 Jan;18(1):47-62. [doi: 10.1080/17434440.2021.1860750] [Medline: 33283563]
10. Chen FQ, Leng YF, Ge JF, Wang DW, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: meta-analysis. J Med Internet Res 2020 Sep 15;22(9):e18290 [FREE Full text] [doi: 10.2196/18290] [Medline: 32930664]
11. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]
12. Han J, Kang HJ, Kim M, Kwon GH. Mapping the intellectual structure of research on surgery with mixed reality: bibliometric network analysis (2000-2019). J Biomed Inform 2020 Sep;109:103516 [FREE Full text] [doi: 10.1016/j.jbi.2020.103516] [Medline: 32736125]
13. Dadario NB, Quinoa T, Khatri D, Boockvar J, Langer D, D'Amico RS. Examining the benefits of extended reality in neurosurgery: a systematic review. J Clin Neurosci 2021 Dec;94:41-53. [doi: 10.1016/j.jocn.2021.09.037] [Medline: 34863461]
14. Dellazizzo L, Potvin S, Luigi M, Dumais A. Evidence on virtual reality-based therapies for psychiatric disorders: meta-review of meta-analyses. J Med Internet Res 2020 Aug 19;22(8):e20889 [FREE Full text] [doi: 10.2196/20889] [Medline: 32812889]
15. Wu J, Zeng A, Chen Z, Wei Y, Huang K, Chen J, et al. Effects of virtual reality training on upper limb function and balance in stroke patients: systematic review and meta-meta-analysis. J Med Internet Res 2021 Oct 12;23(10):e31051 [FREE Full text] [doi: 10.2196/31051] [Medline: 34636735]
16. Zhang Q, Fu Y, Lu Y, Zhang Y, Huang Q, Yang Y, et al. Impact of virtual reality-based therapies on cognition and mental health of stroke patients: systematic review and meta-analysis. J Med Internet Res 2021 Nov 17;23(11):e31007 [FREE Full text] [doi: 10.2196/31007] [Medline: 34787571]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Gilbert B. Microsoft's ambitious HoloLens headset ships on March 30. Insider. 2016 Feb 29. URL: https://www.businessinsider.com/hololens-release-date-march-30-2016-2 [accessed 2022-10-30]
19. Kim K, Billinghurst M, Bruder G, Duh HB, Welch GF. Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008-2017). IEEE Trans Vis Comput Graph 2018 Nov;24(11):2947-2962. [doi: 10.1109/TVCG.2018.2868591] [Medline: 30188833]
20. Falconer J. Removing duplicates from an EndNote library. London School of Hygiene and Tropical Medicine. 2018 Dec 7. URL: https://blogs.lshtm.ac.uk/library/2018/12/07/removing-duplicates-from-an-endnote-library/ [accessed 2022-07-25]
21. Kairy D, Veras M, Archambault P, Hernandez A, Higgins J, Levin MF, et al. Maximizing post-stroke upper limb rehabilitation using a novel telerehabilitation interactive virtual reality system in the patient's home: study protocol of a randomized clinical trial. Contemp Clin Trials 2016 Mar;47:49-53. [doi: 10.1016/j.cct.2015.12.006] [Medline: 26655433]
22. Aloyuni S, Alharbi R, Kashoo F, Alqahtani M, Alanazi A, Alzhrani M, et al. Knowledge, attitude, and barriers to telerehabilitation-based physical therapy practice in Saudi Arabia. Healthcare (Basel) 2020 Nov 04;8(4):460 [FREE Full text] [doi: 10.3390/healthcare8040460] [Medline: 33158298]
23. Held JP, Ferrer B, Mainetti R, Steblin A, Hertler B, Moreno-Conde A, et al. Autonomous rehabilitation at stroke patients home for balance and gait: safety, usability and compliance of a virtual reality system. Eur J Phys Rehabil Med 2018 Aug;54(4):545-553 [FREE Full text] [doi: 10.23736/S1973-9087.17.04802-X] [Medline: 28949120]
24. Triandafilou KM, Tsoupikova D, Barry AJ, Thielbar KN, Stoykov N, Kamper DG. Development of a 3D, networked multi-user virtual reality environment for home therapy after stroke. J Neuroeng Rehabil 2018 Oct 05;15(1):88 [FREE Full text] [doi: 10.1186/s12984-018-0429-0] [Medline: 30290777]
25. Zhou RJ, Hondori HM, Khademi M, Cassidy JM, Wu KM, Yang DZ, et al. Predicting gains with visuospatial training after stroke using an EEG measure of frontoparietal circuit function. Front Neurol 2018 Jul 24;9:597 [FREE Full text] [doi: 10.3389/fneur.2018.00597] [Medline: 30087653]

26. Fluet GG, Qiu Q, Patel J, Cronce A, Merians AS, Adamovich SV. Autonomous use of the home virtual rehabilitation system: a feasibility and pilot study. Games Health J 2019 Dec;8(6):432-438 [FREE Full text] [doi: 10.1089/g4h.2019.0012] [Medline: 31769724]
27. Maresca G, Maggio MG, Latella D, Cannavò A, De Cola MC, Portaro S, et al. Toward improving poststroke aphasia: a pilot study on the growing use of telerehabilitation for the continuity of care. J Stroke Cerebrovasc Dis 2019 Oct;28(10):104303. [doi: 10.1016/j.jstrokecerebrovasdis.2019.104303] [Medline: 31371144]
28. Sheehy L, Taillon-Hobson A, Sveistrup H, Bilodeau M, Yang C, Welch V, et al. Home-based virtual reality training after discharge from hospital-based stroke rehabilitation: a parallel randomized feasibility trial. Trials 2019 Jun 07;20(1):333 [FREE Full text] [doi: 10.1186/s13063-019-3438-9] [Medline: 31174579]
29. Torrisi M, Maresca G, De Cola MC, Cannavò A, Sciarrone F, Silvestri G, et al. Using telerehabilitation to improve cognitive function in post-stroke survivors: is this the time for the continuity of care? Int J Rehabil Res 2019 Dec;42(4):344-351. [doi: 10.1097/MRR.0000000000000369] [Medline: 31464812]
30. Allegue DR, Kairy D, Higgins J, Archambault P, Michaud F, Miller W, et al. Optimization of upper extremity rehabilitation by combining telerehabilitation with an exergame in people with chronic stroke: protocol for a mixed methods study. JMIR Res Protoc 2020 May 21;9(5):e14629 [FREE Full text] [doi: 10.2196/14629] [Medline: 32097119]
31. Amorim P, Santos BS, Dias P, Silva S, Martins H. Serious games for stroke telerehabilitation of upper limb - a review for future research. Int J Telerehabil 2020 Dec 08;12(2):65-76 [FREE Full text] [doi: 10.5195/ijt.2020.6326] [Medline: 33520096]
32. Burdea GC, Grampurohit N, Kim N, Polistico K, Kadaru A, Pollack S, et al. Feasibility of integrative games and novel therapeutic game controller for telerehabilitation of individuals chronic post-stroke living in the community. Top Stroke Rehabil 2020 Jul;27(5):321-336 [FREE Full text] [doi: 10.1080/10749357.2019.1701178] [Medline: 31875775]
33. Kim WS, Cho S, Ku J, Kim Y, Lee K, Hwang HJ, et al. Clinical application of virtual reality for upper limb motor rehabilitation in stroke: review of technologies and clinical evidence. J Clin Med 2020 Oct 21;9(10):3369 [FREE Full text] [doi: 10.3390/jcm9103369] [Medline: 33096678]
34. Klaic M, Galea MP. Using the technology acceptance model to identify factors that predict likelihood to adopt tele-neurorehabilitation. Front Neurol 2020 Dec 2;11:580832 [FREE Full text] [doi: 10.3389/fneur.2020.580832] [Medline: 33343488]
35. Qiu Q, Cronce A, Patel J, Fluet GG, Mont AJ, Merians AS, et al. Development of the Home based Virtual Rehabilitation System (HoVRS) to remotely deliver an intense and customized upper extremity training. J Neuroeng Rehabil 2020 Nov 23;17(1):155 [FREE Full text] [doi: 10.1186/s12984-020-00789-w] [Medline: 33228709]
36. Sramka M, Lacko J, Ruzicky E, Masan J. Combined methods of rehabilitation of patients after stroke: virtual reality and traditional approach. Neuro Endocrinol Lett 2020 Sep;41(3):123-133. [Medline: 33201645]
37. Thielbar KO, Triandafilou KM, Barry AJ, Yuan N, Nishimoto A, Johnson J, et al. Home-based upper extremity stroke therapy using a multiuser virtual reality environment: a randomized trial. Arch Phys Med Rehabil 2020 Feb;101(2):196-203. [doi: 10.1016/j.apmr.2019.10.182] [Medline: 31715140]
38. Allegue DR, Kairy D, Higgins J, Archambault PS, Michaud F, Miller WC, et al. A personalized home-based rehabilitation program using exergames combined with a telerehabilitation app in a chronic stroke survivor: mixed methods case study. JMIR Serious Games 2021 Aug 31;9(3):e26153 [FREE Full text] [doi: 10.2196/26153] [Medline: 34132649]
39. Jonsdottir J, Baglio F, Gindri P, Isernia S, Castiglioni C, Gramigna C, et al. Virtual reality for motor and cognitive rehabilitation from clinic to home: a pilot feasibility and efficacy study for persons with chronic stroke. Front Neurol 2021 Apr 7;12:601131 [FREE Full text] [doi: 10.3389/fneur.2021.601131] [Medline: 33897579]
40. Ostrowska PM, Śliwiński M, Studnicki R, Hansdorfer-Korzon R. Telerehabilitation of post-stroke patients as a therapeutic solution in the era of the COVID-19 pandemic. Healthcare (Basel) 2021 May 31;9(6):654 [FREE Full text] [doi: 10.3390/healthcare9060654] [Medline: 34072939]
41. Ferreira B, Menezes P. Gamifying motor rehabilitation therapies: challenges and opportunities of immersive technologies. Information 2020 Feb 07;11(2):88 [FREE Full text] [doi: 10.3390/info11020088]
42. Kato N, Tanaka T, Sugihara S, Shimizu K, Kudo N. Trial operation of a cloud service-based three-dimensional virtual reality tele-rehabilitation system for stroke patients. In: Proceedings of the 11th International Conference on Computer Science & Education. 2016 Presented at: ICCSE '16; August 23-25, 2016; Nagoya, Japan p. 285-290. [doi: 10.1109/iccse.2016.7581595]
43. Caminero RV, Méndez-Herrero LA, Diaz-Pernasl FJ, del Campo JC, Rodriguez MA, Martinez-Zarzuelal M. Telekin - tele-rehabilitation system for musculoskeletal and cognitive disorders using natural movement interface devices. In: Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health. 2017 Presented at: ICT4AWE '17; April 28-29, 2017; Porto, Portugal p. 198-203. [doi: 10.5220/0006366701980203]
44. Gandolfi M, Geroin C, Dimitrova E, Boldrini P, Waldner A, Bonadiman S, et al. Virtual reality telerehabilitation for postural instability in Parkinson's disease: a multicenter, single-blind, randomized, controlled trial. Biomed Res Int 2017;2017:7962826 [FREE Full text] [doi: 10.1155/2017/7962826] [Medline: 29333454]
45. O'Neil O, Fernandez MM, Herzog J, Beorchia M, Gower V, Gramatica F, et al. Virtual reality for neurorehabilitation: insights from 3 European clinics. PM R 2018 Sep;10(9 Suppl 2):S198-S206. [doi: 10.1016/j.pmrj.2018.08.375] [Medline: 30121365]

46. Keshner EA, Weiss PT, Geifman D, Raban D. Tracking the evolution of virtual reality applications to rehabilitation as a field of study. J Neuroeng Rehabil 2019 Jun 21;16(1):76 [FREE Full text] [doi: 10.1186/s12984-019-0552-6] [Medline: 31226995]
47. Isernia S, Di Tella S, Pagliari C, Jonsdottir J, Castiglioni C, Gindri P, et al. Effects of an innovative telerehabilitation intervention for people with Parkinson's disease on quality of life, motor, and non-motor abilities. Front Neurol 2020 Aug 13;11:846 [FREE Full text] [doi: 10.3389/fneur.2020.00846] [Medline: 32903506]
48. Kalron A, Achiron A, Pau M, Cocco E. The effect of a telerehabilitation virtual reality intervention on functional upper limb activities in people with multiple sclerosis: a study protocol for the TEAMS pilot randomized controlled trial. Trials 2020 Aug 12;21(1):713 [FREE Full text] [doi: 10.1186/s13063-020-04650-2] [Medline: 32787896]
49. Manenti R, Gobbi E, Baglio F, Macis A, Ferrari C, Pagnoni I, et al. Effectiveness of an innovative cognitive treatment and telerehabilitation on subjects with mild cognitive impairment: a multicenter, randomized, active-controlled study. Front Aging Neurosci 2020 Nov 16;12:585988 [FREE Full text] [doi: 10.3389/fnagi.2020.585988] [Medline: 33304267]
50. Mantovani E, Zucchella C, Bottiroli S, Federico A, Giugno R, Sandrini G, et al. Telemedicine and virtual reality for cognitive rehabilitation: a roadmap for the COVID-19 pandemic. Front Neurol 2020 Sep 15;11:926 [FREE Full text] [doi: 10.3389/fneur.2020.00926] [Medline: 33041963]
51. Maresca G, Maggio MG, De Luca R, Manuli A, Tonin P, Pignolo L, et al. Tele-neuro-rehabilitation in Italy: state of the art and future perspectives. Front Neurol 2020 Sep 30;11:563375 [FREE Full text] [doi: 10.3389/fneur.2020.563375] [Medline: 33101176]
52. Matamala-Gomez M, Maisto M, Montana JI, Mavrodiev PA, Baglio F, Rossetto F, et al. The role of engagement in teleneurorehabilitation: a systematic review. Front Neurol 2020 May 6;11:354 [FREE Full text] [doi: 10.3389/fneur.2020.00354] [Medline: 32435227]
53. Meca-Lallana V, Prefasi D, Alabarcez W, Hernández T, García-Vaz F, Portaña A, et al. A pilot study to explore patient satisfaction with a virtual rehabilitation program in multiple sclerosis: the RehabVR study protocol. Front Neurol 2020 Aug 21;11:900 [FREE Full text] [doi: 10.3389/fneur.2020.00900] [Medline: 33162924]
54. Stasolla F, Matamala-Gomez M, Bernini S, Caffò AO, Bottiroli S. Virtual reality as a technological-aided solution to support communication in persons with neurodegenerative diseases and acquired brain injury during COVID-19 pandemic. Front Public Health 2020 Feb 16;8:635426 [FREE Full text] [doi: 10.3389/fpubh.2020.635426] [Medline: 33665181]
55. da Silva TD, da Silva PL, de Jesus Valenzuela E, Dias ED, Simcsik AO, de Carvalho MG, et al. Serious game platform as a possibility for home-based telerehabilitation for individuals with cerebral palsy during COVID-19 quarantine - a cross-sectional pilot study. Front Psychol 2021 Feb 02;12:622678 [FREE Full text] [doi: 10.3389/fpsyg.2021.622678] [Medline: 33633648]
56. Langer A, Gassner L, Flotz A, Hasenauer S, Gruber J, Wizany L, et al. How COVID-19 will boost remote exercise-based treatment in Parkinson's disease: a narrative review. NPJ Parkinsons Dis 2021 Mar 08;7(1):25 [FREE Full text] [doi: 10.1038/s41531-021-00160-3] [Medline: 33686074]
57. Lasaponara S, Marson F, Doricchi F, Cavallo M. A scoping review of cognitive training in neurodegenerative diseases via computerized and virtual reality tools: what we know so far. Brain Sci 2021 Apr 21;11(5):528 [FREE Full text] [doi: 10.3390/brainsci11050528] [Medline: 33919244]
58. Varela-Aldás J, Buele J, Ramos Lorente P, García-Magariño I, Palacios-Navarro G. A virtual reality-based cognitive telerehabilitation system for use in the COVID-19 pandemic. Sustainability 2021 Feb 18;13(4):2183 [FREE Full text] [doi: 10.3390/su13042183]
59. Anton D, Berges I, Bermúdez J, Goñi A, Illarramendi A. A telerehabilitation system for the selection, evaluation and remote management of therapies. Sensors (Basel) 2018 May 08;18(5):1459 [FREE Full text] [doi: 10.3390/s18051459] [Medline: 29738442]
60. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 03;20(1):506 [FREE Full text] [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]
61. Berton A, Longo UG, Candela V, Fioravanti S, Giannone L, Arcangeli V, et al. Virtual reality, augmented reality, gamification, and telerehabilitation: psychological impact on orthopedic patients' rehabilitation. J Clin Med 2020 Aug 07;9(8):2567 [FREE Full text] [doi: 10.3390/jcm9082567] [Medline: 32784745]
62. Prvu Bettger J, Green CL, Holmes DN, Chokshi A, Mather 3rd RC, Hoch BT, et al. Effects of virtual exercise rehabilitation in-home therapy compared with traditional care after total knee arthroplasty: VERITAS, a randomized controlled trial. J Bone Joint Surg Am 2020 Jan 15;102(2):101-109. [doi: 10.2106/JBJS.19.00695] [Medline: 31743238]
63. Borresen A, Wolfe C, Lin CK, Tian Y, Raghuraman S, Nahrstedt K, et al. Usability of an immersive augmented reality based telerehabilitation system with haptics (ARTESH) for synchronous remote musculoskeletal examination. Int J Telerehabil 2019 Jun 12;11(1):23-32 [FREE Full text] [doi: 10.5195/ijt.2019.6275] [Medline: 31341544]
64. Rothgangel A, Braun S, Smeets R, Beurskens A. Feasibility of a traditional and teletreatment approach to mirror therapy in patients with phantom limb pain: a process evaluation performed alongside a randomized controlled trial. Clin Rehabil 2019 Oct;33(10):1649-1660. [doi: 10.1177/0269215519846539] [Medline: 31066315]

65. Afyouni I, Murad A, Einea A. Adaptive rehabilitation bots in serious games. Sensors (Basel) 2020 Dec 09;20(24):7037 [FREE Full text] [doi: 10.3390/s20247037] [Medline: 33316916]
66. Birckhead B, Eberlein S, Alvarez G, Gale R, Dupuy T, Makaroff K, et al. Home-based virtual reality for chronic pain: protocol for an NIH-supported randomised-controlled trial. BMJ Open 2021 Jun 15;11(6):e050545 [FREE Full text] [doi: 10.1136/bmjopen-2021-050545] [Medline: 34130965]
67. Logan DE, Simons LE, Caruso TJ, Gold JI, Greenleaf W, Griffin A, et al. Leveraging virtual reality and augmented reality to combat chronic pain in youth: position paper from the interdisciplinary network on virtual and augmented technologies for pain management. J Med Internet Res 2021 Apr 26;23(4):e25916 [FREE Full text] [doi: 10.2196/25916] [Medline: 33667177]
68. Yeo SM, Lim JY, Do JG, Lim JY, In Lee J, Hwang JH. Effectiveness of interactive augmented reality-based telerehabilitation in patients with adhesive capsulitis: protocol for a multi-center randomized controlled trial. BMC Musculoskelet Disord 2021 Apr 26;22(1):386 [FREE Full text] [doi: 10.1186/s12891-021-04261-1] [Medline: 33902546]
69. Annaswamy TM, Pradhan GN, Chakka K, Khargonkar N, Borresen A, Prabhakaran B. Using biometric technology for telehealth and telerehabilitation. Phys Med Rehabil Clin N Am 2021 May;32(2):437-449. [doi: 10.1016/j.pmr.2020.12.007] [Medline: 33814068]
70. Piech J, Czernicki K. Virtual reality rehabilitation and exergames—physical and psychological impact on fall prevention among the elderly—a literature review. Appl Sci 2021 Apr 30;11(9):4098 [FREE Full text] [doi: 10.3390/app11094098]
71. Gmez-Portes C, Lacave C, Molina AI, Vallejo D. Home rehabilitation based on gamification and serious games for young people: a systematic mapping study. Appl Sci 2020 Dec 10;10(24):8849 [FREE Full text] [doi: 10.3390/app10248849]
72. Tamplin J, Loveridge B, Clarke K, Li Y, Berlowitz DJ. Development and feasibility testing of an online virtual reality platform for delivering therapeutic group singing interventions for people living with spinal cord injury. J Telemed Telecare 2020 Jul;26(6):365-375. [doi: 10.1177/1357633X19828463] [Medline: 30823854]
73. Cerdán de Las Heras J, Tulppo M, Kiviniemi AM, Hilberg O, Løkke A, Ekholm S, et al. Augmented reality glasses as a new tele-rehabilitation tool for home use: patients' perception and expectations. Disabil Rehabil Assist Technol 2022 May;17(4):480-486. [doi: 10.1080/17483107.2020.1800111] [Medline: 32750254]
74. Jung T, Moorhouse N, Shi X, Amin MF. A virtual reality-supported intervention for pulmonary rehabilitation of patients with chronic obstructive pulmonary disease: mixed methods study. J Med Internet Res 2020 Jul 07;22(7):e14178 [FREE Full text] [doi: 10.2196/14178] [Medline: 32673224]
75. Lorenzini MC, Wittich W. Head-mounted visual assistive technology-related quality of life changes after telerehabilitation. Optom Vis Sci 2021 Jun 01;98(6):582-591 [FREE Full text] [doi: 10.1097/OPX.0000000000001705] [Medline: 34081648]
76. Matamala-Gomez M, Bottiroli S, Realdon O, Riva G, Galvagni L, Platz T, et al. Telemedicine and virtual reality at time of COVID-19 pandemic: an overview for future perspectives in neurorehabilitation. Front Neurol 2021 Mar 25;12:646902 [FREE Full text] [doi: 10.3389/fneur.2021.646902] [Medline: 33841313]
77. Asadzadeh A, Samad-Soltani T, Rezaei-Hachesu P. Applications of virtual and augmented reality in infectious disease epidemics with a focus on the COVID-19 outbreak. Inform Med Unlocked 2021;24:100579 [FREE Full text] [doi: 10.1016/j.imu.2021.100579] [Medline: 33937503]
78. Liu P, Li C, Xiao C, Zhang Z, Ma J, Gao J, et al. A wearable augmented reality navigation system for surgical telementoring based on Microsoft HoloLens. Ann Biomed Eng 2021 Jan;49(1):287-298. [doi: 10.1007/s10439-020-02538-5] [Medline: 32504141]
79. Kutzin JM, Milligan Z, Chawla S. Using simulation to conduct a usability study of wearable technology. Clin Simul Nurs 2017 Feb;13(2):64-70. [doi: 10.1016/j.ecns.2016.12.003]
80. Pérez Alonso N, Pardo Rios M, Juguera Rodriguez L, Vera Catalan T, Segura Melgarejo F, Lopez Ayuso B, et al. Randomised clinical simulation designed to evaluate the effect of telemedicine using Google Glass on cardiopulmonary resuscitation (CPR). Emerg Med J 2017 Nov;34(11):734-738. [doi: 10.1136/emermed-2016-205998] [Medline: 28768700]
81. Andersen D, Popescu V, Cabrera ME, Shanghavi A, Mullis B, Marley S, et al. An augmented reality-based approach for surgical telementoring in austere environments. Mil Med 2017 Mar;182(S1):310-315. [doi: 10.7205/MILMED-D-16-00051] [Medline: 28291491]
82. Greenfield MJ, Luck J, Billingsley ML, Heyes R, Smith OJ, Mosahebi A, et al. Demonstration of the effectiveness of augmented reality telesurgery in complex hand reconstruction in Gaza. Plast Reconstr Surg Glob Open 2018 Mar;6(3):e1708 [FREE Full text] [doi: 10.1097/GOX.0000000000001708] [Medline: 29707463]
83. Harris TE, DeLellis SF, Heneghan JS, Buckman RF, Miller GT, Magee JH, et al. Augmented reality forward damage control procedures for nonsurgeons: a feasibility demonstration. Mil Med 2020 Jan 07;185(Suppl 1):521-525. [doi: 10.1093/milmed/usz298] [Medline: 32074340]
84. Lin C, Rojas-Munoz E, Cabrera ME, Sanchez-Tamayo N, Andersen D, Popescu V, et al. How about the mentor? Effective workspace visualization in AR telementoring. In: Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces. 2020 Presented at: VR '20; March 22-26, 2020; Atlanta, GA, USA p. 212-220. [doi: 10.1109/vr46266.2020.00040]

85. Rojas-Muñoz E, Lin C, Sanchez-Tamayo N, Cabrera ME, Andersen D, Popescu V, et al. Evaluation of an augmented reality platform for austere surgical telementoring: a randomized controlled crossover study in cricothyroidotomies. NPJ Digit Med 2020 May 21;3:75 [FREE Full text] [doi: 10.1038/s41746-020-0284-9] [Medline: 32509972]
86. Weibel N, Gasques D, Johnson J, Sharkey T, Xu ZR, Zhang X, et al. ARTEMIS: mixed-reality environment for immersive surgical telementoring. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI, USA p. 1-4. [doi: 10.1145/3334480.3383169]
87. Rojas-Muñoz E, Cabrera ME, Lin C, Andersen D, Popescu V, Anderson K, et al. The System for Telementoring with Augmented Reality (STAR): a head-mounted display to improve surgical coaching and confidence in remote areas. Surgery 2020 Apr;167(4):724-731. [doi: 10.1016/j.surg.2019.11.008] [Medline: 31916990]
88. Glick Y, Avital B, Oppenheimer J, Nahman D, Wagnert-Avraham L, Eisenkraft A, et al. Augmenting prehospital care. BMJ Mil Health 2021 Jun;167(3):158-162. [doi: 10.1136/jramc-2019-001320] [Medline: 32086268]
89. Carbone M, Freschi C, Mascioli S, Ferrari V, Ferrari M. A wearable augmented reality platform for telemedicine. In: Proceedings of the 3rd International Conference on Augmented Reality, Virtual Reality, and Computer Graphics. 2016 Presented at: AVR '16; June 15-18, 2016; Lecce, Italy p. 92-100. [doi: 10.1007/978-3-319-40651-0_8]
90. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016 Feb;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
91. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google Glass in surgical settings: systematic review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
92. Andersen DS, Cabrera ME, Rojas-Muñoz EJ, Popescu VS, Gonzalez GT, Mullis B, et al. Augmented reality future step visualization for robust surgical telementoring. Simul Healthc 2019 Feb;14(1):59-66. [doi: 10.1097/SIH.0000000000000334] [Medline: 30395078]
93. Mitsuno D, Hirota Y, Akamatsu J, Kino H, Okamoto T, Ueda K. Telementoring demonstration in craniofacial surgery with Hololens, Skype, and three-layer facial models. J Craniofac Surg 2019 Jan;30(1):28-32. [doi: 10.1097/SCS.0000000000004899] [Medline: 30439735]
94. Rojas-Muñoz E, Andersen D, Cabrera ME, Popescu V, Marley S, Zarzaur B, et al. Augmented reality as a medium for improved telementoring. Mil Med 2019 Mar 01;184(Suppl 1):57-64. [doi: 10.1093/milmed/usy300] [Medline: 30901394]
95. Rojas-Muñoz E, Cabrera ME, Andersen D, Popescu V, Marley S, Mullis B, et al. Surgical telementoring without encumbrance: a comparative study of see-through augmented reality-based approaches. Ann Surg 2019 Aug;270(2):384-389. [doi: 10.1097/SLA.0000000000002764] [Medline: 29672404]
96. Darrow DP, Spano A, Grande A. The potential for undue patient exposure during the use of telementoring technology. Cureus 2020 Apr 08;12(4):e7594 [FREE Full text] [doi: 10.7759/cureus.7594] [Medline: 32399328]
97. Sadeghi AH, Mathari SE, Abjigitova D, Maat AP, Taverne YJ, Bogers AJ, et al. Current and future applications of virtual, augmented, and mixed reality in cardiothoracic surgery. Ann Thorac Surg 2022 Feb;113(2):681-691 [FREE Full text] [doi: 10.1016/j.athoracsur.2020.11.030] [Medline: 33347848]
98. Cofano F, Di Perna G, Bozzaro M, Longo A, Marengo N, Zenga F, et al. Augmented reality in medical practice: from spine surgery to remote assistance. Front Surg 2021 Mar 30;8:657901 [FREE Full text] [doi: 10.3389/fsurg.2021.657901] [Medline: 33859995]
99. Lareyre F, Chaudhuri A, Adam C, Carrier M, Mialhe C, Raffort J. Applications of head-mounted displays and smart glasses in vascular surgery. Ann Vasc Surg 2021 Aug;75:497-512. [doi: 10.1016/j.avsg.2021.02.033] [Medline: 33823254]
100. Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, et al. Augmented reality as a telemedicine platform for remote procedural training. Sensors (Basel) 2017 Oct 10;17(10):2294 [FREE Full text] [doi: 10.3390/s17102294] [Medline: 28994720]
101. Deldar K, Bahaadinbeigy K, Tara SM. Teleconsultation and clinical decision making: a systematic review. Acta Inform Med 2016 Jul 16;24(4):286-292 [FREE Full text] [doi: 10.5455/aim.2016.24.286-292] [Medline: 27708494]
102. Noorian AR, Bahr Hosseini M, Avila G, Gerardi R, Andrle A, Su M, et al. Use of wearable technology in remote evaluation of acute stroke patients: feasibility and reliability of a Google Glass-based device. J Stroke Cerebrovasc Dis 2019 Oct;28(10):104258. [doi: 10.1016/j.jstrokecerebrovasdis.2019.06.016] [Medline: 31296476]
103. Martin G, Koizia L, Kooner A, Cafferkey J, Ross C, Purkayastha S, PanSurg Collaborative. Use of the HoloLens2 mixed reality headset for protecting health care workers during the COVID-19 pandemic: prospective, observational evaluation. J Med Internet Res 2020 Aug 14;22(8):e21486 [FREE Full text] [doi: 10.2196/21486] [Medline: 32730222]
104. Skolnik AB, Chai PR, Dameff C, Gerkin R, Monas J, Padilla-Jones A, et al. Teletoxicology: patient assessment using wearable audiovisual streaming technology. J Med Toxicol 2016 Dec;12(4):358-364 [FREE Full text] [doi: 10.1007/s13181-016-0567-3] [Medline: 27381429]
105. Kaylor J, Hooper V, Wilson A, Burkert R, Lyda M, Fletcher K, et al. Reliability testing of augmented reality glasses technology: establishing the evidence base for telewound care. J Wound Ostomy Continence Nurs 2019;46(6):485-490. [doi: 10.1097/WON.0000000000000585] [Medline: 31633610]

106. Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M. Technical support by smart glasses during a mass casualty incident: a randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine. J Med Internet Res 2019 Jan 03;21(1):e11939 [FREE Full text] [doi: 10.2196/11939] [Medline: 30609988]
107. Sirilak S, Muneesawang P. A new procedure for advancing telemedicine using the HoloLens. IEEE Access 2018 Oct 11;6:60224-60233. [doi: 10.1109/access.2018.2875558]
108. Hu HZ, Feng XB, Shao ZW, Xie M, Xu S, Wu XH, et al. Application and prospect of mixed reality technology in medical field. Curr Med Sci 2019 Feb;39(1):1-6. [doi: 10.1007/s11596-019-1992-8] [Medline: 30868484]
109. Venkata HS, Alsadoon A, Prasad PW, Alsadoon OH, Haddad S, Deva A, et al. A novel mixed reality in breast and constructive jaw surgical tele-presence. Comput Methods Programs Biomed 2019 Aug;177:253-268. [doi: 10.1016/j.cmpb.2019.05.025] [Medline: 31319954]
110. Zhang Q, Karunanithi M, Kang C. Immersive augmented reality (I am real) – remote clinical consultation. In: Proceedings of the 2019 IEEE EMBS International Conference on Biomedical & Health Informatics. 2019 Presented at: BHI '19; May 19-22, 2019; Chicago, IL, USA p. 1-4. [doi: 10.1109/bhi.2019.8834641]
111. Dananjayan S, Raj GM. 5G in healthcare: how fast will be the transformation? Ir J Med Sci 2021 May;190(2):497-501. [doi: 10.1007/s11845-020-02329-w] [Medline: 32737688]
112. Desselle MR, Brown RA, James AR, Midwinter MJ, Powell SK, Woodruff MA. Augmented and virtual reality in surgery. Comput Sci Eng 2020 May 1;22(3):18-26. [doi: 10.1109/mcse.2020.2972822]
113. Neupane A, Alsadoon A, Prasad PW, Ali RS, Haddad S. A novel modified chaotic simplified advanced encryption system (MCS-AES): mixed reality for a secure surgical tele-presence. Multimed Tools Appl 2020 Aug 08;79(39-40):29043-29067. [doi: 10.1007/s11042-020-09478-1]
114. Pinter C, Lasso A, Choueib S, Asselin M, Fillion-Robin JC, Vimort JB, et al. SlicerVR for medical intervention training and planning in immersive virtual reality. IEEE Trans Med Robot Bionics 2020 May;2(2):108-117 [FREE Full text] [doi: 10.1109/tmrb.2020.2983199] [Medline: 33748693]
115. Higginbotham G. Virtual connections: improving global neurosurgery through immersive technologies. Front Surg 2021 Feb 19;8:629963 [FREE Full text] [doi: 10.3389/fsurg.2021.629963] [Medline: 33681283]
116. Hanna MG, Ahmed I, Nine J, Prajapati S, Pantanowitz L. Augmented reality technology using Microsoft HoloLens in anatomic pathology. Arch Pathol Lab Med 2018 May;142(5):638-644 [FREE Full text] [doi: 10.5858/arpa.2017-0189-OA] [Medline: 29384690]
117. Christensen JK. The emergence and unfolding of telemonitoring practices in different healthcare organizations. Int J Environ Res Public Health 2018 Jan 03;15(1):61 [FREE Full text] [doi: 10.3390/ijerph15010061] [Medline: 29301384]
118. Vidal-Balea A, Blanco-Novoa Ó, Fraga-Lamas P, Fernández-Caramés TM. Developing the next generation of augmented reality games for pediatric healthcare: an open-source collaborative framework based on ARCore for implementing teaching, training and monitoring applications. Sensors (Basel) 2021 Mar 07;21(5):1865 [FREE Full text] [doi: 10.3390/s21051865] [Medline: 33800070]
119. Kowatsch T, Lohse K, Erb V, Schittenhelm L, Galliker H, Lehner R, et al. Hybrid ubiquitous coaching with a novel combination of mobile and holographic conversational agents targeting adherence to home exercises: four design and evaluation studies. J Med Internet Res 2021 Feb 22;23(2):e23612 [FREE Full text] [doi: 10.2196/23612] [Medline: 33461957]
120. Thøgersen M, Andoh J, Milde C, Graven-Nielsen T, Flor H, Petrini L. Individualized augmented reality training reduces phantom pain and cortical reorganization in amputees: a proof of concept study. J Pain 2020;21(11-12):1257-1269. [doi: 10.1016/j.jpain.2020.06.002] [Medline: 32574786]
121. Ponce BA, Brabston EW, Zu S, Watson SL, Baker D, Winn D, et al. Telemedicine with mobile devices and augmented reality for early postoperative care. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:4411-4414. [doi: 10.1109/EMBC.2016.7591705] [Medline: 28269256]
122. McKnight RR, Pean CA, Buck JS, Hwang JS, Hsu JR, Pierrie SN. Virtual reality and augmented reality-translating surgical training into surgical technique. Curr Rev Musculoskelet Med 2020 Dec;13(6):663-674 [FREE Full text] [doi: 10.1007/s12178-020-09667-3] [Medline: 32779019]
123. Nikouline A, Jimenez MC, Okrainec A. Feasibility of remote administration of the fundamentals of laparoscopic surgery (FLS) skills test using Google wearable device. Surg Endosc 2020 Jan;34(1):443-449. [doi: 10.1007/s00464-019-06788-w] [Medline: 31037339]
124. Chakrabarti S. Usefulness of telepsychiatry: a critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015 Sep 22;5(3):286-304 [FREE Full text] [doi: 10.5498/wjp.v5.i3.286] [Medline: 26425443]
125. Levy F, Leboucher P, Rautureau G, Jouvent R. E-virtual reality exposure therapy in acrophobia: a pilot study. J Telemed Telecare 2016 Jun;22(4):215-220. [doi: 10.1177/1357633X15598243] [Medline: 26253746]
126. Jebbar Y, Belqasmi F, Glitho R, Alfandi O. A fog-based architecture for remote phobia treatment. In: Proceedings of the 2019 IEEE International Conference on Cloud Computing Technology and Science. 2019 Presented at: CloudCom '19; December 11-13, 2019; Sydney, Australia p. 271-278. [doi: 10.1109/cloudcom.2019.00047]
127. Cikajlo I, Cizman Staba U, Vrhovac S, Larkin F, Roddy M. A cloud-based virtual reality app for a novel telemindfulness service: rationale, design and feasibility evaluation. JMIR Res Protoc 2017 Jun 05;6(6):e108 [FREE Full text] [doi: 10.2196/resprot.6849] [Medline: 28583904]

128. Baccon LA, Chiarovano E, MacDougall HG. Virtual reality for teletherapy: avatars may combine the benefits of face-to-face communication with the anonymity of online text-based communication. Cyberpsychol Behav Soc Netw 2019 Feb;22(2):158-165. [doi: 10.1089/cyber.2018.0247] [Medline: 30540490]
129. Yuen EK, Goetter EM, Stasio MJ, Ash P, Mansour B, McNally E, et al. A pilot of acceptance and commitment therapy for public speaking anxiety delivered with group videoconferencing and virtual reality exposure. J Contextual Behav Sci 2019 Apr;12:47-54. [doi: 10.1016/j.jcbs.2019.01.006]
130. Pedram S, Palmisano S, Perez P, Mursic R, Farrelly M. Examining the potential of virtual reality to deliver remote rehabilitation. Comput Human Behav 2020 Apr;105:106223. [doi: 10.1016/j.chb.2019.106223]
131. Galunder SS, Gottlieb JF, Ladwig J, Hamell J, Keller PK, Wu P. A VR ecosystem for telemedicine and non-intrusive cognitive and affective assessment. In: Proceedings of the IEEE 6th International Conference on Serious Games and Applications for Health. 2018 Presented at: SeGAH '18; May 16-18, 2018; Vienna, Austria p. 1-6. [doi: 10.1109/segah.2018.8401347]
132. Shao D, Lee IJ. Acceptance and influencing factors of social virtual reality in the urban elderly. Sustainability 2020 Nov 10;12(22):9345 [FREE Full text] [doi: 10.3390/su12229345]
133. Nakagawa K, Yellowlees PM. University of California Technology Wellness Index: a physician-centered framework to assess technologies' impact on physician well-being. Psychiatr Clin North Am 2019 Dec;42(4):669-681. [doi: 10.1016/j.psc.2019.08.005] [Medline: 31672216]
134. Sampaio M, Haro MV, De Sousa B, Melo WV, Hoffman HG. Therapists make the switch to telepsychology to safely continue treating their patients during the COVID-19 pandemic. Virtual reality telepsychology may be next. Front Virtual Real 2021 Jan;1:576421 [FREE Full text] [doi: 10.3389/frvir.2020.576421] [Medline: 33585834]
135. Mohan A, Wara UU, Arshad Shaikh MT, Rahman RM, Zaidi ZA. Telesurgery and robotics: an improved and efficient era. Cureus 2021 Mar 26;13(3):e14124 [FREE Full text] [doi: 10.7759/cureus.14124] [Medline: 33927932]
136. Jiménez Moreno R, Espinosa Valcárcel FA, Amaya Hurtado D. Teleoperated systems: a perspective on telesurgery applications. Rev Ing Biomed 2013;7(14):30-41 [FREE Full text]
137. Guo J, Guo S, Tamiya T, Hirata H, Ishihara H. A virtual reality-based method of decreasing transmission time of visual feedback for a tele-operative robotic catheter operating system. Int J Med Robot 2016 Mar;12(1):32-45. [doi: 10.1002/rcs.1642] [Medline: 25693866]
138. Bhattacherjee S, Chattopadhayay S, Rao V, Seth S, Mukherjee S, Sengupta A, et al. Kinematics and teleoperation of tendon driven continuum robot. Procedia Comput Sci 2018;133:879-886. [doi: 10.1016/j.procs.2018.07.106]
139. Abu-Kheil Y, Al Trad O, Seneviratne L, Dias J. A proposed clinical evaluation of a simulation environment for magnetically-driven active endoscopic capsules. Adv Exp Med Biol 2019;1170:87-94. [doi: 10.1007/978-3-030-24230-5_4] [Medline: 32067204]
140. Ho DK. Using smartphone-delivered stereoscopic vision in microsurgery: a feasibility study. Eye (Lond) 2019 Jun;33(6):953-956 [FREE Full text] [doi: 10.1038/s41433-019-0356-8] [Medline: 30755728]
141. Richter F, Zhang Y, Zhi Y, Orosco RK, Yip MC. Augmented reality predictive displays to help mitigate the effects of delayed telesurgery. In: Proceedings of the 2019 International Conference on Robotics and Automation. 2019 Presented at: ICRA '19; May 20-24, 2019; Montreal, Canada p. 444-450. [doi: 10.1109/icra.2019.8794051]
142. De Silva R, Siriwardhana Y, Samarasinghe T, Ylianttila M, Liyanage M. Local 5G operator architecture for delay critical telehealth applications. In: Proceedings of the IEEE 3rd 5G World Forum. 2020 Presented at: 5GWF '20; September 10-12, 2020; Bangalore, India p. 257-262. [doi: 10.1109/5GWF49715.2020.9221292]
143. Tuladhar S, AlSallami N, Alsadoon A, Prasad PW, Alsadoon OH, Haddad S, et al. A recent review and a taxonomy for hard and soft tissue visualization-based mixed reality. Int J Med Robot 2020 Oct;16(5):1-22. [doi: 10.1002/rcs.2120] [Medline: 32388923]
144. Piccolo D, Smolle J, Wolf IH, Peris K, Hofmann-Wellenhof R, Dell'Eva G, et al. Face-to-face diagnosis vs telediagnosis of pigmented skin tumors: a teledermoscopic study. Arch Dermatol 1999 Dec;135(12):1467-1471. [doi: 10.1001/archderm.135.12.1467] [Medline: 10606051]
145. telediagnosis. Farlex. URL: https://medical-dictionary.thefreedictionary.com/telediagnosis [accessed 2022-08-01]
146. Ji H, Wang J, Gao J, Liu X. Research on telemedicine technology and implement based on virtual reality. In: Proceedings of the 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference. 2016 Presented at: IMCEC '16; October 3-5, 2016; Xi'an, China p. 1581-1586. [doi: 10.1109/imcec.2016.7867484]
147. Tian Y, Raghuraman S, Annaswamy T, Borresen A, Nahrstedt K, Prabhakaran B. H-TIME: haptic-enabled tele-immersive musculoskeletal examination. In: Proceedings of the 25th ACM international conference on Multimedia. 2017 Presented at: MM '17; October 23-27, 2017; Mountain View, CA, USA p. 137-145. [doi: 10.1145/3123266.3123395]
148. Filippeschi A, Brizzi F, Ruffaldi E, Jacinto Villegas JM, Landolfi L, Avizzano CA. Evaluation of diagnostician user interface aspects in a virtual reality-based tele-ultrasonography simulation. Adv Robot 2019 Jul 04;33(15-16):840-852. [doi: 10.1080/01691864.2019.1635909]
149. Baribeau Y, Sharkey A, Chaudhary O, Krumm S, Fatima H, Mahmood F, et al. Handheld point-of-care ultrasound probes: the new generation of POCUS. J Cardiothorac Vasc Anesth 2020 Nov;34(11):3139-3145 [FREE Full text] [doi: 10.1053/j.jvca.2020.07.004] [Medline: 32736998]

150. Syawaludin MF, Lee M, Hwang JI. Foveation pipeline for 360° video-based telemedicine. Sensors (Basel) 2020 Apr 16;20(8):2264 [FREE Full text] [doi: 10.3390/s20082264] [Medline: 32316257]
151. Ramsingh D, Van Gorkom C, Holsclaw M, Nelson S, De La Huerta M, Hinson J, et al. Use of a smartphone-based augmented reality video conference app to remotely guide a point of care ultrasound examination. Diagnostics (Basel) 2019 Oct 24;9(4):159 [FREE Full text] [doi: 10.3390/diagnostics9040159] [Medline: 31652998]
152. Wang GG. Definition and review of virtual prototyping. J Comput Inf Sci Eng 2002;2(3):232-236. [doi: 10.1115/1.1526508]
153. Trenholme D, Smith SP. Computer game engines for developing first-person virtual environments. Virtual Real 2008 Mar 28;12(3):181-187. [doi: 10.1007/s10055-008-0092-z]
154. Sherman WR, Craig AB. Interface to the virtual world–output. In: Sherman WR, Craig AB, editors. Understanding Virtual Reality: Interface, Application, and Design. Cambridge, MA, USA: Elsevier; 2003:114-203.
155. Cheng L, Tavakoli M. COVID-19 pandemic spurs medical telerobotic systems: a survey of applications requiring physiological organ motion compensation. Front Robot AI 2020 Nov 9;7:594673 [FREE Full text] [doi: 10.3389/frobt.2020.594673] [Medline: 33501355]
156. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertain Comput 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
157. Nickel F, Cizmic A, Chand M. Telestration and augmented reality in minimally invasive surgery: an invaluable tool in the age of COVID-19 for remote proctoring and telementoring. JAMA Surg 2022 Feb 01;157(2):169-170. [doi: 10.1001/jamasurg.2021.3604] [Medline: 34705030]
158. Liston M, Genna G, Maurer C, Kikidis D, Gatsios D, Fotiadis D, et al. Investigating the feasibility and acceptability of the HOLOBalance system compared with standard care in older adults at risk for falls: study protocol for an assessor blinded pilot randomised controlled study. BMJ Open 2021 Feb 12;11(2):e039254 [FREE Full text] [doi: 10.1136/bmjopen-2020-039254] [Medline: 33579762]
159. Cassidy D. Broadband connection types explained. Bonkers Money. 2022 Feb 3. URL: https://www.bonkers.ie/guides/broadband-phone-tv/broadband-connection-types-explained/ [accessed 2022-09-11]
160. What is telehealth? NEJM Catal Innov Care Deliv 2018 Feb 1;4(1). [doi: 10.1056/CAT.18.0268]
161. Jay A. Cloud-based vs web-based applications: a comparison of features & key aspects. Finances Online. 2019 May 16. URL: https://financesonline.com/cloud-based-web-based-applications-a-comparison-of-features-key-aspects/ [accessed 2022-09-11]
162. Gunshin M, Doi K, Morimura N. Use of high-fidelity simulation technology in disasters: an integrative literature review. Acute Med Surg 2020 Dec 20;7(1):e596 [FREE Full text] [doi: 10.1002/ams2.596] [Medline: 33364034]
163. Persky S. A virtual home for the virtual clinical trial. J Med Internet Res 2020 Jan 03;22(1):e15582 [FREE Full text] [doi: 10.2196/15582] [Medline: 31899455]
164. Koledova E, Le Masne Q, Spataru A, Bagha M, Dixon D. Digital health in the management of pediatric growth hormone therapy - 10 years of developments. Stud Health Technol Inform 2021 May 27;281:926-930. [doi: 10.3233/SHTI210314] [Medline: 34042809]
165. Threapleton K, Drummond A, Standen P. Virtual rehabilitation: what are the practical barriers for home-based research? Digit Health 2016 Apr 29;2:2055207616641302 [FREE Full text] [doi: 10.1177/2055207616641302] [Medline: 29942551]
166. Yu J, Ferniany W, Guthrie B, Parekh SG, Ponce B. Lessons learned from Google Glass: telemedical spark or unfulfilled promise? Surg Innov 2016 Apr;23(2):156-165. [doi: 10.1177/1553350615597085] [Medline: 26224576]
167. Vogiatzaki E, Krukowski A. Maintaining mental wellbeing of elderly at home. In: Ganchev I, Garcia NM, Dobre C, Mavromoustakis CX, Goleva R, editors. Enhanced Living Environments: Algorithms, Architectures, Platforms, and Systems. Cham, Switzerland: Springer; 2019:177-209.

Abbreviations

AR: augmented reality
HMD: head-mounted display
H-TIME: Haptic Enable Tele-Immersion Musculoskeletal Examination
ICT: information and communication technology
LCD: liquid crystal display
LED: light emitting diode
MR: mixed reality
POCUS: point-of-focus ultrasonography
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
REBOA: resuscitative endovascular balloon occlusion of the aorta
ReCoVR: Realizing Collaborative virtual reality for well-being and self-healing
STAR: System for Telementoring with Augmented Reality
USG: ultrasonography

VE: virtual environment
VERGE: Virtual Environment for Rehabilitative Gaming Exercise
VR: virtual reality

Edited by A Mavragani; submitted 14.09.22; peer-reviewed by J Hogg, A Asadzadeh; comments to author 14.10.22; revised version received 03.11.22; accepted 16.11.22; published 24.03.23.

Please cite as:
Worlikar H, Coleman S, Kelly J, O'Connor S, Murray A, McVeigh T, Doran J, McCabe I, O'Keeffe D
Mixed Reality Platforms in Telehealth Delivery: Scoping Review
JMIR Biomed Eng 2023;8:e42709
URL: https://biomedeng.jmir.org/2023/1/e42709
doi:10.2196/42709
PMID:38875694

Original Paper

An Algorithm to Classify Real-World Ambulatory Status From a Wearable Device Using Multimodal and Demographically Diverse Data: Validation Study

Sara Popham[1], PhD; Maximilien Burq[1], PhD; Erin E Rainaldi[1], MSc; Sooyoon Shin[1], PhD; Jessilyn Dunn[2,3,4], PhD; Ritu Kapur[1], PhD

[1]Verily Life Sciences, South San Francisco, CA, United States

[2]Department of Biomedical Engineering, Duke University, Durham, NC, United States

[3]Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, United States

[4]Duke Clinical Research Institute, Durham, NC, United States

Corresponding Author:
Sara Popham, PhD
Verily Life Sciences
269 E Grand Ave
South San Francisco, CA, 94080
United States
Phone: 1 650 253 0000
Email: spopham@verily.com

Abstract

Background: Measuring the amount of physical activity and its patterns using wearable sensor technology in real-world settings can provide critical insights into health status.

Objective: This study's aim was to develop and evaluate the analytical validity and transdemographic generalizability of an algorithm that classifies binary ambulatory status (yes or no) on the accelerometer signal from wrist-worn biometric monitoring technology.

Methods: Biometric monitoring technology algorithm validation traditionally relies on large numbers of self-reported labels or on periods of high-resolution monitoring with reference devices. We used both methods on data collected from 2 distinct studies for algorithm training and testing, one with precise ground-truth labels from a reference device (n=75) and the second with participant-reported ground-truth labels from a more diverse, larger sample (n=1691); in total, we collected data from 16.7 million 10-second epochs. We trained a neural network on a combined data set and measured performance in multiple held-out testing data sets, overall and in demographically stratified subgroups.

Results: The algorithm was accurate at classifying ambulatory status in 10-second epochs (area under the curve 0.938; 95% CI 0.921-0.958) and on daily aggregate metrics (daily mean absolute percentage error 18%; 95% CI 15%-20%) without significant performance differences across subgroups.

Conclusions: Our algorithm can accurately classify ambulatory status with a wrist-worn device in real-world settings with generalizability across demographic subgroups. The validated algorithm can effectively quantify users' walking activity and help researchers gain insights on users' health status.

(JMIR Biomed Eng 2023;8:e43726) doi:10.2196/43726

KEYWORDS

digital measurement; wearable sensor; machine learning; ambulatory status; Project Baseline Health Study; physical activity

Introduction

Quantifying physical activity can be highly informative about both general health status and the condition of people with specific diseases [1,2]. Characteristics of physical activity have been shown to be prognostic factors in various chronic conditions [3-13]. Yet reliably producing research-grade measurements of physical activity in real-world settings remains a challenge. Traditionally, the validation of such measurements often relies on individual self-reports or is performed

episodically and in artificial laboratory environments. These approaches suffer from known challenges, such as subjectivity, assessment bias, and unreliability [14-16].

Recently, the advent of wearable technology has made it possible to measure physical activity to a previously untenable extent [17,18]. Ambulatory activity in particular, namely whether individuals are walking and how much, is a basic aspect of physical activity that can be investigated in general populations and in specific clinical settings. Wearable devices can collect information passively during daily living and generate a vast quantity of digital measurements that allow researchers to probe functional physical activity generally and ambulatory activity specifically. Using these digital measures in research studies, however, requires analytical validation [19]. In their design, validation studies have to balance factors such as feasibility and the resource-intensiveness of their data collection approach with demonstrating validity in representative populations.

To date, the majority of measurements in validation studies have come from either short observation periods in laboratory settings [20,21] or self-reported labels in real-world settings [22]. Laboratory measurements often render observations with exceptionally clean and easy-to-use ground-truth labels, but algorithms trained on data of this kind do not always generalize to everyday activities [23]. On the other hand, using self-reported labels as the ground truth yields a closer reflection of individual everyday activities, but these labels are often noisy and less accurate [15,16,24]. There have been some examples of reference devices deployed to generate accurate truth labels in generalizable real-world settings [25,26], but this came at the cost of intrusiveness and resource-intensive data processing steps after collection, such as manual video footage tagging. With all these considerations in mind, validation studies tend to be highly heterogeneous, and need to be interpreted in context.

Herein we report on the development and analytical validation of an ambulatory status classification algorithm. This algorithm classifies the ambulatory status of users of a wrist-worn device in real-world environments. We carried out 2 separate studies including participants from independent populations with distinct sources of ground-truth labels for a deeper characterization of the algorithm performance. One of the studies, the pilot program study, used a relatively small and demographically homogeneous cohort, where participants provided a highly accurate ground-truth source from a reference device. The other study was derived from the Project Baseline Health Study (PBHS), a prospective, multicenter, longitudinal study with participants of diverse backgrounds who were representative of the entire health spectrum [27]; this was a demographically diverse cohort that provided self-reported labels as the ground-truth source. This cohort was also relatively large, and we therefore expected it to yield results less susceptible to outlier readouts. We present analytical validation results of the performance of our algorithm against the highly accurate ground-truth source (from the pilot), and we examine the generalizability of the results across a study population of demographically diverse individuals (in the PBHS).

Methods

Participant Cohorts

Two distinct studies were conducted, with training and testing groups identified a priori within each study. Participants in both studies wore the smartwatch (the Verily Study Watch) [27-30].

The first study was a pilot program (n=75) of adult volunteer participants recruited among Verily Life Sciences employees in 2 locations (South San Francisco, California, and Cambridge, Massachusetts) without specific selection criteria. For this group, ground-truth labels were collected from an ankle-worn reference device (StepWatch 4). The Verily Study Watch and reference device were worn simultaneously for 7 consecutive days to ensure capture of both weekday and weekend behavior; for each participant, days were included as evaluable if both devices were worn synchronously for a minimum of 8 hours. No demographic information on race or ethnicity was collected in this study. The observation period ran from June to December 2019.

In order to expand the demographic representativeness of the overall validation effort, the second study included a large and diverse cohort (n=1691) consisting of participants from the PBHS who consented to participate in this substudy [27]. The period for data collection ran from May to December 2019.

Ethics Approval

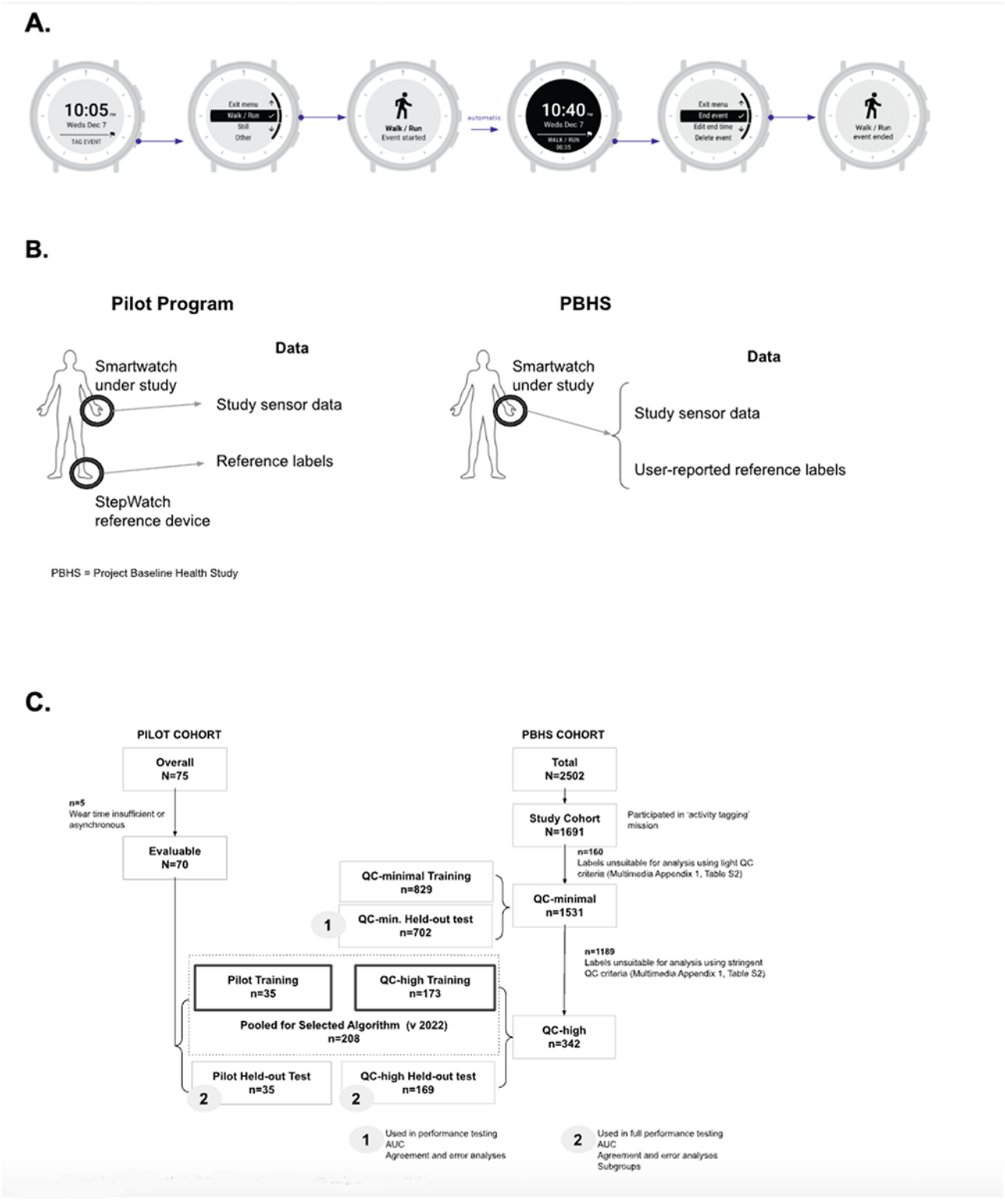
The pilot program was determined to be exempt research that did not require institutional review board review. Written informed consent was obtained from all participants enrolled in the PBHS; the PBHS was approved by both the WCG institutional review board (approval tracking number 20170163, work order number 1-1506365-1) and the institutional review boards of each participating institution (Stanford University, Duke University, and the California Health and Longevity Institute) [27]. The PBHS was registered at ClinicalTrials.gov (NCT03154346).

All methods complied with relevant guidelines and regulations; the research involving human participants was performed in accordance with relevant guidelines and regulations. Experimental protocols were approved by appropriate committees from Verily Life Sciences and by PBHS governance (participating institutions are above).

Wearable Devices

The Verily Study Watch recorded acceleration data in both cohorts via an onboard inertial measurement unit with a 30 Hz 3-axis accelerometer. For the PBHS population, the smartwatch also contained a user interface allowing participants to tag their activities on the watch (Figure 1A).

Figure 1. (A) Sketch of the user interface of the study device used in the Project Baseline Health Study. (B) Data elements for the 2 studies. (C) Flow of participant inclusion for the different cohorts and data sets in the 2 studies. AUC: area under the receiver operating characteristic curve; PBHS: Project Baseline Health Study; QC: quality control.



The reference device for the pilot program was an ankle-worn single-axis accelerometer (Modus StepWatch 4) that provided step count as a reference label for algorithm development.

Reference Labels

In the pilot program, we generated reference labels on data collected from the ankle-worn StepWatch: 10-second windows were considered "ambulatory" if they had ≥3 steps on the wearing foot and "nonambulatory" if they had <3 steps [31]. The default window size returned by this device was 10 seconds, and this was deemed to provide good temporal granularity.

For the self-reported reference labels from the PBHS, participants tagged their activities as 1 of 3 options listed by the wrist device: "walk/run," "still," and "other." Participants could tag the start and end of an activity period directly on the

watch, which enabled precise synchronization of the labels to the raw sensor data stream. When necessary, participants could edit or delete tags as needed (Figure 1B). For the purpose of this analysis, "still" and "other" were grouped together under the "nonambulatory" label, while "walk/run" was equated to "ambulatory."

The amount of data used from each of these studies is summarized in Multimedia Appendix 1, Table S3.

Algorithm Development

Data from each study (pilot program, n=75; PBHS, n=1691) were split into nonoverlapping training and testing data sets at the participant level. For each study, data from approximately half the participants were used for training the algorithm and data from the other half were held out for algorithm testing. We decided on a 50-50 split in order to retain statistical power in the testing data, particularly considering the intended additional analyses of different demographic subgroups (discussed below).

In the pilot program, the split into training and testing data sets was based on participants' daily step counts in order to mitigate potential algorithmic biases caused by training primarily on data from participants with either very low or very high activity levels. The difference in the mean daily step counts between the 2 halves of the split was 234 steps. For the PBHS cohort, the split into training and testing data sets was done randomly, as participants did not have daily aggregated results. We trained multiple versions of the algorithm with combinations of different subsets of training data and compared performance across these different algorithms (Figure 1).

We developed an algorithm that classifies the ambulatory status of device users in 10-second epochs (as ambulatory vs nonambulatory). First, the following 14 features were extracted from the Verily Study Watch's acceleration data, in 10-second epochs: 3 features related to deviations of the signal, 5 features derived from power spectral density energy in frequency bands typically associated with user ambulation (ie, walking or running), 2 features that are signal percentiles (ie, 95th percentiles), and 4 features that are differences between signal percentiles (ie, IQR). These features were fed into a shallow neural network model with 2 dense layers: ReLu nonlinearities and softmax of outputs. The neural network was trained with a batch size of 32. The Adam optimizer was used with a learning rate of 0.001, and loss was calculated using categorical cross-entropy. Training ran for 10 epochs. Alternative features and neural network architectures were explored using the training data, but larger feature sets or more complex architectures did not result in higher performance, so this algorithm was chosen.

The classifier threshold was optimized to minimize absolute percentage error on daily ambulatory time on the training data from the pilot study (vs the data from the reference device used as the ground-truth source, as discussed above). For this optimization process, we performed 5-fold cross-validation at the participant level within the training data. We found the minimum daily mean absolute percentage error (MAPE) across the aggregated held-out data from all folds using a 1D grid search procedure.

The signal-processing, feature selection, model training, and hyperparameter tuning were all performed on training data sets identified a priori.

Analyses

The demographic characteristics of the study cohorts were analyzed using descriptive statistics.

We analyzed the following metrics to characterize the performance of the algorithm, calculated on the held-out test sets: area under the receiver operating characteristic curve (AUC) for the overall study cohorts and across different demographic subcohorts within the PBHS cohort (this was chosen as the metric for comparison because, unlike other measures, such as F_1-score or accuracy, it is not susceptible to differences in the chosen classifier threshold), mean accuracy, and MAPE of daily ambulatory time, defined as the summing of all 10-second windows that were labeled as "ambulatory" in a day.

Analyses were performed in python using *NumPy* (version 1.21.5), *pandas* (version 1.1.5), *SciPy* (version 1.2.1), *scikit-learn* (version 1.0.2), and *tensorflow* (version 2.10.0).

Confidence intervals were calculated using the bootstrap method with 1000 resampling iterations. Resampling was done at the participant level to ensure that all data from a single participant were either included or excluded within each resampling iteration.

Results

Characteristics of Participants From the Pilot Study and the PBHS Cohort

Participants in the pilot study were mostly male (45/75, 64%), with a mean age of 33 (SD 8.5) years. Participants from the PBHS were more often female (1366/2502, 55%), with a mean age of 54 (SD 17) years (Multimedia Appendix 1, Table S1).

Algorithm Training

Data from each study were separately split (approximately 50-50) into nonoverlapping training and testing data sets (Figure 1); this allocation was done at the participant level (n=75 from the pilot study and n=1691 from the PBHS population). Out of 16.769 million 10-second epochs collected from the 2 studies, 8.841 million 10-second epochs were used for training across all algorithm iterations generated (the data sets are described in Multimedia Appendix 1, Table S3).

From the pilot program study, a total of 1,641,272 nonoverlapping 10-second epochs were collected (n=70 participants; Figure 1), of which 228,721 (13.9%) were "ambulatory" according to the reference device–based labels. We used 879,593 10-second epochs (from 35 unique participants) for training (118,730, 13.5% of which were "ambulatory"; Multimedia Appendix 1, Table S3).

We collected a total of 14,814,910 nonoverlapping 10-second epochs from the PBHS (n=1531 participants; Figure 1), of which 7,079,216 (47.8%) were "ambulatory" according to the participant-reported reference labels. The proportion of

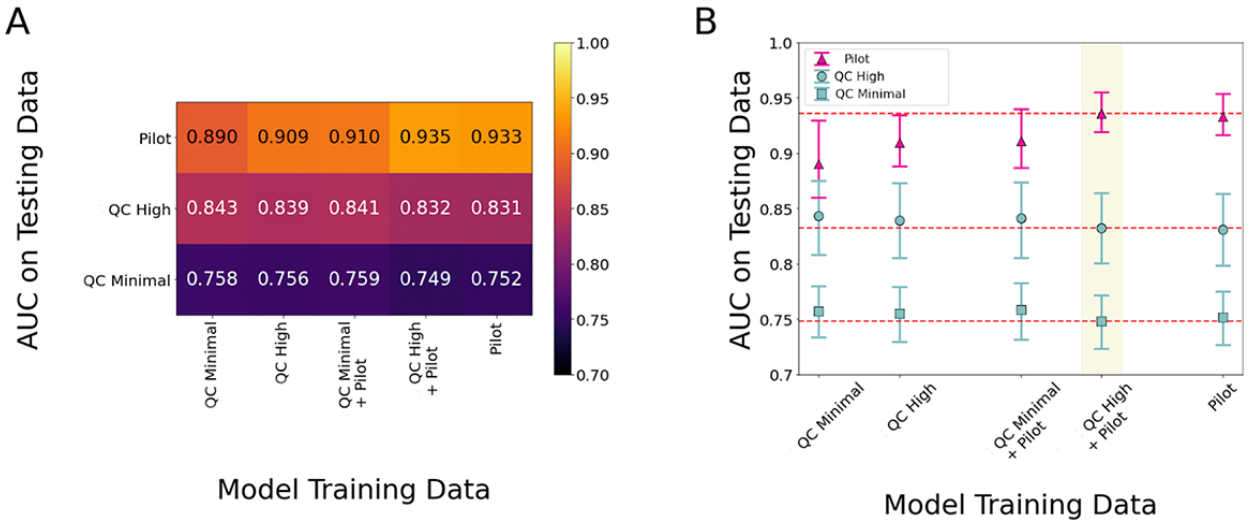
"ambulatory" labels in the PBHS was higher than in the pilot program study (47.8% vs 13.9%), which is likely attributable to the different labeling methods across studies. We expect that labeling from the pilot study was more stringent to show true ambulatory epochs, because these were determined directly by the reference device readouts (ie, any 10-second epoch with greater than or equal to 6 steps, relative to all 10-second epochs collected during the wear time). In the PBHS, the proportion of ambulatory labels was determined based on participant self-reported, manually entered walk/run tags relative to all entered tags. PBHS tagging, therefore, can be more vulnerable to selection bias toward "ambulatory," since participants may favor reporting active over inactive states.

Data from the PBHS were not only divided into training and testing sets, but, across each set, we considered 2 quality control (QC) strata to test the impact of data quality on the development and performance of the algorithm. An extremely light QC selection, eliminating labels with gross apparent user errors (such as tags that were longer than a full day), was applied to generate the "QC-minimal" sub–data set, which therefore included virtually all labels suitable for evaluation (10,264/104,212, 9.8% of user-tagged events were eliminated, and another 12,010/104,212, 11.5% were truncated); a more stringent selection was applied to generate the "QC-high" sub–data set (80,852/104,212, 77.6% of user-tagged events were eliminated, and all tags were truncated to some degree; Figure 1 and Multimedia Appendix 1, Table S2). The 2 strata aimed to parse out performance variability due to noise generated by imperfectly self-reported reference labels (this was not a factor for the labels from the reference device in the pilot program). The resulting size of these QC training sub–data sets was 160,778 10-second epochs for QC-high (n=173 participants) and 7,802,829 for QC-minimal (n=829 participants). Of these labeled epochs, 102,783 (63.7%) and 3,863,964 (49.5%), respectively, were ambulatory according to the participant-reported tags (Figure 1 and Multimedia Appendix 1, Table S3).

Effect of Raw Data Quality on Algorithm Performance

We tested each of the algorithm iterations from the training process above (originated using the 2 PBHS QC sub–data sets and the pilot data set) across data from the held-out QC sub–data sets from the PBHS and the pilot program by calculating AUC values across all combinations. Namely, we tested each of the following algorithms against the held-out data sets from the pilot study and the PBHS QC-high and QC-minimal sub–data sets (Figure 2): (1) trained with the PBHS QC-high sub–data set, (2) trained with the PBHS QC-minimal sub–data set, (3) trained with the pooled PBHS QC-high plus pilot data set, (4) trained with the pooled PBHS QC-minimal plus pilot data set, and (5) trained with just the pilot data set. For each algorithm iteration, AUC values varied across the testing sub–data sets (QC sub–data sets from the PBHS and pilot program), with differences ranging from 0.047 to 0.187. For each test data set, the AUC variations across the algorithm iterations (1) through (5) were narrower, with differences ranging between 0.001 and 0.045. Therefore, data quality differences across the training sub–data sets did not appear to affect algorithm performance, as reflected in AUC variability, as much as data quality in the testing sub–data sets.

Figure 2. (A) Heat map of AUC values for the algorithm iterations generated via different training sub–data sets from the PBHS when tested on each of the separate testing cohorts. (B) AUC values for the algorithm iterations generated via different training sub–data sets from the PBHS when tested on each of the separate testing cohorts, with error bars based on the 95% CI. Each testing cohort is shown with a different color or symbol. From top to bottom, the red dotted lines indicate mean AUC values for the pilot, PBHS QC-high, and PBHS QC-minimal test data sets, respectively. The model trained on combined PBHS QC-high and pilot training data (highlighted in yellow) was the version of the algorithm used for further analyses. AUC: area under the receiver operating characteristic curve; PBHS: Project Baseline Health Study; QC: quality control.



Based on the testing results described above, we selected an algorithm trained using combined data from one of the PBHS sub–data sets (QC-high) plus the pilot program data set to proceed to further analysis. This algorithm iteration (termed "version 2022") showed the highest testing performance (evaluated by AUC) calculated with data from the pilot program (the most precise and cleanest data set) without substantially reduced performance on PBHS data (Figure 2). With this approach, we prioritized testing the accuracy of the algorithm against participants' actual ambulatory status based on the

reference device, not against the type of labels that are most feasible to obtain (ie, self-reported labels), although we report accuracy on both types of labels.

Algorithm Testing

Tested against the held-out data set from the pilot program (Table 1), the selected algorithm had a sensitivity of 71% and a specificity of 95%, for an overall accuracy of 91.5% (95% CI 90.3%-92.9%; Figure 3A) and an AUC of 0.938 (95% CI 0.921-0.958; Figure 3B) when classifying the ambulatory status of 10-second epochs. When tested on the held-out data set from the PBHS QC-high sub–data set, the selected algorithm had an overall accuracy of 75.7% (95% CI 72.5%-78.6%) and an AUC of 0.832 (95% CI 0.800-0.864).

Table 1. Algorithm performance measures.

	Accuracy	Sensitivity	Specificity	PPV[a]	F_1-score	AUC-ROC[b]	AUC-PRC[c]
Pilot study	91.3%	0.706	0.948	0.696	0.701	0.938	0.781
PBHS[d] QC[e]-high	75.8%	0.731	0.802	0.885	0.788	0.832	0.901

[a]PPV: positive predictive value.

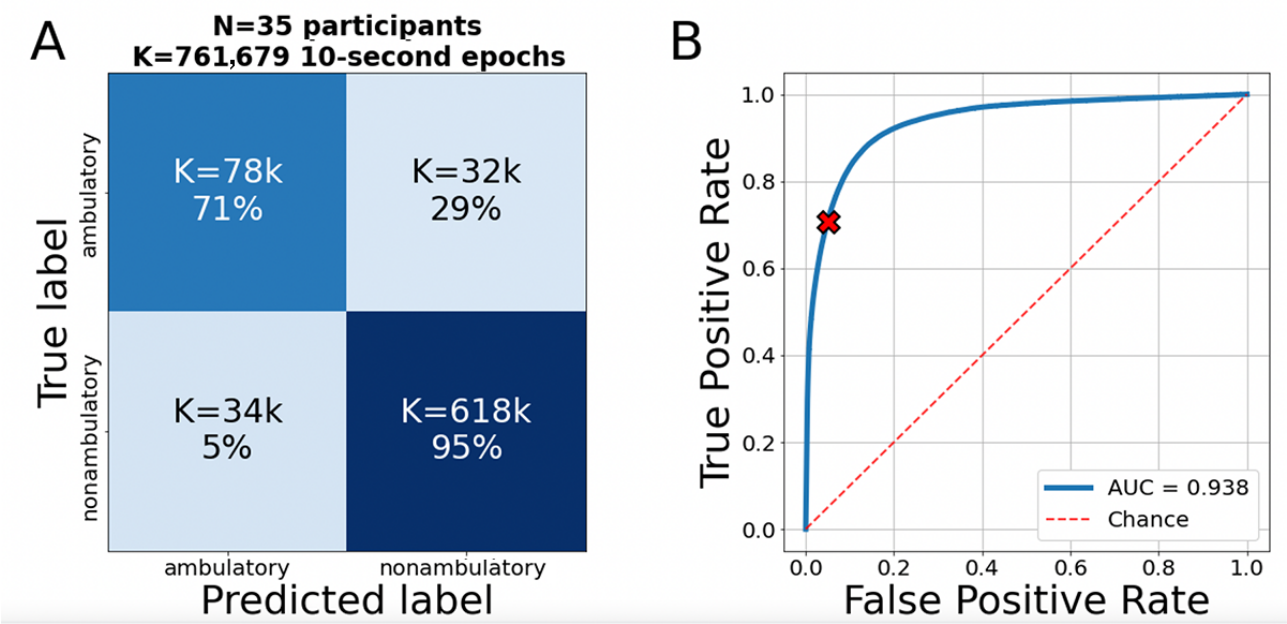
[b]AUC-ROC: area under the receiver operating characteristic curve.

[c]AUC-PRC: area under the precision-recall curve.

[d]PBHS: Project Baseline Health Study.

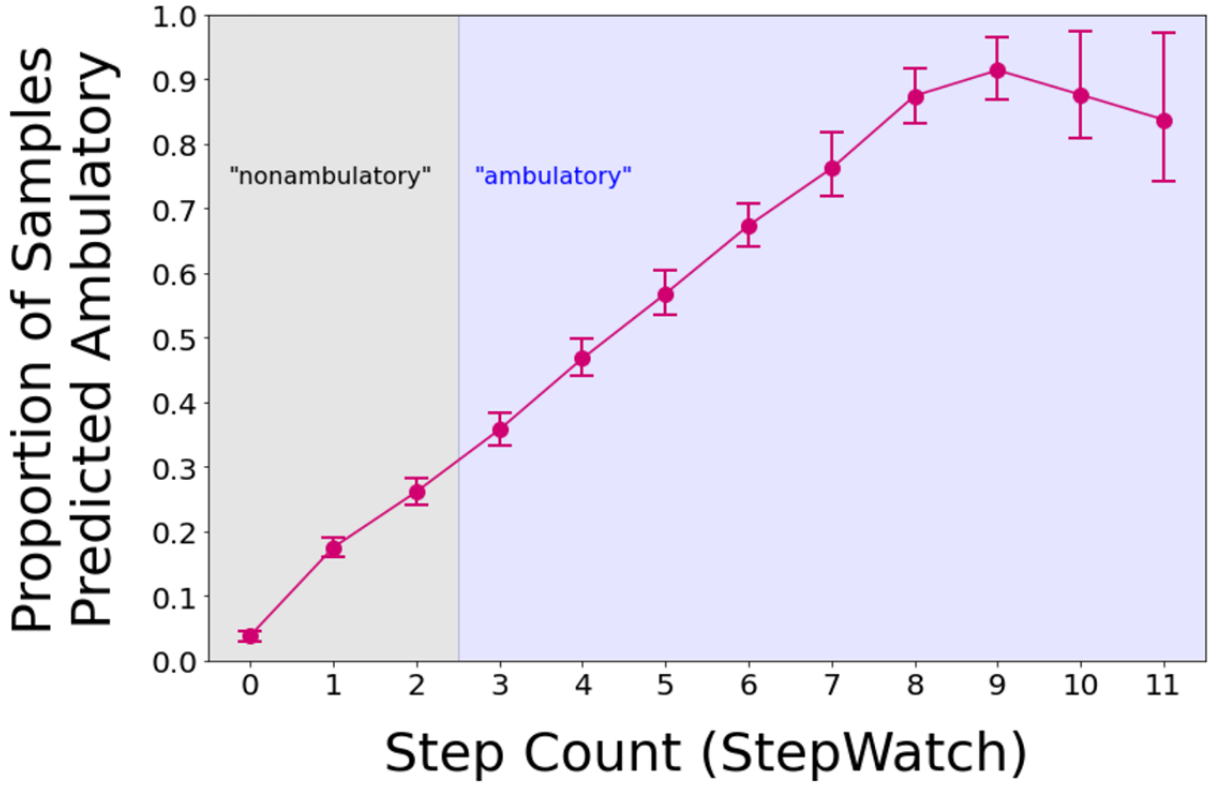
[e]QC: quality control.

Figure 3. (A) Accuracy of the algorithm selected for full analysis, as evaluated in the pilot cohort. Here, the color map denotes K, the number of 10-second epochs. Percentages are normalized across rows, which allows easy reading of the sensitivity and specificity values. (B) Receiver operating characteristic curve and area under the curve of the algorithm selected for full validation, as evaluated in the pilot cohort. The red X denotes the true positive rate and false positive rate of the algorithm at the chosen classifier threshold. AUC: area under the receiver operating characteristic curve.



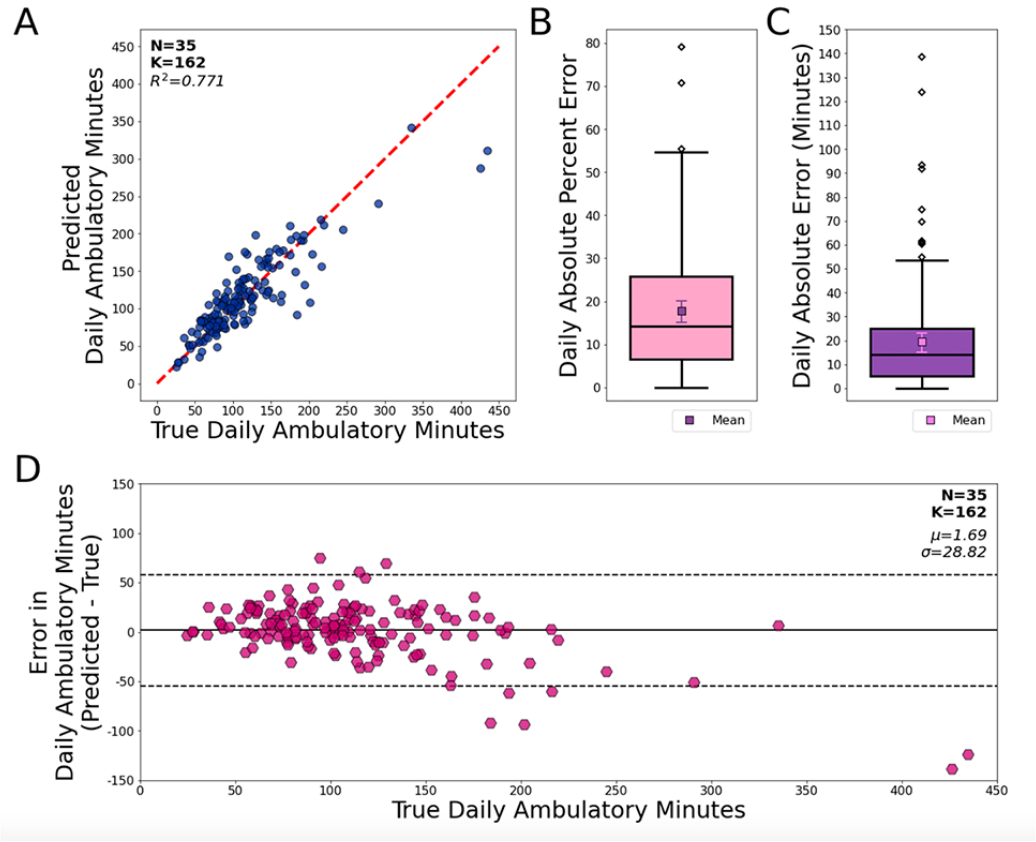
The proportion of predicted ambulatory epochs of the selected algorithm varied with the number of steps in the 10-second epochs (Figure 4). The lowest proportion of predicted ambulatory epochs happened in the 3 to 5 step range (36%-57% sensitivity, ie, correct predictions as “ambulatory: yes”), and the proportion of epochs classified as ambulatory grew with additional steps in the 10-second window (67%-91% correct predictions). Note that data from epochs with more than 11 recorded single-leg steps are not shown due to their low frequency (the number of samples per step count is shown in Multimedia Appendix 1, Figure S1).

Figure 4. Predictions of the selected algorithm to classify 10-second epochs as ambulatory (or not) according to the number of steps in the 10-second epochs based on the reference device data from the pilot program study. A perfectly performing algorithm would predict "ambulatory" for all epochs with 3 or more steps on the wearing foot (indicated by the blue shadow), and nonambulatory for all epochs with fewer steps (indicated by the gray shadow). Epochs with more than 11 recorded steps are not shown due to their small sample size (Multimedia Appendix 1, Figure S1).



When considering daily step aggregates as the metric of interest, there was good agreement between the algorithm classifications and the reference (R^2=0.771), with a MAPE in daily ambulatory time (minutes) of 18% (95% CI 15%-20%) and a median absolute percentage error of 14% (Figure 5A and Figure 5B). The mean absolute error (MAE) of daily ambulatory time was 19.5 (95% CI 15.0-23.2) minutes, and the median absolute error was 14 minutes (Figure 5C). Consistent with the observations at the 10-second epoch level, the magnitude of error in daily ambulatory time (ie, the difference between algorithm-predicted and actual values) was dependent on the actual daily ambulatory time (as computed by the StepWatch; Figure 5D): the chance for underestimating daily ambulatory time (in minutes) grew as the reference daily ambulatory time increased. The largest underestimation we observed was 138.5 minutes in absolute time (relative error 32.5%).

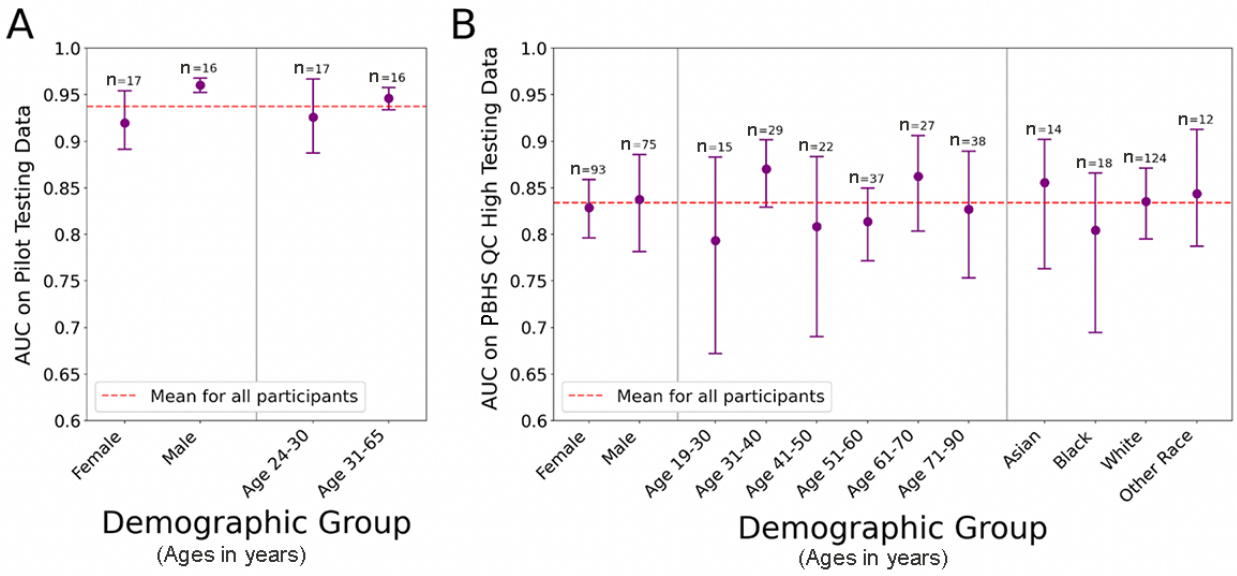
Figure 5. Agreement and error rates of the algorithm predictions. K is the number of user-days. (A) Agreement between the selected algorithm's predictions and the ground-truth source for daily ambulatory time in the pilot-testing data set. (B) Absolute percent error in daily ambulatory time: median (pink box) and mean (purple box). (C) Absolute error in daily ambulatory time in minutes: median (purple box) and mean (pink box). (D) Modified Bland-Altman plot showing error in daily ambulatory time (in minutes) as it relates to the ground-truth daily ambulatory time.



Performance of the Ambulatory Status Classification Algorithm Across Demographic Subgroups

In order to characterize the generalizability of the algorithm's performance, we calculated AUC values for the selected algorithm across demographic subgroups based on gender, age, and race. Initially, in the testing data set from the pilot program (Figure 6A), the results suggested a possible difference in performance between male and female participants, as seen in the lack of overlap of the 95% CIs. However, in a similar analysis using the larger and more diverse testing data set from the PBHS, which enabled subanalyses by participant gender, race, and age, that difference was no longer present and the results showed no meaningful performance difference across any of the subgroups of age, gender, or race, as evidenced by the overlapping 95% CIs (Figure 6B). A replication of the majority population from the pilot study within the PBHS showed an AUC of 0.8166 (95% CI 0.7501-0.8666) for White males aged 31 to 65 years in the PBHS cohort, which was not significantly different from the AUC of the PBHS cohort as a whole (AUC 0.8339).

Figure 6. Performance (AUC values) of the selected algorithm across different demographic subgroups. (A) The pilot study testing data set. (B) The PBHS QC-high testing data set. AUC: area under the receiver operating characteristic curve; PBHS: Project Baseline Health Study; QC: quality control.



Discussion

This study presents the analytical validation in a real-world setting of an algorithm to classify the ambulatory status of users wearing a smartwatch. The algorithm performs well, distinguishing between ambulatory and nonambulatory states with high accuracy (75.7%-91.5% depending on the testing data set). Furthermore, the approach taken to analytic validation allowed us to investigate multiple subgroups, including age, gender, and race, demonstrating that the high performance of the algorithm is generalizable across a broad range of demographics.

All existing validation studies of ambulatory status classification from wrist-worn sensors have been either performed on young and healthy populations [25] or in the laboratory or clinic [20,21]. Yet measuring ambulatory status or daily ambulatory time is most clinically relevant for people with walking impairments—whether due to age, movement disorders, cardiovascular illness, or other circumstances—and most informative when done in an individual's own environment (ie, their real-world setting). Thus, a key innovation in this work is our focus on using data captured in real-world settings (as opposed to highly controlled clinic or laboratory settings) from demographically diverse cohorts for the actual development and validation of this algorithm.

Therefore, the novel contributions of this work are 2-fold. First, we introduce a scalable framework for collecting reference labels on ambulatory status via a reference device and via user-reported data for training and validation. As part of that approach, we used 2 separate and different modalities to measure ground-truth status. This strategy enabled us to handle both comprehensive and highly precise labels (in the pilot program), as well as a larger volume of inherently noisy ones (user-reported tags from the PBHS), both in real-world settings. Our strong results across both sets of data indicate that this innovative multimodal approach contributed to a robust development scheme that may have boosted the performance of the resulting algorithm. The long-term practical convenience of a wrist-worn device (as opposed to an ankle-worn device or a dedicated assessment period) may be advantageous for this type of continuous generalizable monitoring [32-35], although a thorough side-by-side analysis of these 2 reference standard measurement methods to fully understand their correlation remains as a topic for future studies.

Second, we leveraged this framework to provide large-scale validation of the performance of the selected algorithm iteration, addressing shortcomings in terms of generalization previously reported in the literature [20,21,32]. Prior studies have used algorithms to report on differences in physical activity by different demographic subgroups but lacked validation data for those algorithms across demographic subgroups [25,36-38]. To our knowledge, this is one of the first studies to show a proper validation approach to develop and test a generalizable algorithm across demographic subgroups where algorithm output could have differed by subgroup.

In addition, our approach highlights several points of interest when developing validation methodologies for this type of algorithm. The increased sample sizes and variability in data quality achieved by combining 2 distinct data sets enabled deeper characterization of the algorithm's performance. One of our studies generated data sets where truth labels were of high quality and accuracy but were collected from a study population limited in scope; the other study collected data from a large and demographically diverse cohort (albeit a somewhat engaged and self-selected participant group who volunteered and expressed interest in the PBHS and its health technology aspects), which allowed us to conduct subgroup analyses for both training and testing. Our results reinforce the well-established fact that modern machine-learning algorithms can sometimes perform well even when trained on a noisy data set [39]. This observation may be useful for researchers navigating study design decisions and tradeoffs, including

sample sizes and data labeling methods. For future research, determining the role of data quality factors in the development and characterization of this type of algorithm is an open issue [18].

Our approach to the generation of reference labels was pragmatic, using deployment-friendly ankle-worn devices or user-reported tags. Neither of these was as resource-demanding as other intensive approaches (ie, video monitoring), but generated information of sufficient quality to conduct our validation with relatively high time resolution (10-second epochs). Of note, the intrinsic nature of the 2 methods used for the generation of reference labels probably contributed to the noticeable difference in the proportions of "ambulatory" labels between the 2 studies (discussed in the Results), with the proportion observed in the pilot program study being the one closest to other literature reports [40].

When interpreting our results in the context of existing literature, it is worth noting that most validation studies for this type of algorithm have used step counts as the metric of interest [31,36,41-53], while ambulatory time (or a related metric) is the focus of a minority of reports [54,55]. In general, considering the close correlation between step count and ambulatory time, the performance of our algorithm could be placed on par with other algorithms, yet detailed side-by-side appraisals of results remain challenging; this research field is in need of standardization [19,56,57].

This study also had limitations. First, in principle, the StepWatch readouts used as ground truth may not have provided perfect accuracy, even though there is extensive literature supporting the use of StepWatch as a reference device [31,50,51,56,58-60]. Second, we observed fluctuations in the ambulatory status classification algorithm performance based on daily ambulatory time; this fluctuation was present when the algorithm detected 10-second epochs as ambulatory (or not) and was also manifested in the daily aggregates of ambulatory time. While this trend (shown in Figure 5) may have been driven, partially, by outlying data points with high step counts in our sample, which would be of little relevance in hypothetical clinical scenarios, it may also have been due to low-step periods containing mixed activities in which walking was not the only or dominant source of hand motion. In addition, while the cutoff used to read the StepWatch ambulatory classification relied on existing literature [61], it may not be perfect in itself. In this regard, it could be reassuring that the algorithm handled epochs with step counts between 4 and 8 as a continuum, as this is possibly reflective of the complexities of organic movement.

In sum, we have developed an accurate algorithm for the detection of the ambulatory status of users of a wrist-worn device in a free-living, real-world setting; the output is generalizable across several user demographic characteristics. The characterization of this algorithm was conducted in 2 distinct data sets, which lends credibility to the robustness and applicability of the performance results obtained in this study and illustrates the advantages of similar approaches to future research in this field.

Acknowledgments

The authors wish to thank David Andresen, Anthony Chan, Chen Chen, Robin Lin, Stephen Lanham, David Miller, Shannon Fong, and the Project Baseline Health Study team and participants for their contributions to this work. The authors also wish to acknowledge writing and editing support from Julia Saiz (Verily). This study was sponsored by Verily Life Sciences, which was responsible for data collection. Authors employed by Verily Life Sciences had access to the raw study data in full. All authors were responsible for interpretation of results and writing and review of the manuscript; all authors approved the final manuscript for submission.

Data Availability

The deidentified PBHS data corresponding to this study are available upon request for the purpose of examining their reproducibility. Interested investigators should direct requests to jsaiz@verily.com. Requests are subject to approval by the Project Baseline Health Study governance. Data from the pilot program are not available due to the nature of this program. Participants in this program did not consent for their data to be shared publicly.

Authors' Contributions

ER and RK contributed to study concept and design; ER contributed to data collection; and SP, MB, SS, ER, and RK contributed to data analysis and interpretation.

Conflicts of Interest

SP, MB, ER, SS, and RK report employment and equity ownership in Verily Life Sciences. JD is a scientific advisor to Veri, Inc.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 61 KB - biomedeng_v8i1e43726_app1.docx]

References

1. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
2. Katzmarzyk P, Powell KE, Jakicic JM, Troiano RP, Piercy K, Tennant B, 2018 Physical Activity Guidelines Advisory Committee. Sedentary Behavior and Health: Update from the 2018 Physical Activity Guidelines Advisory Committee. Med Sci Sports Exerc 2019 Jun;51(6):1227-1241 [FREE Full text] [doi: 10.1249/MSS.0000000000001935] [Medline: 31095080]
3. Johansen KL, Kaysen GA, Dalrymple LS, Grimes BA, Glidden DV, Anand S, et al. Association of physical activity with survival among ambulatory patients on dialysis: the Comprehensive Dialysis Study. Clin J Am Soc Nephrol 2013 Feb;8(2):248-253 [FREE Full text] [doi: 10.2215/CJN.08560812] [Medline: 23124787]
4. Waschki B, Kirsten A, Holz O, Müller KC, Meyer T, Watz H, et al. Physical activity is the strongest predictor of all-cause mortality in patients with COPD: a prospective cohort study. Chest 2011 Aug;140(2):331-342. [doi: 10.1378/chest.10-2521] [Medline: 21273294]
5. Walsh JT, Charlesworth A, Andrews R, Hawkins M, Cowley AJ. Relation of daily activity levels in patients with chronic heart failure to long-term prognosis. Am J Cardiol 1997 May 15;79(10):1364-1369. [doi: 10.1016/s0002-9149(97)00141-0] [Medline: 9165159]
6. Garcia-Aymerich J, Lange P, Benet M, Schnohr P, Antó JM. Regular physical activity reduces hospital admission and mortality in chronic obstructive pulmonary disease: a population based cohort study. Thorax 2006 Sep;61(9):772-778 [FREE Full text] [doi: 10.1136/thx.2006.060145] [Medline: 16738033]
7. Garcia-Rio F, Rojo B, Casitas R, Lores V, Madero R, Romero D, et al. Prognostic value of the objective measurement of daily physical activity in patients with COPD. Chest 2012 Aug;142(2):338-346. [doi: 10.1378/chest.11-2014] [Medline: 22281798]
8. Alahmari AD, Patel AR, Kowlessar BS, Mackay AJ, Singh R, Wedzicha JA, et al. Daily activity during stability and exacerbation of chronic obstructive pulmonary disease. BMC Pulm Med 2014 Jun 02;14:98 [FREE Full text] [doi: 10.1186/1471-2466-14-98] [Medline: 24885188]
9. Hayata A, Minakata Y, Matsunaga K, Nakanishi M, Yamamoto N. Differences in physical activity according to mMRC grade in patients with COPD. Int J Chron Obstruct Pulmon Dis 2016;11:2203-2208 [FREE Full text] [doi: 10.2147/COPD.S109694] [Medline: 27695306]
10. Demeyer H, Gimeno-Santos E, Rabinovich RA, Hornikx M, Louvaris Z, de Boer WI, PROactive consortium. Physical activity characteristics across GOLD quadrants depend on the questionnaire used. PLoS One 2016;11(3):e0151255 [FREE Full text] [doi: 10.1371/journal.pone.0151255] [Medline: 26974332]
11. Chow E, Abdolell M, Panzarella T, Harris K, Bezjak A, Warde P, et al. Predictive model for survival in patients with advanced cancer. J Clin Oncol 2008 Dec 20;26(36):5863-5869. [doi: 10.1200/JCO.2008.17.1363] [Medline: 19018082]
12. Maltoni M, Caraceni A, Brunelli C, Broeckaert B, Christakis N, Eychmueller S, Steering Committee of the European Association for Palliative Care. Prognostic factors in advanced cancer patients: evidence-based clinical recommendations--a study by the Steering Committee of the European Association for Palliative Care. J Clin Oncol 2005 Sep 01;23(25):6240-6248. [doi: 10.1200/JCO.2005.06.866] [Medline: 16135490]
13. Ballard-Barbash R, Friedenreich CM, Courneya KS, Siddiqi SM, McTiernan A, Alfano CM. Physical activity, biomarkers, and disease outcomes in cancer survivors: a systematic review. J Natl Cancer Inst 2012 Jun 06;104(11):815-840 [FREE Full text] [doi: 10.1093/jnci/djs207] [Medline: 22570317]
14. Pitta F, Troosters T, Probst VS, Spruit MA, Decramer M, Gosselink R. Quantifying physical activity in daily life with questionnaires and motion sensors in COPD. Eur Respir J 2006 May;27(5):1040-1055 [FREE Full text] [doi: 10.1183/09031936.06.00064105] [Medline: 16707399]
15. Fiedler J, Eckert T, Burchartz A, Woll A, Wunsch K. Comparison of self-reported and device-based measured physical activity using measures of stability, reliability, and validity in adults and children. Sensors (Basel) 2021 Apr 10;21(8):2672 [FREE Full text] [doi: 10.3390/s21082672] [Medline: 33920145]
16. Skender S, Ose J, Chang-Claude J, Paskow M, Brühmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health 2016 Jun 16;16(1):515 [FREE Full text] [doi: 10.1186/s12889-016-3172-0] [Medline: 27306667]
17. Düking P, Fuss FK, Holmberg H, Sperlich B. Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e102 [FREE Full text] [doi: 10.2196/mhealth.9341] [Medline: 29712629]
18. Poli A, Cosoli G, Scalise L, Spinsante S. Impact of Wearable Measurement Properties and Data Quality on ADLs Classification Accuracy. IEEE Sensors J 2021 Jul 1;21(13):14221-14231. [doi: 10.1109/jsen.2020.3009368]
19. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]

20. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
21. Moore C, McCullough AK, Aguiar EJ, Ducharme SW, Tudor-Locke C. Toward harmonized treadmill-based validation of step-counting wearable technologies: a scoping review. J Phys Act Health 2020 Jul 11:1-13 [FREE Full text] [doi: 10.1123/jpah.2019-0205] [Medline: 32652514]
22. Knaier R, Höchsmann C, Infanger D, Hinrichs T, Schmidt-Trucksäss A. Validation of automatic wear-time detection algorithms in a free-living setting of wrist-worn and hip-worn ActiGraph GT3X. BMC Public Health 2019 Feb 28;19(1):244 [FREE Full text] [doi: 10.1186/s12889-019-6568-9] [Medline: 30819148]
23. Poli A, Spinsante S, Nugent C, Cleland I. Improving the collection and understanding the quality of datasets for the aim of human activity recognition. In: Chen F, García-Betances R, Chen L, Cabrera-Umpiérrez M, Nugent C, editors. Smart Assisted Living. Cham, Switzerland: Springer; 2020.
24. Prince SA, Adamo KB, Hamel M, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
25. Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. Sci Rep 2018 May 21;8(1):7961 [FREE Full text] [doi: 10.1038/s41598-018-26174-1] [Medline: 29784928]
26. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank Study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
27. Arges K, Assimes T, Bajaj V, Balu S, Bashir MR, Beskow L, et al. The Project Baseline Health Study: a step towards a broader mission to map human health. NPJ Digit Med 2020;3:84 [FREE Full text] [doi: 10.1038/s41746-020-0290-y] [Medline: 32550652]
28. Bloem B, Marks WJ, Silva de Lima AL, Kuijf ML, van Laar T, Jacobs BPF, et al. The Personalized Parkinson Project: examining disease progression through broad biomarkers in early Parkinson's disease. BMC Neurol 2019 Jul 17;19(1):160 [FREE Full text] [doi: 10.1186/s12883-019-1394-3] [Medline: 31315608]
29. Burq M, Rainaldi E, Ho KC, Chen C, Bloem BR, Evers LJW, et al. Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. NPJ Digit Med 2022 May 23;5(1):65 [FREE Full text] [doi: 10.1038/s41746-022-00607-8] [Medline: 35606508]
30. McLean SA, Ressler K, Koenen KC, Neylan T, Germine L, Jovanovic T, et al. The AURORA Study: a longitudinal, multimodal library of brain biology and function after traumatic stress exposure. Mol Psychiatry 2020 Feb;25(2):283-296 [FREE Full text] [doi: 10.1038/s41380-019-0581-3] [Medline: 31745239]
31. Toth L, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-recorded validation of wearable step counters under free-living conditions. Med Sci Sports Exerc 2018 Jun;50(6):1315-1322. [doi: 10.1249/MSS.0000000000001569] [Medline: 29381649]
32. Cho J. Current status and prospects of health-related sensing technology in wearable devices. J Healthc Eng 2019;2019:3924508 [FREE Full text] [doi: 10.1155/2019/3924508] [Medline: 31316740]
33. Keogh A, Dorn JF, Walsh L, Calvo F, Caulfield B. Comparing the usability and acceptability of wearable sensors among older Irish adults in a real-world context: observational study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15704 [FREE Full text] [doi: 10.2196/15704] [Medline: 32310149]
34. 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 Jul 30;15:55 [FREE Full text] [doi: 10.1186/s12905-015-0212-3] [Medline: 26223521]
35. Chu AHY, Ng SHX, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
36. Golbus JR, Pescatore NA, Nallamothu BK, Shah N, Kheterpal S. Lancet Digit Health 2021 Nov;3(11):e707-e715 [FREE Full text] [doi: 10.1016/S2589-7500(21)00138-2] [Medline: 34711377]
37. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol 2014 Mar 15;179(6):781-790 [FREE Full text] [doi: 10.1093/aje/kwt330] [Medline: 24500862]
38. Gupta S, Gupta A. Dealing with noise problem in machine learning data-sets: a systematic review. Procedia Computer Science 2019;161:466-474 [FREE Full text] [doi: 10.1016/j.procs.2019.11.146]
39. Wallmann-Sperlich B, Chau JY, Froboese I. Self-reported actual and desired proportion of sitting, standing, walking and physically demanding tasks of office employees in the workplace setting: do they fit together? BMC Res Notes 2017 Nov 17;10(1):504 [FREE Full text] [doi: 10.1186/s13104-017-2829-9] [Medline: 29145883]

40. Giansanti D, Tiberi Y, Silvestri G, Maccioni G. New wearable system for step-counting telemonitoring and telerehabilitation based on the Codivilla spring. Telemed J E Health 2008 Dec;14(10):1096-1100. [doi: 10.1089/tmj.2008.0035] [Medline: 19119833]
41. Magistro D, Brustio PR, Ivaldi M, Esliger DW, Zecca M, Rainoldi A, et al. Validation of the ADAMO Care Watch for step counting in older adults. PLoS One 2018;13(2):e0190753 [FREE Full text] [doi: 10.1371/journal.pone.0190753] [Medline: 29425196]
42. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
43. Chen M, Kuo CC, Pellegrini CA, Hsu MJ. Accuracy of wristband activity monitors during ambulation and activities. Med Sci Sports Exerc 2016 Oct;48(10):1942-1949. [doi: 10.1249/MSS.0000000000000984] [Medline: 27183123]
44. Park S, Marcotte RT, Toth LP, Paulus P, Lauricella LP, Kim AH, et al. Free-living validation and harmonization of 10 wearable step count monitors. Transl J ACSM 2021 Sep 3;6(4):e000172 [FREE Full text] [doi: 10.1249/tjx.0000000000000172]
45. Polese JC, E Faria GS, Ribeiro-Samora GA, Lima LP, Coelho de Morais Faria CD, Scianni AA, et al. Google fit smartphone application or Gt3X Actigraph: Which is better for detecting the stepping activity of individuals with stroke? A validity study. J Bodyw Mov Ther 2019 Jul;23(3):461-465. [doi: 10.1016/j.jbmt.2019.01.011] [Medline: 31563356]
46. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. J Gerontol A Biol Sci Med Sci 2017 Feb;72(2):229-236 [FREE Full text] [doi: 10.1093/gerona/glw098] [Medline: 27257217]
47. Toth LP. Validity of activity tracker step counts during walking, running, and activities of daily living. Transl J Am Coll Sports Med 2018;3:52-59 [FREE Full text] [doi: 10.1249/TJX.0000000000000057]
48. Korpan S, Schafer JL, Wilson KC, Webber SC. Effect of ActiGraph GT3X+ position and algorithm choice on step count accuracy in older adults. J Aging Phys Act 2015 Jul;23(3):377-382. [doi: 10.1123/japa.2014-0033] [Medline: 25102469]
49. Webber SC, St John PD. Comparison of ActiGraph GT3X+ and StepWatch step count accuracy in geriatric rehabilitation patients. J Aging Phys Act 2016 Jul;24(3):451-458. [doi: 10.1123/japa.2015-0234] [Medline: 26751505]
50. Feito Y, Bassett DR, Thompson DL. Evaluation of activity monitors in controlled and free-living environments. Med Sci Sports Exerc 2012 Apr;44(4):733-741. [doi: 10.1249/MSS.0b013e3182351913] [Medline: 21904249]
51. Bender CG, Hoffstot JC, Combs BT, Hooshangi S, Cappos J. Measuring the fitness of fitness trackers. 2017 Presented at: 2017 IEEE Sensors Applications Symposium (SAS); March 13-15, 2017; Glassboro, NJ p. 1-6. [doi: 10.1109/sas.2017.7894077]
52. Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of activity monitor step detection is related to movement patterns. J Phys Act Health 2016 Feb;13(2):145-153. [doi: 10.1123/jpah.2015-0203] [Medline: 26107045]
53. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021;16(5):e0251975 [FREE Full text] [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
54. Burnham J, Lu C, Yaeger LH, Bailey TC, Kollef MH. Using wearable technology to predict health outcomes: a literature review. J Am Med Inform Assoc 2018 Sep 01;25(9):1221-1227 [FREE Full text] [doi: 10.1093/jamia/ocy082] [Medline: 29982520]
55. Walmsley R, Chan S, Smith-Byrne K, Ramakrishnan R, Woodward M, Rahimi K, et al. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. Br J Sports Med 2021 Sep 06;56(18):1008-1017 [FREE Full text] [doi: 10.1136/bjsports-2021-104050] [Medline: 34489241]
56. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Kowalsky RJ, Perera S, Brach JS. Accuracy of Objective Physical Activity Monitors in Measuring Steps in Older Adults. Gerontol Geriatr Med 2018;4:2333721418781126 [FREE Full text] [doi: 10.1177/2333721418781126] [Medline: 29977979]
57. Bent B, Wang K, Grzesiak E, Jiang C, Qi Y, Jiang Y, et al. The digital biomarker discovery pipeline: An open-source software platform for the development of digital biomarkers using mHealth and wearables data. J Clin Transl Sci 2020 Jul 14;5(1):e19 [FREE Full text] [doi: 10.1017/cts.2020.511] [Medline: 33948242]
58. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of Different Activity Monitors to Count Steps in an Inpatient Rehabilitation Setting. Phys Ther 2017 May 01;97(5):581-588. [doi: 10.1093/ptj/pzx010] [Medline: 28339904]
59. Moy ML, Danilack VA, Weston NA, Garshick E. Daily step counts in a US cohort with COPD. Respir Med 2012 Jul;106(7):962-969 [FREE Full text] [doi: 10.1016/j.rmed.2012.03.016] [Medline: 22521225]
60. Schmidt A, Pennypacker ML, Thrush AH, Leiper CI, Craik RL. Validity of the StepWatch Step Activity Monitor: preliminary findings for use in persons with Parkinson disease and multiple sclerosis. J Geriatr Phys Ther 2011;34(1):41-45. [doi: 10.1519/JPT.0b013e31820aa921] [Medline: 21937891]
61. Kluge F, Del Din S, Cereatti A, Gaßner H, Hansen C, Helbostad JL, Mobilise-D consortium. Consensus based framework for digital mobility monitoring. PLoS One 2021;16(8):e0256541 [FREE Full text] [doi: 10.1371/journal.pone.0256541] [Medline: 34415959]

Abbreviations

AUC: area under the receiver operating characteristic curve
IRB: institutional review board
MAE: mean absolute error
MAPE: mean absolute percentage error
PBHS: Project Baseline Health Study
PPV: positive predictive value
PRC: precision-recall curve
QC: quality control
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 21.10.22; peer-reviewed by M Kraus, A Gangadhara Rao; comments to author 24.11.22; revised version received 05.12.22; accepted 19.01.23; published 07.03.23.

Please cite as:
Popham S, Burq M, Rainaldi EE, Shin S, Dunn J, Kapur R
An Algorithm to Classify Real-World Ambulatory Status From a Wearable Device Using Multimodal and Demographically Diverse Data: Validation Study
JMIR Biomed Eng 2023;8:e43726
URL: https://biomedeng.jmir.org/2023/1/e43726
doi:10.2196/43726
PMID:38875664

Original Paper

Development and Testing of a Data Capture Device for Use With Clinical Incentive Spirometers: Testing and Usability Study

Michael L Burns[1], MD, PhD; Anik Sinha[1], MSc; Alexander Hoffmann[1], BSc; Zewen Wu[1], MSc; Tomas Medina Inchauste[1], MSc; Aaron Retsky[1], BSc; David Chesney[2], PhD; Sachin Kheterpal[1], MBA, MD; Nirav Shah[1], MBA, MD

[1]Department of Anesthesiology, University of Michigan Health System, Ann Arbor, MI, United States

[2]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Michael L Burns, MD, PhD
Department of Anesthesiology
University of Michigan Health System
1500 E Medical Center Drive
Ann Arbor, MI, 48109
United States
Phone: 1 734 936 4171
Email: mlburns@med.umich.edu

Abstract

Background: The incentive spirometer is a basic and common medical device from which electronic health care data cannot be directly collected. As a result, despite numerous studies investigating clinical use, there remains little consensus on optimal device use and sparse evidence supporting its intended benefits such as prevention of postoperative respiratory complications.

Objective: The aim of the study is to develop and test an add-on hardware device for data capture of the incentive spirometer.

Methods: An add-on device was designed, built, and tested using reflective optical sensors to identify the real-time location of the volume piston and flow bobbin of a common incentive spirometer. Investigators manually tested sensor level accuracies and triggering range calibrations using a digital flowmeter. A valid breath classification algorithm was created and tested to determine valid from invalid breath attempts. To assess real-time use, a video game was developed using the incentive spirometer and add-on device as a controller using the Apple iPad.

Results: In user testing, sensor locations were captured at an accuracy of 99% (SD 1.4%) for volume and 100% accuracy for flow. Median and average volumes were within 7.5% (SD 6%) of target volume sensor levels, and maximum sensor triggering values seldom exceeded intended sensor levels, showing a good correlation to placement on 2 similar but distinct incentive spirometer designs. The breath classification algorithm displayed a 100% sensitivity and a 99% specificity on user testing, and the device operated as a video game controller in real time without noticeable interference or delay.

Conclusions: An effective and reusable add-on device for the incentive spirometer was created to allow the collection of previously inaccessible incentive spirometer data and demonstrate Internet-of-Things use on a common hospital device. This design showed high sensor accuracies and the ability to use data in real-time applications, showing promise in the ability to capture currently inaccessible clinical data. Further use of this device could facilitate improved research into the incentive spirometer to improve adoption, incentivize adherence, and investigate the clinical effectiveness to help guide clinical care.

(JMIR Biomed Eng 2023;8:e46653) doi:10.2196/46653

KEYWORDS

incentive; spirometry; Internet-of-Things; electronic health records; web-based intervention; medical device; medical tool; data collection; spirometry data; incentive spirometer; data analysis; algorithm; effectiveness

Introduction

Pulmonary complications after major surgery, including pneumonia, atelectasis, respiratory failure, prolonged supplemental oxygen requirements, and reintubation, are common, expensive, and harmful to patients [1]. Studies estimated these complications in the range of 2%-39% [2,3], though atelectasis alone has been found to affect up to 92% of

postsurgical patients [4]. Originating in the 1970s, the incentive spirometer was designed to mimic the physiology of a sigh or a yawn—a slow voluminous inhalation [5,6], and this basic medical device is often used in postoperative care to aid lung expansion to prevent or reduce respiratory complications.

To complicate the matter, correct use and device adherence is low among patients. Guidelines on how to properly use the incentive spirometer device outline best practices [7,8], but over 26% of patients fail to use their incentive spirometer correctly, and over 38% deny ever using their device in their postoperative care [9], highlighting the need for more evidence-based recommendations. There is sparse evidence for the use of the incentive spirometer device [10] for postoperative pulmonary complication prevention [11] with only a few studies demonstrating clinical effectiveness when used properly [12,13]. As a result of the lack of high-level evidence, some clinical practice guidelines do not support its routine postoperative use [14]. Additionally, there remains disagreement as to the most effective way to use the device, as studies have been unable to demonstrate its superiority over other techniques such as deep breathing techniques, directed coughing, early mobilization, and optimal analgesia [15]. Uncertainty around the effective spirometry use is partially due to the scarcity of spirometer compliance data [16]. Compliance measurements, made through self-reporting and staff observation, are difficult to obtain, and when captured, they have demonstrated low patient adherence to the incentive spirometer device [12]. Though data remain elusive, 86% of health care providers believe patient adherence is poor, and 95.4% believe it should be improved [17], demonstrating the perceived use of incentive spirometry. The first step in determining the optimal use of the incentive spirometer is to improve data collection. Automated capture of spirometry data may improve the quality of research studies and ultimately determine the incentive spirometer's use in improving lung function and minimizing postoperative complications.

Incentive spirometers are simple plastic meters to measure inhalational breath flow and volume; they lack the ability to record data. Digital flowmeters can be used to replicate the incentive spirometer [18]. While an option for improving data capture, the digital flowmeter can be clinically infeasible in most practices due to its complexity and cost. Collecting data directly from incentive spirometers falls into a technology category called the "Internet-of-Things" [19]. There have been great advances in the miniaturization of computing devices and the evolution of the "Internet-of-Things" into mainstream health devices [20,21]. As a result, wearable technologies and web-based platforms are capturing more clinical data now than ever before [22,23]. These technologies give clinicians and researchers access to otherwise inaccessible patient data and the ability to investigate new data interactions, such as enabling patients to exercise with wearable device hardware (eg, chest monitors and watches) and incorporating and communicating these data with patient physiologic and movement data [24,25]. The goal was to use this technology for the incentive spirometer.

We hypothesize an add-on device can accurately measure flow and volume data from a common incentive spirometer. This paper describes the creation and testing of an incentive spirometer add-on device to measure flow and volume data. An Internet-of-Things approach was adopted to enable this device to work with an existing incentive spirometer to capture physiological data and communicate externally on a closed private internet connection using hospital Wi-Fi and Bluetooth technologies. Captured data were tested for use by developing a classification algorithm to determine valid from invalid breaths.

Methods

Ethics Approval

Institutional review board approval was obtained for this prospective exploratory study on April 15, 2021 (HUM00196543, University of Michigan, Ann Arbor). Informed consent was waived as participants were limited to authors within this study, and interventions for this study are limited to existing approved uses of the incentive spirometer.

Add-On Device Creation

An add-on device was designed and created for use with the Hudson RCI Voldyne 5000 Volumetric Exerciser (Teleflex Medical) incentive spirometer (Figures 1-3 and Multimedia Appendix 1) and composed of photoelectric reflective infrared optical sensors (Xingyheng) positioned lateral to the spirometer volume and flow columns. Ten sensors were placed along the spirometer volume column with each sensor along a 500 mL marking on the incentive spirometer device (500-5000 mL). Three sensors were placed along the flow column corresponding to the middle of flow spirometer markings ("best," "better," and "good"). Sensors were connected to a microcontroller by a solderless breadboard, using an ESP32 development board (Dorhea), a low-cost low-power system-on-a-chip microcontroller with integrated Wi-Fi and dual-mode Bluetooth wireless communication capability. Components were soldered on a breadboard with breadboard jumper wires and resistors connecting components. Interactions between the microcontroller and the sensors were coded using C++ (Bell Laboratories of American Telephone and Telegraph). Components were encased in a 3D-printed base situated beneath the incentive spirometer with a layer of plexiglass fixed overtop.

Data were sent directly to an iOS application running on an iPad Pro (12.9″ and 10.2″, fifth generation, Apple) through ESP32 Bluetooth as well as stored on the add-on device in a 32 gigabyte microsecure digital card (Kootion) and card reader module (HiLetGo). A clock module (Melife) time-stamped data. Further interconnection was made possible using an analog digital multiplexer breakout board (Xie QianJin) and included a 3.7-V 2400-mAh rechargeable lithium battery (Akzytue) and charging module (MakerFocus) to power the device. An alternative design for the add-on base can be found in Multimedia Appendix 2.

Figure 1. The incentive spirometer piston positions were read using reflective infrared optical sensors (1) and relayed to an ESP32 in the base of the add-on device (2). The device both stored the data internally and used Bluetooth technology to transmit spirometer data to an Apple iPad (3). Data were then transmitted from the iPad to dedicated servers for further data processing and storage (4).

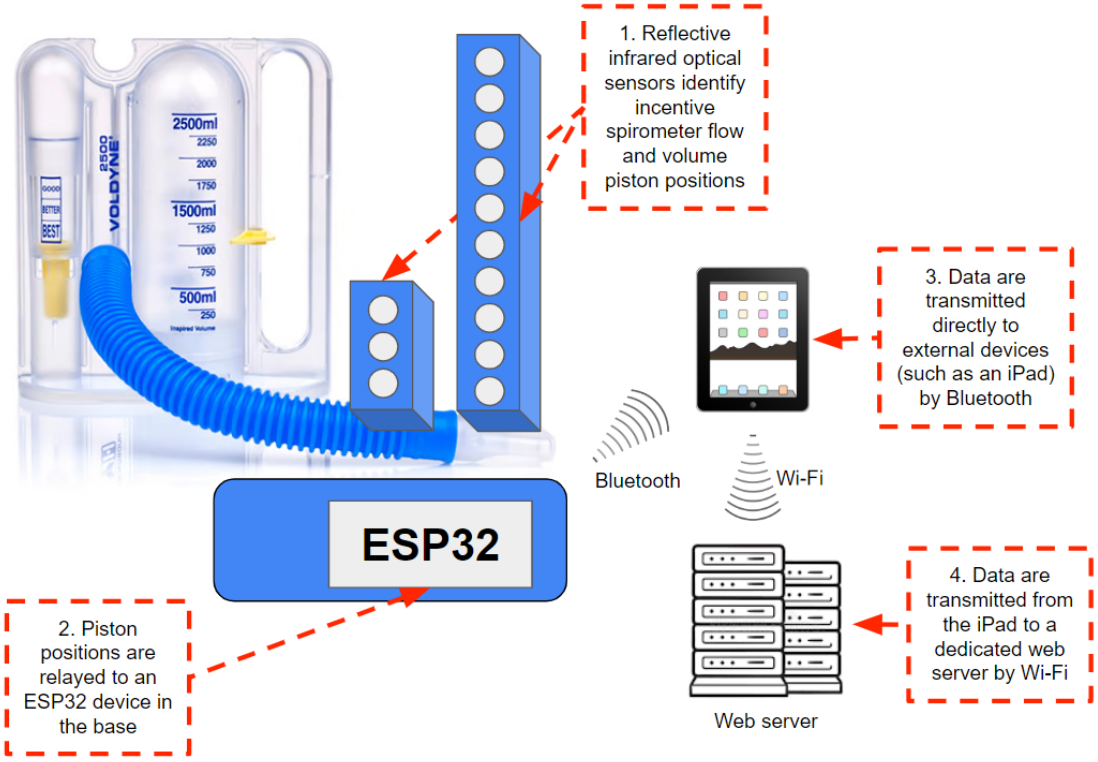


Figure 2. Ten photoelectric reflective infrared optical sensors were placed along the spirometer volume column with each sensor along a 500 mL marking on the incentive spirometer device (500-5000 mL); 3 sensors placed along the flow column. An on/off switch is featured on the front of the device. An ESP32 development board, rechargeable lithium battery, and real-time clock module are labeled. Connections between components were made with breadboard jumper wires and resistors, and all components are encased in a 3D-printed base with a layer of plexiglass overtop. A microsecure digital card reader module is in the back of the base (not shown).

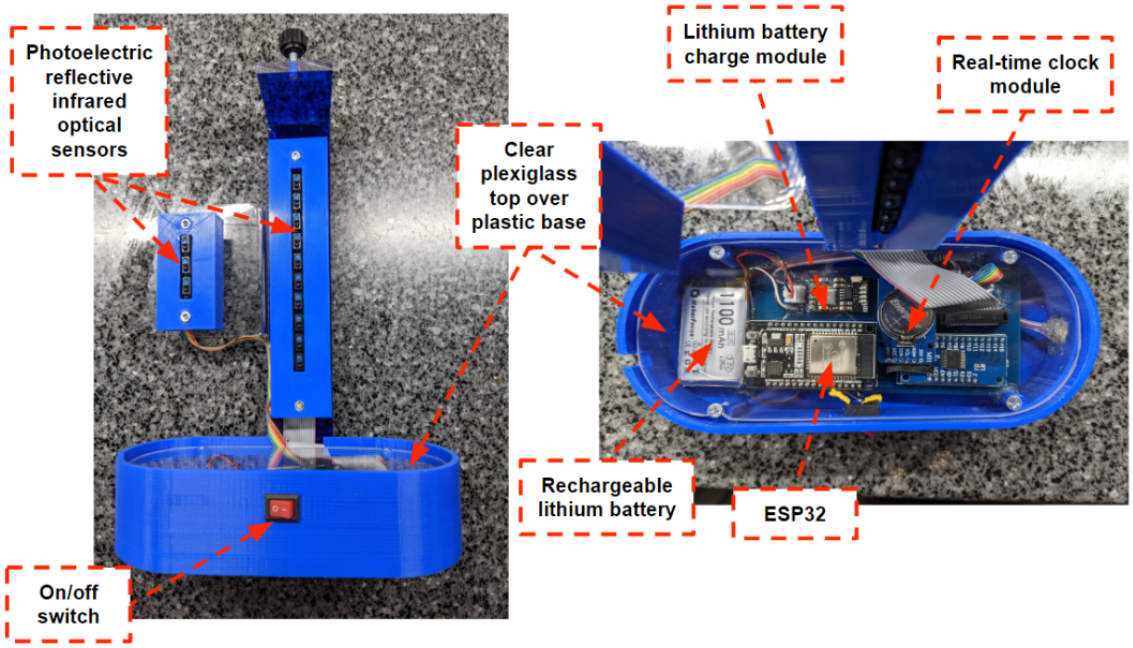


Figure 3. The Hudson RCI Voldyne 5000 Volumetric Exerciser (Teleflex Medical) incentive spirometer is situated over the base and locked into place with an overlying base lip and a screw on top of the 3D-printed posterior arm. A charging port is located in the base (right picture).

Software and Video Game Creation

An iOS video game and analytics application was built using Unity and C# (Microsoft). A web server receives the data and handles application programming interface (API) requests written in Java (Sun Microsystems). The ESP32 reads the sensor data using custom C++ code and sends processed data to an Apache web server via Wi-Fi through an iPad device connected via Bluetooth (Figure 1). Python (Python Software Foundation) applications were exposed by the web server, and downstream applications were networked with the server through an API to allow data collection and further use. A web server receives the data and handles API requests. A video game was developed specifically for use with the incentive spirometer and add-on device serving as a controller, designed for use with an Apple iPad Pro.

Add-On Device Sensor Testing

Two investigators (AS and MLB) tested the add-on device without crossover using the Hudson RCI Voldyne 5000 Volumetric Exerciser incentive spirometer. In these tests, 5 user breaths were attempted at each of the 10 volume sensor positions (100-5000 mL) for a total of 50 breaths per user. The volume goal was to get the top of the volume piston to the desired volume marking, while the flow goal was to get the top of the flow bobbin to the middle of the desired flow marking. Flow readings were also tested with every volume test with an additional 5 flow tests at each of the 3 flow sensor positions. To attain the desired flow or volume piston levels, users were allowed to breathe through or tilt the incentive spirometer device.

A single investigator (MLB) tested flow and volume sensor ranges with the incentive spirometer connected to the Puritan-Bennet PTS-2000 Ventilator Analyzer Tester (Mallinckrodt) digital flowmeter. In this testing, the investigator attempted increasing volume and flow values to identify sensor ranges. Volume measurements from the flowmeter were corrected using the body temperature, pressure, water vapor saturated method (correcting for body temperature [37 °C], ambient pressure, and gas saturated with water vapor). The flowmeter volume was calculated to be 183.75 mL by calibrating the flowmeter readings precisely at the 1000 mL display level of the Voldyne 5000, averaged across 10 breaths. This volume was subtracted from each raw volume reading from the flowmeter. Flowmeter testing was attempted with a minimum of 25 breaths around each flow and volume sensor level, independently, and the volume achieved from the flowmeter and the sensor level attained from the iPad were recorded. Due to limitations continually achieving breath volumes above 3250 mL (sensor level 7), additional testing using the 2500 mL volume Hudson RCI Voldyne 2500 Volumetric Exerciser (Teleflex Medical) incentive spirometer was conducted. The add-on device was created with each sensor aligning to the 500 mL markings of the Voldyne 5000. When using the Voldyne 2500, the sensors were near but incompletely aligned to the 250-mL markings.

Breath Algorithm Development and Testing

While sensors in the add-on device can detect where a piston is located, they cannot determine if a breath was conducted to achieve the sensor level. It is important to be able to accurately identify when a breath moves the incentive spirometer piston

and bobbin as opposed to tilting the incentive spirometer causing them to fall to the desired level. To aid in distinguishing a breath attempt, an algorithm was developed in Python (version 3.1) to process and classify each user's breath using flow and volume data over time. This algorithm reads spirometer log data into a pandas data frame and parses data by identifying the start and end of breaths. Breath start was identified by a zero flow value that precedes a positive flow value, while breath end is identified by 2 consecutive zero flow values. Once identified, each breath is classified as "valid" or "invalid." A valid breath must have the following volume criteria: volume starts at 0, increases within 1.5 seconds of breath start, >0 throughout the breath, does not decrease while there is positive flow, and the length of the breath is between 2 and 15 seconds. Breaths that do not meet these requirements were deemed invalid.

To test this algorithm, separate from device testing, 2 investigators (AS and MLB) each attempted to create 5 valid and 5 invalid breaths at each volume sensor level. These breath data were processed through the classification algorithm and evaluated. Invalid breaths were created by manually tipping the incentive spirometers or starting breaths at a starting position >0. All data validation was retrospectively validated using expert opinion from one of 2 users (AS and MLB).

Statistical Analysis

Summary and descriptive statistics were calculated for collected data using basic statistical techniques to assess models created for breath analysis including accuracy (as defined by user identification as gold standard), as well as mean, median, and SD for volume and flow measurements in sensor range testing.

Results

Add-On Device Sensor Testing

An add-on incentive spirometer device was created to independently measure real-time spirometer flow and volume piston positions (Figures 1-3). From 2 investigators, 99 of 100 volume readings were measured at the correct corresponding volume sensor (accuracy 99%, SD 1.4%). Only a single volume reading was inaccurate, failing once to capture at the lowest volume sensor (500 mL). All 130 flow bobbin readings corresponded to the correct flow sensor (100% accuracy).

Volume and flow sensor ranges were determined using a digital flowmeter in line with the 5000 and 2500 mL incentive spirometers independently (Figure 4 and Tables 1 and 2). Add-on sensors were designed with placement at every 500 mL marking on the Voldyne 5000 incentive spirometer. These sensor placements were not adjusted for the 2500 mL incentive spirometer testing. Investigator breath limitations were reached at volume levels approaching 4000 mL. Between the 2 devices tested, median and average volumes were within 7.5% (SD 6%) of target volume sensor levels. Maximum triggering values seldom exceeded the intended sensor level, showing a good correlation to placement. Flow levels corresponding to "best," "better," and "good" levels differed significantly between devices with the 5000 mL flow values more than double the 2500-mL incentive spirometer at each level.

Figure 4. User testing of add-on device connected to a digital flowmeter for both the Voldyne 5000 (A) and the Voldyne 2500 (B) incentive spirometers. Circles represent individual breath attempts with digital flowmeter volume readings (L, x-axis) and add-on device sensor iPad readings (y-axis). Sensor levels were designed to be placed at 500 mL levels of the Voldyne 5000 incentive spirometer and closely aligned to 250 mL markings of the Voldyne 2500.

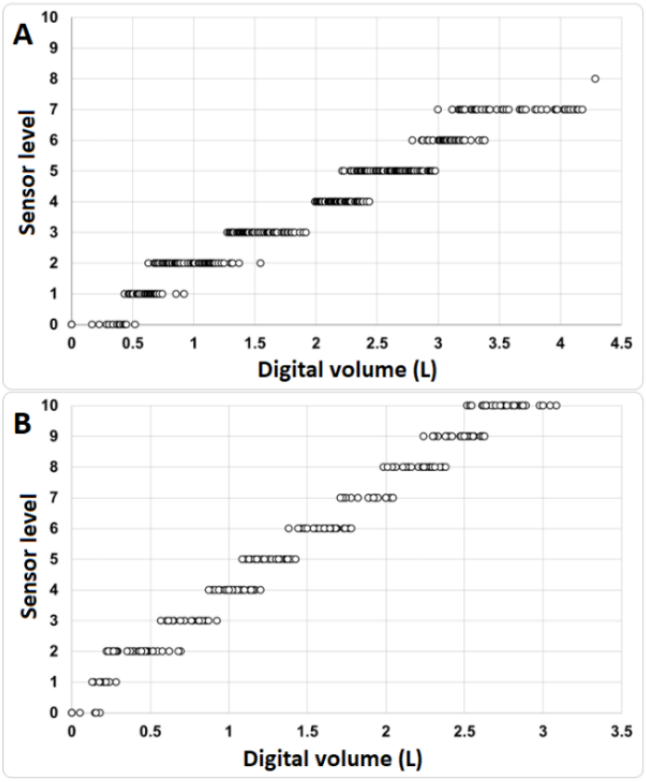


Table 1. Volume sensor ranges.[a]

Number	Breaths, n	Volume (L)			
		Mean (SD)	Median (IQR)	Minimum	Maximum
Voldyne 5000 volume sensor					
1	43	0.61 (0.10)	0.62 (0.54-0.67)	0.43	0.92
2	71	0.98 (0.20)	0.99 (0.81-1.13)	0.63	1.55
3	74	1.55 (0.18)	1.54 (1.40-1.68)	1.27	1.92
4	60	2.20 (0.12)	2.17 (2.11-2.28)	1.99	2.44
5	88	2.59 (0.20)	2.59 (2.42-2.76)	2.21	2.98
6	37	3.09 (0.13)	3.10 (3.02-3.16)	2.79	3.38
7	48	3.59 (0.35)	3.53 (3.29-3.96)	3.00	4.18
8	1	4.29 (N/A[b])	4.29 (N/A)	4.29	4.29
9	0	N/A (N/A)	N/A (N/A)	N/A	N/A
10	0	N/A (N/A)	N/A (N/A)	N/A	N/A
Voldyne 2500 volume sensor					
1	13	0.21 (0.04)	0.22 (0.18-0.23)	0.13	0.28
2	35	0.41 (0.13)	0.43 (0.29-0.48)	0.22	0.69
3	21	0.73 (0.11)	0.72 (0.64-0.82)	0.57	0.92
4	29	1.03 (0.09)	1.02 (0.99-1.09)	0.87	1.20
5	23	1.26 (0.10)	1.27 (1.18-1.34)	1.09	1.43
6	24	1.60 (0.11)	1.61 (1.53-1.67)	1.38	1.78
7	16	1.89 (0.11)	1.91 (1.81-1.96)	1.71	2.05
8	25	2.21 (0.11)	2.24 (2.13-2.28)	1.99	2.38
9	26	2.47 (0.11)	2.49 (2.40-2.55)	2.24	2.63
10	37	2.75 (0.14)	2.74 (2.64-2.85)	2.52	3.09

[a]Results from investigations of increasing volume sensor readings from the Voldyne 2500 and Voldyne 5000 incentive spirometers. The number of breaths (n) at each sensor level (1-10) and the average, median (IQR), minimum, and maximum readings were obtained from a digital flowmeter in liters (L).

[b]N/A: not applicable.

Table 2. Flow sensor ranges.[a]

	Breaths, n	Flow (L/minute)			
		Mean (SD)	Median (IQR)	Minimum	Maximum
Voldyne 5000 flow sensor					
Best	30	26.36 (5.56)	26.04 (22.51-27.76)	19.21	40.21
Better	30	44.31 (5.49)	43.56 (40.44-48.25)	36.10	56.69
Good	31	57.75 (7.74)	58.24 (53.99-61.27)	39.36	71.38
Voldyne 2500 flow sensor					
Best	30	12.24 (2.36)	12.31 (10.69-13.12)	8.43	19.63
Better	30	18.70 (1.92)	19.08 (17.45-19.56)	15.07	22.60
Good	31	28.45 (3.34)	28.13 (25.93-29.96)	22.61	38.30

[a]Results from investigations of increasing flow sensor readings from the Voldyne 2500 and Voldyne 5000 incentive spirometers. The number of breaths (n) at each sensor level ("best," "better," and "good") and the average, median (IQR), minimum, and maximum readings were obtained from a digital flowmeter in liters per minute (L/minute).

Breath Algorithm Development and Testing

After data collection, an algorithm was developed to classify breath data to determine when the spirometer was actively used. This algorithm works by classifying each individual breath using criteria to determine their validity as described in the Methods section. Independent of sensor testing, 2 members of the study team attempted 5 valid and 5 invalid breaths at each of ten 500-mL volume levels (500-5000 mL). Investigator limitations in achievable volumes resulted in a total of 65 valid breaths and 100 invalid breaths. The breath classification algorithm resulted in a 100% sensitivity and a 99% specificity for the classification of "valid" versus "invalid" user breaths. A single valid reading was inappropriately classified by the algorithm, occurring at the 500 mL volume sensor level. Example breath algorithm results are shown in Figure 5.

To investigate the downstream applicability of breath data, a video game was developed using Unity software, Apple iOS, with a single final application installed on Apple iPad devices (Figure 6). The game was a Kirby (Nintendo)-based side scroller, where the character would approach an obstacle and traverse the obstacle after a successful breath. This game was developed specifically for use with the incentive spirometry device with game-play centered around proper use [7,8]. The incentive spirometer with an add-on device was successfully used to control the created game. A video showing the game played in real time with an incentive spirometer device can be found in Multimedia Appendix 3.

Figure 5. Example results from the breath algorithm used to classify "valid" from "invalid" breath attempts. The y-axis represents sensor levels for flow and volume readings in the incentive spirometer device: flow (0-3; corresponding to none: 0; "good": 1; "better": 2; and "best": 3); volume (0-10, representing each 500 mL increment from 0 to 5000 mL). The x-axis represents the time from breath start in seconds.

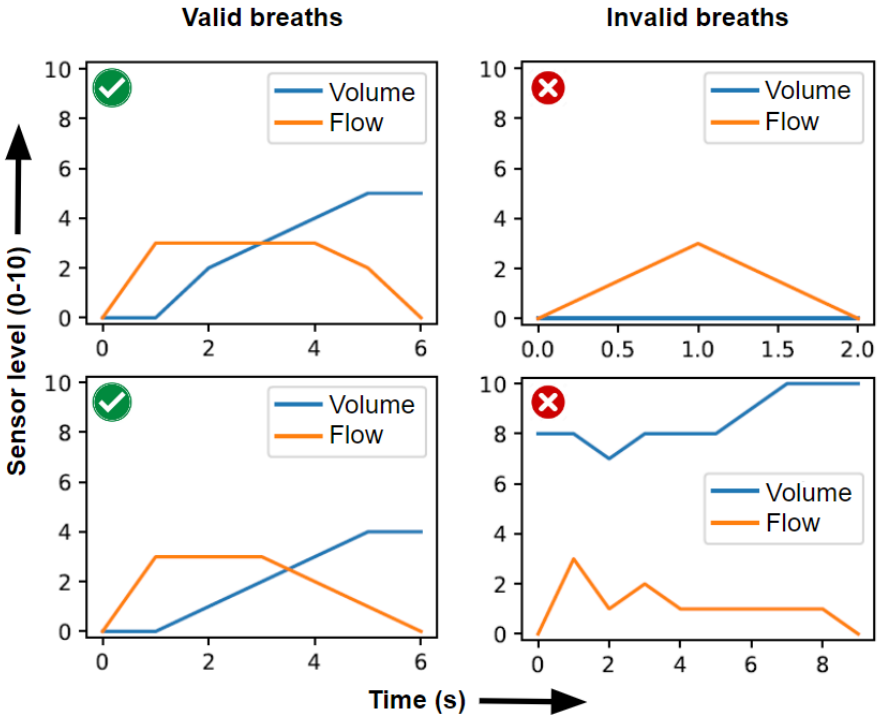
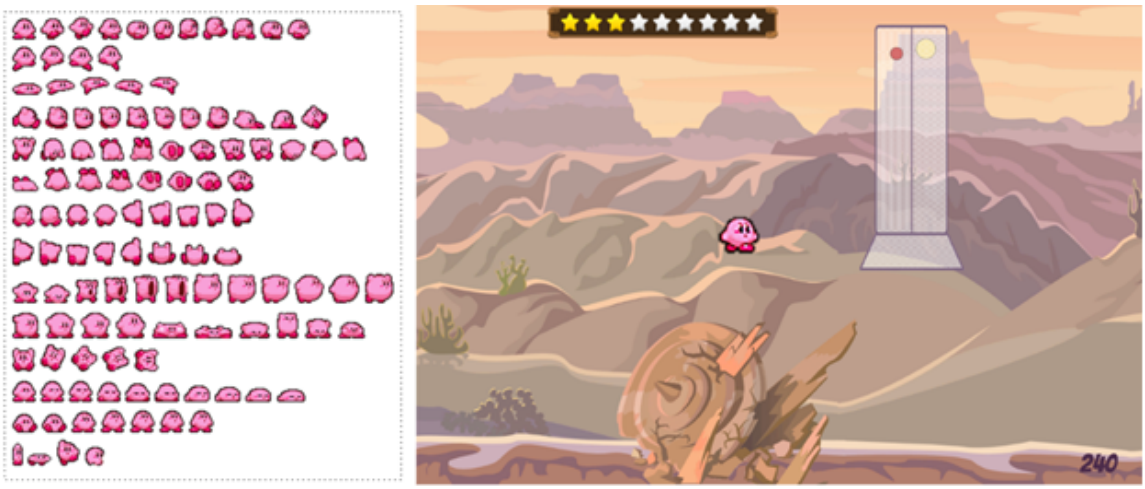


Figure 6. A video game was specifically designed for use with the incentive spirometer and add-on device based on the Kirby Nintendo character. A sprite sheet for the character development within the video game is shown on the left with a screenshot of the video game shown on the right. Multimedia Appendix 3 shows the game being controlled using an incentive spirometer.



Discussion

Principle Findings

In this study, an add-on device was created to allow the capture of incentive spirometry data with high accuracy (99%, SD 1.4% volume and 100% flow when tested at incentive spirometer level markings). Maximum triggering values were rarely exceeding intended sensor levels, proving excellent differentiation of levels in the design. Furthermore, while sensor placements were designed around the Voldyne 5000 incentive spirometer, the similarly shaped Voldyne 2500 performed well in testing without sensor modifications, suggesting use of the add-on device to similar incentive spirometers without significant redesign. While range testing was not intended to study volume sensor accuracy, between the 2 spirometers tested, median and average volumes were within 7.5% (SD 6%) of target volume levels, with the worst individual readings at lower volumes (250-500 mL sensors) with readings maximally 40 to 120 mL off target, respectively.

To differentiate quality breath attempts from errors, a classification breath algorithm was developed. The device allowed identification of start and end of breath attempts, valid from invalid breaths assessment, maximum flow, maximum volume, and volume or time ascents and descent calculations. The breath classification algorithm used device data to discern valid versus invalid breath attempts, showing 100% sensitivity and 99% specificity. Identifying valid breath data from noise is critically important for downstream applications and, while the classification algorithm yielded results, further efforts for improvement could be made using additional rule-based systems or machine learning algorithms.

This system is intended to enable clinical providers access to previously inaccessible spirometry data to improve spirometer instruction and use protocols, study patient compliance, and incentivize use. Using an add-on device similar to what was created in this study would increase the granularity of spirometer compliance data and could be used to provide insight into proper incentive spirometer use. Additionally, the add-on device can allow focused interventions to improve adherence. Reminder notifications alone have been shown to improve incentive spirometer use. In one study, an add-on use-tracking device was equipped with a bell that sounded for up to 2 minutes every hour as a reminder for the patient to use their incentive spirometer [12]. This study demonstrated that patients using the reminder device had a greater number of mean daily inspiratory breaths and a percentage of recorded hours with an inspiratory breath event. More importantly, patients with the reminder displayed significantly lower mean atelectasis severity scores measured by chest radiography, reduced median postoperative and intensive care unit length of stay, and had a lower mortality rate at 6 months. These findings support postoperative incentive spirometer use and show effectiveness of a simple intervention to improve incentive spirometer adherence.

Gamification

To demonstrate the real-time use of incentive spirometry data, iPad video game was created to be controlled by the add-on incentive spirometer device. In testing, the game showed no appreciable lag and continued connectivity during use, proving electronic spirometer data collected by the add-on device to be capable of real-time gamification applications. Gamification of medical interventions is an exciting concept for improving medical care adherence. Breathing games for the incentive spirometer is a familiar idea, with one group brainstorming a suite of games for asthmatics focusing on breathing metaphors as incentives for spirometer use [26], while others developed video games to incentivize breathing exercises and peak expiratory flow using digital flowmeters [18,27]. There exists an abandoned patent around the use of the incentive spirometer as a game controller [28] and an active patent around use of the flowmeter in video games [29], further supporting the popularity of the idea. The device in this study was created leveraging recent technologies and focusing design on clinical care use. Using the add-on device, as opposed to a digital flowmeter, maintains the current use of incentive spirometers in medical settings to allow native data capture. Potential reuse of the add-on device limits additional costs such as those incurred using digital flowmeters.

Limitations

Limitations to the add-on device design include contamination risks, costs, and technical and workflow implementation constraints. First, while completely enclosed, the add-on device is designed to be reused and carries the risk of infection—an especially important consideration in a respiratory pandemic such as COVID-19. To improve sterility, the device was enclosed in plastic, and sensors were placed behind the incentive spirometer, removing the need to expose the base to breaths from sensors placed below. Improvements can be made to close remaining gaps in the plastic encasing to further enclose the device and allow cleaning like an iPad or PlayStation controller, commonly used in the hospital settings. Second, routine incentive spirometer postoperative care has been estimated to carry a US $107.36 cost per patient above material cost of the spirometer device, which totals US $1.04 billion in total US annual costs [30]. This is a significant cost for a device with sparse evidence around use and poor patient compliance. While incentive spirometer devices are not reusable from patient to patient, the add-on device was designed to be reusable, lowering its effective cost. Overall, the add-on device carries a material price of approximately US $150 per unit (Multimedia Appendix 4), which could be reduced by bulk purchasing and further investigation into alternative individual components. While add-on device cost is an addition to the already significant price, identifying compliance and improving adherence will facilitate improved use and function of the incentive spirometer. Further studies of the incentive spirometer are required to investigate the prevention of postoperative breathing complications and their associated health care costs. These studies are dependent upon accurate compliance data and would benefit from the capabilities of the add-on device. Third, there exist integration and maintenance requirements of the add-on device, and it may be feasible for use only in hospital systems with existing technical support structures. The device was designed to minimize technical requirements, but more investigation is required. Future studies are required to trial the add-on device

in clinical settings and test for improved adherence using strategies such as gamification compared to traditional incentive spirometry.

Conclusions

Incentive spirometers are routinely used in hospital settings, specifically in postoperative clinical care, but recommendations for proper routine use lack thorough investigation due to a general lack of data on device use. Creating a low-cost, effective, and reusable add-on device for the incentive spirometer allows native collection of previously inaccessible incentive spirometer compliance data. These data can facilitate research into incentive spirometer use to guide clinical care, incentivize adherence, and draw conclusions about the clinical effectiveness of the incentive spirometer.

Acknowledgments

The authors gratefully acknowledge the valuable contribution of Mikele Garrett and dedicate this paper in his memory. Mikele was an incredible person and teammate who passed away prior to the completion of this paper. He was critical in device development, overseeing and creating the add-on device designs. The authors also acknowledge the Adult Clinical Research Group at the University of Michigan for the valuable contribution to protocol development and final paper review; the University of Michigan Hacks with Friends competition for initial idea development; and the University of Michigan Software Development for Accessibility course EECS 495 and its students for video game ideas and creation, Amanda Bahamonde for her assistance with preparing this paper for submission, Ruth Cassidy for statistical support, and Kent Miller for access to the digital flowmeter used in device testing. Partial funding can be attributed to the Department of Anesthesiology, University of Michigan Medical School, the Society for Technology in Anesthesia 2021 Neurowave Research Grant, and the National Institute for General Medical Sciences of the National Institutes of Health (MLB; grant T32GM103730).

Data Availability

Deidentified data were collected from spirometer use by the study team. Data were collected using the incentive spirometer, add-on device, and digital flowmeter. Data sets generated and analyzed within this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

This work has been declared through the University of Michigan Innovation Partnerships.

Multimedia Appendix 1
Add-on device design and schematics.
[DOCX File , 294 KB - biomedeng_v8i1e46653_app1.docx]

Multimedia Appendix 2
Alternative design schematic for add-on device.
[DOCX File , 436 KB - biomedeng_v8i1e46653_app2.docx]

Multimedia Appendix 3
Video displaying an investigator (TM) using the incentive spirometer with the add-on device to control the created video game.
[MP4 File (MP4 Video), 7764 KB - biomedeng_v8i1e46653_app3.mp4]

Multimedia Appendix 4
Add-on device cost.
[DOCX File , 13 KB - biomedeng_v8i1e46653_app4.docx]

References

1. Rock P, Rich PB. Postoperative pulmonary complications. Curr Opin Anaesthesiol 2003;16(2):123-131. [doi: 10.1097/00001503-200304000-00004] [Medline: 17021450]
2. Fisher BW, Majumdar SR, McAlister FA. Predicting pulmonary complications after nonthoracic surgery: a systematic review of blinded studies. Am J Med 2002;112(3):219-225. [doi: 10.1016/s0002-9343(01)01082-8] [Medline: 11893349]
3. Gatti G, Cardu G, Lusa AM, Pugliese P. Predictors of postoperative complications in high-risk octogenarians undergoing cardiac operations. Ann Thorac Surg 2002;74(3):671-677 [FREE Full text] [doi: 10.1016/s0003-4975(02)03741-4] [Medline: 12238822]
4. Carter AR, Sostman HD, Curtis AM, Swett HA. Thoracic alterations after cardiac surgery. AJR Am J Roentgenol 1983;140(3):475-481 [FREE Full text] [doi: 10.2214/ajr.140.3.475] [Medline: 6600537]

5. Bartlett RH, Krop P, Hanson EL, Moore FD. Physiology of yawning and its application to postoperative care. Surg Forum 1970;21:222-224. [Medline: 4936971]
6. Bartlett RH, Gazzaniga AB, Geraghty TR. Respiratory maneuvers to prevent postoperative pulmonary complications. A critical review. JAMA 1973;224(7):1017-1021. [Medline: 4574097]
7. Franklin E, Anjum F. Incentive Spirometer and Inspiratory Muscle Training. Treasure Island, FL: StatPearls Publishing; 2022.
8. Pullen RL. Teaching bedside incentive spirometry. Nursing 2003;33(8):24. [doi: 10.1097/00152193-200308000-00018] [Medline: 12918478]
9. Martin TJ, Patel SA, Tran M, Eltorai AS, Daniels AH, Eltorai AEM. Patient factors associated with successful incentive spirometry. R I Med J (2013) 2018;101(9):14-18. [Medline: 30384513]
10. Strickland SL, Rubin BK, Drescher GS, Haas CF, O'Malley CA, Volsko TA, American Association for Respiratory Care, Irving, Texas. AARC clinical practice guideline: effectiveness of nonpharmacologic airway clearance therapies in hospitalized patients. Respir Care 2013;58(12):2187-2193 [FREE Full text] [doi: 10.4187/respcare.02925] [Medline: 24222709]
11. Freitas RFS, Soares GOS, Cardoso RC, Atallah Á. Incentive spirometry for preventing pulmonary complications after coronary artery bypass graft. Cochrane Database Syst Rev 2012;2012(9):CD004466 [FREE Full text] [doi: 10.1002/14651858.CD004466.pub3] [Medline: 22972072]
12. Eltorai AEM, Baird GL, Eltorai AS, Healey TT, Agarwal S, Ventetuolo CE, et al. Effect of an incentive spirometer patient reminder after coronary artery bypass grafting: a randomized clinical trial. JAMA Surg 2019;154(7):579-588 [FREE Full text] [doi: 10.1001/jamasurg.2019.0520] [Medline: 30969332]
13. Amin R, Alaparthi GK, Samuel SR, Bairapareddy KC, Raghavan H, Vaishali K. Effects of three pulmonary ventilation regimes in patients undergoing coronary artery bypass graft surgery: a randomized clinical trial. Sci Rep 2021;11(1):6730 [FREE Full text] [doi: 10.1038/s41598-021-86281-4] [Medline: 33762655]
14. Malik PRA, Fahim C, Vernon J, Thomas P, Schieman C, Finley CJ, et al. Incentive spirometry after lung resection: a randomized controlled trial. Ann Thorac Surg 2018;106(2):340-345 [FREE Full text] [doi: 10.1016/j.athoracsur.2018.03.051] [Medline: 29702071]
15. Restrepo RD, Wettstein R, Wittnebel L, Tracy M. Incentive spirometry: 2011. Respir Care 2011;56(10):1600-1604 [FREE Full text] [doi: 10.4187/respcare.01471] [Medline: 22008401]
16. Narayanan ALT, Hamid SRGS, Supriyanto E. Evidence regarding patient compliance with incentive spirometry interventions after cardiac, thoracic and abdominal surgeries: a systematic literature review. Can J Respir Ther 2016;52(1):17-26 [FREE Full text] [Medline: 26909010]
17. Eltorai AEM, Baird GL, Eltorai AS, Pangborn J, Antoci V, Cullen HA, et al. Incentive spirometry adherence: a national survey of provider perspectives. Respir Care 2018;63(5):532-537 [FREE Full text] [doi: 10.4187/respcare.05882] [Medline: 29362219]
18. Lange B, Flynn S, Rizzo A, Bolas M, Silverman M, Huerta A. Breath: a game to motivate the compliance of postoperative breathing exercises. 2009 Presented at: 2009 Virtual Rehabilitation International Conference; June 29, 2009-July 02, 2009; Haifa, Israel. [doi: 10.1109/icvr.2009.5174212]
19. Ashton K. That ‘internet of things’ thing. Radio Freq Identif J 2009;22(7):97-114 [FREE Full text]
20. Yin Y, Zeng Y, Chen X, Fan Y. The internet of things in healthcare: an overview. J Ind Inf Integration 2016;1:3-13. [doi: 10.1016/j.jii.2016.03.004]
21. Dimitrov DV. Medical internet of things and big data in healthcare. Healthc Inform Res 2016;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
22. Fekr AR, Radecka K, Zilic Z. Design and evaluation of an intelligent remote tidal volume variability monitoring system in E-health applications. IEEE J Biomed Health Inform 2015;19(5):1532-1548. [doi: 10.1109/JBHI.2015.2445783] [Medline: 26087508]
23. Janidarmian M, Fekr AR, Radecka K, Zilic Z. A comprehensive analysis on wearable acceleration sensors in human activity recognition. Sensors (Basel) 2017;17(3):529 [FREE Full text] [doi: 10.3390/s17030529] [Medline: 28272362]
24. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
25. Park T, Hwang I, Lee U, Lee SI, Yoo C, Lee Y, et al. ExerLink: enabling pervasive social exergames with heterogeneous exercise devices. 2012 Presented at: MobiSys '12: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services; June 25-29, 2012; Low Wood Bay Lake District, UK p. 15-28. [doi: 10.1145/2307636.2307639]
26. van Delden R, Bos DPO, de With AJV, Vogel K, Klaassen R, Zwart N, et al. SpiroPlay, a suite of breathing games for spirometry by kids & experts. 2020 Presented at: CHI PLAY '20: Proceedings of the Annual Symposium on Computer-Human Interaction in Play; November 2-4, 2020; Virtual Event, Canada p. 400-413. [doi: 10.1145/3410404.3414223]
27. Chelabi K, Balli F, Bransi M, Gervais Y, Marthe C, Tse SM. Validation of a portable game controller to assess peak expiratory flow against conventional spirometry in children: cross-sectional study. JMIR Serious Games 2021;9(1):e25052 [FREE Full text] [doi: 10.2196/25052] [Medline: 33512326]

28. Elefteriades J, Bango J, Dziekan M. Game-based incentive spirometer and a method of quantifying and recording performance. US Patent. URL: https://patentimages.storage.googleapis.com/b6/69/8c/ea18a2f81ebf10/US20130303930A1.pdf [accessed 2022-12-28]
29. Luttrell RS. Respiratory therapy instrument offering game-based incentives, training, and telemetry collection. US Patent. 2019. URL: https://patentimages.storage.googleapis.com/60/ba/5c/51478d0167ed70/US10506950.pdf [accessed 2022-12-28]
30. Eltorai AEM, Baird GL, Pangborn J, Eltorai AS, Antoci V, Paquette K, et al. Financial impact of incentive spirometry. Inquiry 2018;55:46958018794993 [FREE Full text] [doi: 10.1177/0046958018794993] [Medline: 30175643]

Abbreviations

API: application programming interface

Edited by T Leung; submitted 20.02.23; peer-reviewed by P McCormick, J Mandel; comments to author 16.06.23; revised version received 07.07.23; accepted 27.07.23; published 07.09.23.

Please cite as:
Burns ML, Sinha A, Hoffmann A, Wu Z, Medina Inchauste T, Retsky A, Chesney D, Kheterpal S, Shah N
Development and Testing of a Data Capture Device for Use With Clinical Incentive Spirometers: Testing and Usability Study
JMIR Biomed Eng 2023;8:e46653
URL: https://biomedeng.jmir.org/2023/1/e46653
doi:10.2196/46653
PMID:38875693

Original Paper

A Wearable Vibratory Device (The Emma Watch) to Address Action Tremor in Parkinson Disease: Pilot Feasibility Study

Alissa Pacheco[1], MOT, OTR/L; Tempest A van Schaik[2], PhD; Nadzeya Paleyes[2], BEcon, BEng; Miguel Blacutt[1], EdM; Julio Vega[3], PhD; Abigail R Schreier[1], BA; Haiyan Zhang[2], MA; Chelsea Macpherson[1], PT, DPT, NCS; Radhika Desai[1], MS, PhD; Gavin Jancke[2], BSc; Lori Quinn[1], PT, EdD

[1]Teachers College, Columbia University, New York, NY, United States

[2]Microsoft Research, Cambridge, United Kingdom

[3]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Lori Quinn, PT, EdD
Teachers College
Columbia University
525 West 120th Street
New York, NY, 10027
United States
Phone: 1 2126783424
Email: lq2165@tc.columbia.edu

Abstract

Background: Parkinson disease (PD) is a neurodegenerative disease that has a wide range of motor symptoms, such as tremor. Tremors are involuntary movements that occur in rhythmic oscillations and are typically categorized into rest tremor or action tremor. Action tremor occurs during voluntary movements and is a debilitating symptom of PD. As noninvasive interventions are limited, there is an ever-increasing need for an effective intervention for individuals experiencing action tremors. The Microsoft Emma Watch, a wristband with 5 vibrating motors, is a noninvasive, nonpharmaceutical intervention for tremor attenuation.

Objective: This pilot study investigated the use of the Emma Watch device to attenuate action tremor in people with PD.

Methods: The sample included 9 people with PD who were assessed on handwriting and hand function tasks performed on a digitized tablet. Tasks included drawing horizontal or vertical lines, tracing a star, spiral, writing “elelelel” in cursive, and printing a standardized sentence. Each task was completed 3 times with the Emma Watch programmed at different vibration intensities, which were counterbalanced: high intensity, low intensity (sham), and no vibration. Digital analysis from the tablet captured kinematic, dynamic, and spatial attributes of drawing and writing samples to calculate mathematical indices that quantify upper limb motor function. APDM Opal sensors (APDM Wearable Technologies) placed on both wrists were used to calculate metrics of acceleration and jerk. A questionnaire was provided to each participant after using the Emma Watch to gain a better understanding of their perspectives of using the device. In addition, drawings were compared to determine whether there were any visual differences between intensities.

Results: In total, 9 people with PD were tested: 4 males and 5 females with a mean age of 67 (SD 9.4) years. There were no differences between conditions in the outcomes of interest measured with the tablet (duration, mean velocity, number of peaks, pause time, and number of pauses). Visual differences were observed within a small subset of participants, some of whom reported perceived improvement. The majority of participants (8/9) reported the Emma Watch was comfortable, and no problems with the device were reported.

Conclusions: There were visually depicted and subjectively reported improvements in handwriting for a small subset of individuals. This pilot study was limited by a small sample size, and this should be taken into consideration with the interpretation of the quantitative results. Combining vibratory devices, such as the Emma Watch, with task specific training, or personalizing the frequency to one’s individual tremor may be important steps to consider when evaluating the effect of vibratory devices on hand function or writing ability in future studies. While the Emma Watch may help attenuate action tremor, its efficacy in improving fine motor or handwriting skills as a stand-alone tool remains to be demonstrated.

(JMIR Biomed Eng 2023;8:e40433) doi:10.2196/40433

KEYWORDS

Parkinson's disease; action tremor; Emma Watch; vibration; haptic feedback; handwriting; drawing; spirals; hand function

Introduction

Parkinson disease (PD) is a progressive neurodegenerative disease caused by the degeneration of dopamine in the substantia nigra and striatum areas of the brain that results in a decrease in the ability to control and coordinate movement [1,2]. PD currently affects more than 6 million people worldwide [3] and the incidence of PD is expected to double in the next 2 decades [2]. The cause of PD appears to be multifactorial, with behavioral, environmental, genetic, and lifestyle factors playing a role [2]. PD is characterized by both motor and nonmotor symptoms, however, the cardinal features include rigidity, tremor, bradykinesia, and postural instability [4]. While many forms of tremor may be present in PD, action tremor, which includes postural, isometric, and kinetic tremor [5,6], occurs during active, voluntary movement and is an impairment that affects writing, hand function, activities of daily living, and quality of life [4,7]. Furthermore, people with PD experience psychosocial implications resulting from their tremor, including negatively impacted relationships, self-image, and overall well-being [8]. Action tremor is reported in 46% of individuals with Hoehn and Yahr Stages 1 and 2, and up to 93% of people in Stages 1-5 [7].

The pathophysiology of action tremor is uncertain but has a clear difference from other motor symptoms of PD [9] as it is thought to be modulated by nondopaminergic pathways and does not correlate with dopamine depletion in the striatum; rather, serotonin, noradrenaline, and acetylcholine appear to play more of a role [9]. It may be caused by oscillations that occur within internal sensorimotor feedback circuits during movement [10,11] or by abnormal activity within the basal ganglia and cerebello-thalamo-cortical circuit [9,12] Action tremor frequency usually displays a 1.5 Hz higher frequency than that of rest tremor, which is ~4-6 Hz [9]. Interventions for the management of action tremor have been limited. Deep brain stimulation to the subthalamic nucleus and globus pallidus have shown improvement in action tremor severity at 6 and 12 months post implantation, but is invasive, can lead to adverse events, and is not suitable for everyone [11]. Dopaminergic medications also have limited efficacy in improving action tremor-related motor dysfunction [5,8,9,13], leaving a pressing need to address these functional sequelae.

Equivocal findings have been reported for total body vibration to improve motor function in people with PD [14], however recent studies suggest targeted vibration methods may be beneficial [15]. High-frequency vibration stimulation (also known as haptic feedback), along with medication, have improved movement initiation, movement speed, precision, and decreased tremor for people with PD [15-17]. Vibration is a form of sensory stimulation that results in increased sensory input and activation of the muscle spindle fibers, and may improve neuromotor functions in people with sensorimotor deficits [18]. Providing high frequency vibration over the forearm activates the muscle spindles and interrupts the central nervous system's interpretation of the proprioceptive position of the limb in space, interpreting the vibration as sensory information and producing a muscle contraction [15].

Use of haptic feedback may reduce resting tremor and is considered safe and well tolerated when delivered in short durations via wearable devices, but its effects on action tremor has not been well studied [19,20]. The Emma Watch is a wrist-worn wearable device developed by Microsoft Research that provides constant high frequency vibration to each side of the wrist; preliminary findings suggest it may reduce movement speed and improve precision of performance in drawing and tracing tasks in people with PD [17]. However, efficacy of the Emma Watch on people with PD is unclear; therefore, this pilot study evaluated the use of the Emma Watch in people with PD who present with symptomatic and disruptive action tremor during handwriting, drawing, and hand function tasks.

Methods

Participants

In total, 9 people with PD were recruited for this pilot study, which began just prior to the COVID-19 pandemic. We had expected to enroll a total of 20 participants; however, human subjects research was suspended in New York City during the pandemic. The inclusion criteria were (1) formal diagnosis of PD from a neurologist and (2) the presence of action tremor in one or both hands, with a rating of >1 on the Movement Disorders Society–sponsored Revision of the Unified Parkinson's Disease Rating Scale part III item 3.16 (rating of tremor). Exclusion criteria were: (1) history of comorbid neurological conditions, that is, including stroke or other neurodegenerative disease, (2) acute orthopedic conditions on the dominant hand, (3) chronic orthopedic conditions affecting the ability to write, (4) implantation of a pacemaker or deep brain stimulator, or (5) inability or unwillingness of the participant or legal guardian to give written informed consent. Individuals were recruited from the community via flyers and neurologists from Columbia University Irving Medical Center were informed of our study and could refer their patients.

Ethical Considerations

This study was approved by the institutional review board at Teachers College, Columbia University (IRB#19-266) and all participants signed informed consent. Privacy and confidentiality standards were protected throughout the research study, and all study data collected were deidentified. Participants were compensated US $50 for their participation in this study.

Emma Watch Device

The Emma Watch is a lightweight watch-like device worn around the dominant wrist. The device uses 5 small linear resonant actuators, each with a 205 Hz vibration frequency and controlled by a driver with an auto resonance engine. The vibration's strength and modulation are controlled via a Microsoft Surface tablet app (Microsoft), which connects to the Emma Watch via Bluetooth. The vibration is initiated at the start of movement and delivered throughout the task. The linear

resonant actuators were run by a haptic driver at 100%, 64.7%, and 50% duty cycle, producing 3 vibration amplitudes: ~1.51 g (high intensity), ~0.38 g (low intensity), and 0 g (no vibration) modulated with a 500-millisecond/500-millisecond on/off vibration cycle in a counterbalanced order.

Study Design

Individuals participated in two 90-minute sessions, performed on different days, but only 1 day for each participant was analyzed for this pilot study. Participants were tested within 1 hour after administration of their regular disease-specific medication. All participants donned the Emma Watch, along with 3 wearable inertial measurement units (APDM Opal sensors; APDM Wearable Technologies), one on each wrist and their trunk. All assessments were video recorded using a GoPro camera (GoPro). Participants completed baseline clinical assessments, along with a series of handwriting and fine motor assessments, with the Emma Watch counterbalanced on 3 vibration conditions. The ~0.38 g (low) is a very mild vibration and acted as a control condition [21].

Assessments

We obtained demographic information from each participant, including age, gender, hand dominance, current list of medications, and education level. The effect of tremor on daily living activities was evaluated using the Bain and Finley Activities of Daily Living Scale [22] and self-reported hand function was evaluated using the Manual Ability Measure [23] (see Multimedia Appendix 1).

Tablet Analysis

Participants completed 3 repetitions of handwriting and drawing tasks using a stylus and a digitizing tablet (Microsoft Surface). Participants were instructed to conduct the tasks as quickly and accurately as possible. Tasks included drawing horizontal or vertical lines, tracing a star, spiral, writing “elelelel” in cursive, and printing a standardized sentence. Each task was completed 3 times at 3 different vibration intensities: with the Emma Watch counterbalanced at a high intensity, a low intensity (sham), and their baseline with no vibration. Digital analysis captured kinematic [24], dynamic, and spatial attributes of drawing and writing samples to calculate mathematical indices that quantify upper limb motor function. The tablet recorded the pen’s x- and y-position and timestamp without wires or other attachments. The Windows app stored and converted data into readable files for analysis.

Analysis of the tablet recordings was performed with Python 3.7 (Python Software Foundation). The pen’s position in the x- and y-direction was converted to Euclidian distance at each point. The median sampling frequency of the x- and y-coordinate data was 142 Hz. To ensure a constant sampling frequency, the distance was resampled to a constant 142 Hz, using linear interpolation. Pointwise velocity was calculated as pointwise Euclidian distance divided by pointwise time interval. Velocity data were smoothed using a 3.5-Hz cut off, 5th order, low-pass Butterworth filter to remove high frequency fluctuations. We used this cut off to quantify slow movements related to writing. Measures calculated were (1) duration (seconds) of each drawing task, defined as the time taken to complete the drawing, from stylus down to stylus up, including pauses between strokes; (2) number of pauses, where a pause is any lift off of the stylus while drawing; (3) pause duration (seconds), defined as the sum of all pause times; (4) mean velocity (pixels/second) of each drawing task, defined as the mean of each pointwise velocity; and (5) fluency, defined as the number of local maxima (peaks) in the velocity profile. These peaks were found by comparison of neighboring values without any threshold.

Accelerometry Analysis

Preprocessing of the APDM Opal Sensor data was carried out in MATLAB (MathWorks R2020A). Raw signals from the accelerometers were set to horizontal and vertical coordinates by the sensors’ preexisting algorithms, which use the magnetometers and gyroscopes to identify x-, y-, and z-reference positions. Right and left wrist streams were extracted and processed in MATLAB. Accelerometry streams, sampled at 128 Hz, were filtered through a 3.5 Hz cut off, zero-phase, low-pass Butterworth filter. This filtering profile is consistent with previous uses of APDM inertial measurement unit sensors in identifying anticipatory postural adjustments. The accelerometer and video data were synched through a series of claps, performed to identify the start of each drawing task. The start of a clap was designated if it met a power threshold of 0.17 in the accelerometry stream and was followed by 110 consecutive data points above the threshold.

The processing of the APDM Opal Sensor data was done using R (R Foundation for Statistical Computing) and Python (Python Software Foundation). Further, 32 metrics were calculated for each of the 3 axes of acceleration and jerk signals using *mhealthtools* [25] *R* package. These include the mean, complexity, mobility, roughness, rugosity, Shannon entropy of the frequency probability distribution, mean frequency, and the energy present in the twenty-four 0.5 Hz-bands between 2 Hz and 12 Hz. The signal’s fluency was computed in a similar way to the tablet drawings streams.

Statistical Analysis

Statistical tests were conducted using R (version 4.02). Linear mixed-effects models were used to compare means of task outcomes (duration, number of pauses, pause duration, mean velocity, and number of peaks) between vibration intensities (zero, low, and high) within the 5 tasks (rectangle, spiral, star, elelelel, and handwriting; see Table 1). Linear mixed effects models used vibration frequency as a categorical variable, generated a random intercept, and used the “*lme4*” package on R [26].

Table 1. Duration, velocity, and peaks for tablet tasks.

Task	No vibration, mean (SD)	Low intensity, mean (SD)	High intensity, mean (SD)
Spiral			
Duration (seconds)	38.8 (23.7)	37.2 (20.4)	37.1 (21.1)
Pause number	1.6 (2.8)	0.8 (1.2)	1.2 (2.0)
Pause time (seconds)	0.014 (0.02)	0.008 (0.01)	0.012 (0.02)
Scaled peaks (peaks/s)	2.2 (0.1)	2.2 (0.2)	2.2 (0.2)
Number of peaks	85.8 (50.6)	85.3 (52.1)	82.4 (48.1)
Mean velocity (pixels/s)	168.0 (65.1)	171.0 (73.6)	171.8 (71.2)
Handwriting			
Duration (seconds)	18.2 (4.9)	17.3 (5.0)	18.3 (5.7)
Pause number	25.2 (1.2)	24.6 (1.9)	24.4 (2.0)
Pause time (seconds)	0.2 (0.03)	0.2 (0.02)	0.2 (0.03)
Scaled peaks (peaks/s)	1.4 (0.2)	1.5 (0.2)	1.5 (0.2)
Number of peaks	25.2 (8.3)	24.9 (7.3)	26.5 (9.0)
Mean velocity (pixels/s)	201.5 (80.7)	195.3 (78.3)	195.1 (68.9)
APDM Spiral			
Acceleration			
Complexity	23.5 (3.7)	22.2 (3.4)	23.3 (2.7)
Roughness	0.03 (0.05)	0.08 (0.1)	0.06 (0.1)
Rugosity	0.009 (0.006)	0.01 (0.008)	0.012 (0.008)
Mobility	15.6 (2.4)	14.4 (1.9)	15.0 (1.5)
Frequency	2.3 (0.5)	2.1 (0.4)	2.1 (0.3)
Entropy	0.57 (0.04)	0.55 (0.04)	0.56 (0.02)
Peaks	172.0 (102.6)	135.1 (82.2)	148.5 (82.7)
Normalized peaks	4.7 (0.5)	4.2 (0.5)	4.2 (0.5)
Jerk			
Complexity	41.8 (10.1)	40.3 (21.1)	40.0 (10.3)
Roughness	60.5 (122.5)	121.4 (200.3)	92.9 (173.0)
Rugosity	0.21 (0.12)	0.32 (0.22)	0.30 (0.22)
Mobility	25.1 (4.3)	23.5 (4.6)	24.0 (2.7)
Frequency	3.5 (0.6)	3.3 (0.5)	3.4 (0.3)
Entropy	0.60 (0.03)	0.59 (0.04)	0.60 (0.03)

Results

In total, 9 people with PD were tested: 4 males and 5 females with a mean age of 67 (SD 9.4) years. The results showed no differences in any of the outcomes of interest measured with the tablet (duration, mean velocity, number of peaks, pause time, and number of pauses), or with APDM (acceleration and jerk). The only exception was APDM normalized acceleration peaks, where a main effect of intensity was found [sham: 4.67 (0.52), low: 4.18 (0.53), and high: 4.21 (0.50)]. However, a post hoc Tukey test revealed no pairwise-differences between frequencies.

In total, 3 out of 9 participants reported noticeable or marginal improvement, 4 out of 9 reported enjoyment in device use; 8 out of 9 reported device comfort, and no problems with the device were reported. For those individuals who reported perceived improvement, a stratified sample of tremor severity should be used in future studies to clarify which participants may garner efficacious results.

Figure 1 shows 2 representative participants who demonstrated a visual improvement on the spiral tasks in the high versus sham condition, with corresponding objective data of task duration, pause duration, pause count, mean velocity, and number of peaks. Participant 4 reported perceived improvement and expressed greater functional difficulty on their assessment or

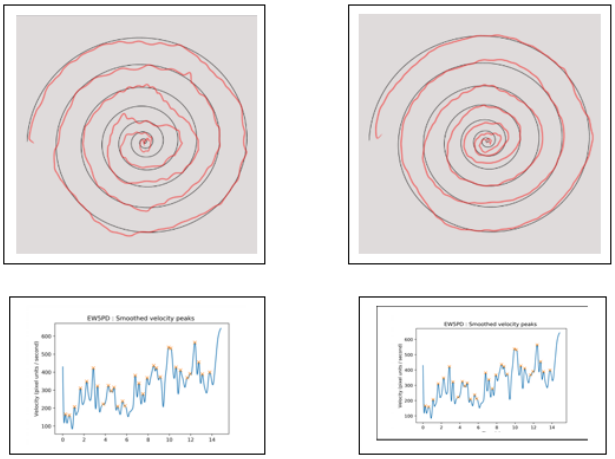
baselines scores (see Multimedia Appendix 1). Alternatively, participant 5 had a visual improvement in spiral quality; however, did not report perceived improvement. This visual difference was observed within a small subset of participants, some of whom reported perceived improvement. Reports included a "benefit on straight lines," "difference on spiral and star," that "continuous motion (was) easier," and the "most impact of tremor (was on) spiral."

Figure 1. This figure represents responders to high vibration for spiral drawings.

Responder 1: spiral images. See Table 1 for corresponding demographics

Participant 4

No vibration High-intensity vibration

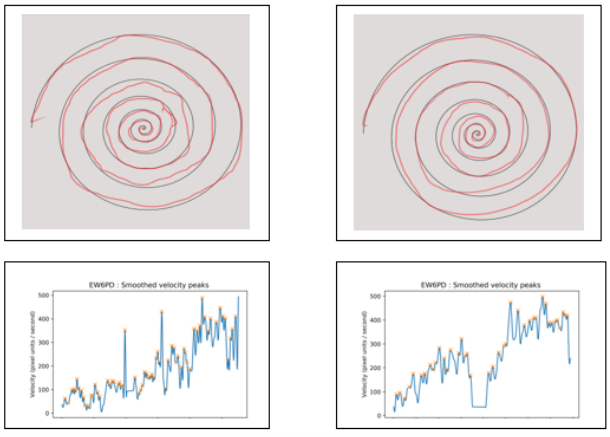


	No vibration	High vibration
Duration (s)	14.83	18.79
Peaks	31	36
Scaled peaks (peaks/s)	2.09	1.92
Mean velocity (pixels/s)	313.28	260.91
Pauses	0	1
Pause time (s)	0	.014

Responder 2: spiral, handwriting and "el" images

Participant 5

No vibration High-intensity vibration



	No vibration	High vibration
Duration (s)	27.57	19.71
Peaks	65	40
Scaled peaks (peaks/s)	2.36	2.03
Mean velocity (pixels/s)	184.13	255.86
Pauses	1	1
Pause time (s)	.01	.01

Discussion

Principal Findings

This pilot study investigated the use of the Emma Watch during fine motor and handwriting tasks for PD-induced action tremor. We recruited a range of people with PD with varying degrees of action tremor. The participants performed 5 different handwriting tasks on a digitized tablet with the use of the Emma Watch on the dominant wrist. The Emma Watch provided vibration at high, low, and zero intensity in a counterbalanced order. The device was found to be safe, and there were no adverse reactions. When interpreting our quantitative results, it is important to consider the small sample size of 9 people with PD. There was a main effect for differences in normalized acceleration peaks measured by APDM Opal sensors, however a post hoc Tukey test revealed no pairwise differences between vibration intensities. Visual differences were observed within a small subset of participants, some of whom reported perceived improvement.

Comparison to Prior Work

The basal ganglia play a critical role in automatic and volitional motor performance, making automatic motor tasks, such as walking and handwriting difficult for people with PD. In addition, the loss of dopamine from PD causes a decrease in

activation of the circuity that runs through the sensorimotor cortex and the basal ganglia; thus, increasing the response from the sensory system may alter the feedback loop and improve automatic motor responses [27]. In previous studies, the use of somatosensory cues has been successful in compensation strategies for improving gait impairments for people with PD [28]. Peripheral vibration is a form of sensory stimulation that provides proprioceptive input and may improve neuromotor function in people with sensorimotor deficits [18]. In a previous study in people with PD who had resting tremor, use of full body vibrotactile stimulation via 4 wearable devices on both wrists and ankles was found to be safe, feasible, and to possibly attenuate resting tremor [20]; however, people with action tremor were not included.

The Emma Watch has been hypothesized to mitigate tremor by mediating sensory signals in the cerebello-thalamo-cortical circuit, which has been linked to the origination of action tremor [29,30]. A recent study found that 80 Hz of vibration was sufficient to demonstrate improved motor performance as well as decreased beta oscillatory activity over the contralateral sensorimotor cortex compared to 20 Hz in people with PD, suggesting that higher frequency peripheral vibration increases the excitability of the sensorimotor cortex [15]. In another study, randomized frequency peripheral vibration from the TheraBracelet, when used in conjunction with a therapy, was found to be safe and feasible for upper extremity motor recovery post stroke [31]. In addition, in a previous study in people with PD, the Emma Watch, at 200 Hz at 60 bpm modulation, was found to have a significant improvement on movement speed and precision of motor performance during tracing motor control tasks when compared to 200 Hz at 20 bpm modulation [17]. Therefore, we hypothesized that the high intensity condition would be sufficient to improve motor function and performance in our study. As per recent evidence suggestions, sensory based strategies, including high-intensity vibration, may potentially improve motor learning and motor performance on automatic motor tasks [15,17,31]. This study is one of the first to report high-intensity vibration during fine motor and handwriting tasks for the primary aim of reducing action tremor in people with PD.

Strengths and Limitations

While we found no difference in task performances with the device on versus off in our 9 participants, a visual difference in accuracy was observed within a small subset, who further reported perceived improvement in handwriting or drawing skills. It is unclear why some participants with action tremor experienced improvements while others did not. Perhaps it may be that individuals who have a lower perceived hand function ability, or greater tremor severity, perform better with use of the device.

There are multiple limitations in this study that should be considered when interpreting the results. First, the results of this study are limited due to the small sample size as we did not reach our planned recruitment goal due to the COVID-19 pandemic; therefore, future studies would benefit from a larger sample size to increase the statistical power. Second, other limitations included technical difficulties spanning app failure, device malfunction, and data-saving issues. Unfortunately, these technical difficulties further reduced the amount of data available.

Third, the Emma Watch was used for a relatively short time period, and hypothetically, it is possible that length of time under stimulus may affect its efficacy. Alternatively, a learning effect might develop while using the Emma Watch as the body adapts to both the vibration stimulus and to the repetition of drawing the same 5 tasks multiple times. Lastly, handwriting was performed on a tablet with a stylus, which has notable differences to handwriting performance using pen and paper, such as paper orientation and feedback from the pen. However, kinematic analysis could not be performed without use of a digitized tablet [24].

Future Directions

Future studies should consider focusing on individuals with greater tremor severity, as 1 participant in our study who demonstrated visual improvement had an action tremor of 2 and was at Hoehn and Yahr Stage 3. Additionally, the vibration may need to be individually tailored to the participant to maximize benefits. A stratified sample of tremor severity should be used to clarify which participants may benefit from this or similar devices. Future studies may also consider including individuals with action tremor who have a diagnosis of Essential Tremor, as there may be a difference in the response generated from the Emma Watch. According to Chen et al [32], there was a difference in the velocity of spiral drawing between patients with essential tremor and those with PD who had similar severity in their action tremor. We initially did include people with essential tremor in our study, and wanted to compare the 2 populations; however, due to technical difficulties and the COVID-19 pandemic halting human research, we did not have sufficient data to analyze and compare to people with PD.

While we recognize that the small sample size and technical difficulties limit the interpretation of our results, there were visually depicted and subjectively reported improvements for a small subset of participants that are important to recognize. As action tremor severely affects quality of life and functional independence for people with PD, it is increasingly important to report on any interventions that may potentially improve functional abilities [8]. Future studies must focus on finding safe and efficacious ways to address this clinical need and should explore the efficacy of combining the Emma Watch with task-specific training or other intervention tools, as this may attenuate action tremor. However, the Emma Watch efficacy in improving fine motor or handwriting skills as a stand-alone tool remains to be demonstrated.

Acknowledgments

All authors declared that they had insufficient or no funding to support open access publication of this paper, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided paper processing fee support for the publication of this paper.

Thank you to all the participants for volunteering their time, and to Greg Youdan for helping with data collection.

Data Availability

The APDM data and tablet data sets generated during and/or analyzed in this study are available from the corresponding author on reasonable request.

Authors' Contributions

TAvS, NP, HZ, and LQ were in charge of conceptualization. Data curation and manuscript writing were carried out by TAvS, NP, LQ, AP, MB, CM, ARS, and RD. AP, CM, ARS, and LQ performed the investigation. TAvS, NP, LQ, AP, MB, CM, ARS, RD, and JV carried out the study’s analysis. The software was managed by GJ and HZ.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Further information on participant demographics, means of tasks outcomes between intensities, responders for “elelelel” drawings, the equations used in the APDM analysis, and participant quotes.
[DOCX File , 1166 KB - biomedeng_v8i1e40433_app1.docx]

References

1. Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008;79(4):368-376 [FREE Full text] [doi: 10.1136/jnnp.2007.131045] [Medline: 18344392]
2. Simon DK, Tanner CM, Brundin P. Parkinson disease epidemiology, pathology, genetics, and pathophysiology. Clin Geriatr Med 2020;36(1):1-12 [FREE Full text] [doi: 10.1016/j.cger.2019.08.002] [Medline: 31733690]
3. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. JAMA 2020;323(6):548-560. [doi: 10.1001/jama.2019.22360] [Medline: 32044947]
4. Abusrair AH, Elsekaily W, Bohlega S. Tremor in Parkinson's disease: from pathophysiology to advanced therapies. Tremor Other Hyperkinet Mov (N Y) 2022;12:29 [FREE Full text] [doi: 10.5334/tohm.712] [Medline: 36211804]
5. Helmich RC, Toni I, Deuschl G, Bloem BR. The pathophysiology of essential tremor and Parkinson's tremor. Curr Neurol Neurosci Rep 2013;13(9):378. [doi: 10.1007/s11910-013-0378-8] [Medline: 23893097]
6. Lenka A, Jankovic J. Tremor syndromes: an updated review. Front Neurol 2021;12:684835 [FREE Full text] [doi: 10.3389/fneur.2021.684835] [Medline: 34381412]
7. Louis ED, Levy G, Côte LJ, Mejia H, Fahn S, Marder K. Clinical correlates of action tremor in Parkinson disease. Arch Neurol 2001;58(10):1630-1634 [FREE Full text] [doi: 10.1001/archneur.58.10.1630] [Medline: 11594921]
8. Heusinkveld LE, Hacker ML, Turchan M, Davis TL, Charles D. Impact of tremor on patients with early stage Parkinson's disease. Front Neurol 2018;9:628 [FREE Full text] [doi: 10.3389/fneur.2018.00628] [Medline: 30123178]
9. Helmich RC, Dirkx MF. Pathophysiology and management of Parkinsonian tremor. Semin Neurol 2017;37(2):127-134. [doi: 10.1055/s-0037-1601558] [Medline: 28511253]
10. Muthuraman M, Heute U, Arning K, Anwar AR, Elble R, Deuschl G, et al. Oscillating central motor networks in pathological tremors and voluntary movements. What makes the difference? Neuroimage 2012;60(2):1331-1339. [doi: 10.1016/j.neuroimage.2012.01.088] [Medline: 22293134]
11. Muthuraman M, Raethjen J, Koirala N, Anwar AR, Mideksa KG, Elble R, et al. Cerebello-cortical network fingerprints differ between essential, Parkinson's and mimicked tremors. Brain 2018;141(6):1770-1781 [FREE Full text] [doi: 10.1093/brain/awy098] [Medline: 29701820]
12. Wong JK, Viswanathan VT, Nozile-Firth KS, Eisinger RS, Leone EL, Desai AM, et al. STN versus GPi deep brain stimulation for action and rest tremor in Parkinson's disease. Front Hum Neurosci 2020;14:578615 [FREE Full text] [doi: 10.3389/fnhum.2020.578615] [Medline: 33192410]
13. Gigante AF, Bruno G, Iliceto G, Guido M, Liuzzi D, Mancino PV, et al. Action tremor in Parkinson's disease: frequency and relationship to motor and non-motor signs. Eur J Neurol 2015;22(2):223-228. [doi: 10.1111/ene.12583] [Medline: 25363380]
14. Dincher A, Schwarz M, Wydra G. Analysis of the effects of whole-body vibration in Parkinson disease—systematic review and meta-analysis. PM R 2019;11(6):640-653. [doi: 10.1002/pmrj.12094] [Medline: 30689308]

15. Macerollo A, Palmer C, Foltynie T, Korlipara P, Limousin P, Edwards M, et al. High-frequency peripheral vibration decreases completion time on a number of motor tasks. Eur J Neurosci 2018;48(2):1789-1802 [FREE Full text] [doi: 10.1111/ejn.14050] [Medline: 29923362]
16. King LK, Almeida QJ, Ahonen H. Short-term effects of vibration therapy on motor impairments in Parkinson's disease. NeuroRehabilitation 2009;25(4):297-306. [doi: 10.3233/NRE-2009-0528] [Medline: 20037223]
17. Macerollo A, Holz C, Cletheror D, Vega J, Moody J, Saul G, et al. Non-invasive intervention for motor signs of Parkinson's disease: the effect of vibratory stimuli. J Neurol Neurosurg Psychiatry 2020;92(1):109-110 [FREE Full text] [doi: 10.1136/jnnp-2020-323427] [Medline: 32917824]
18. Lau RWK, Teo T, Yu F, Chung RCK, Pang MYC. Effects of whole-body vibration on sensorimotor performance in people with Parkinson disease: a systematic review. Phys Ther 2011;91(2):198-209 [FREE Full text] [doi: 10.2522/ptj.20100071] [Medline: 21212374]
19. Buki E, Katz R, Zacksenhouse M, Schlesinger I. Vib-bracelet: a passive absorber for attenuating forearm tremor. Med Biol Eng Comput 2018;56(5):923-930. [doi: 10.1007/s11517-017-1742-7] [Medline: 29101536]
20. Tabacof L, Braren S, Patterson T, Fry A, Putrino D. Safety and tolerability of a wearable, vibrotactile stimulation device for Parkinson's disease. Front Hum Neurosci 2021;15:712621 [FREE Full text] [doi: 10.3389/fnhum.2021.712621] [Medline: 34867237]
21. Jöbges EM, Elek J, Rollnik JD, Dengler R, Wolf W. Vibratory proprioceptive stimulation affects Parkinsonian tremor. Parkinsonism Relat Disord 2002;8(3):171-176. [doi: 10.1016/s1353-8020(01)00016-5] [Medline: 12039427]
22. Bain PG, Findley LJ, Atchison P, Behari M, Vidailhet M, Gresty M, et al. Assessing tremor severity. J Neurol Neurosurg Psychiatry 1993;56(8):868-873 [FREE Full text] [doi: 10.1136/jnnp.56.8.868] [Medline: 8350102]
23. Chen CC, Bode RK. Psychometric validation of the Manual Ability Measure-36 (MAM-36) in patients with neurologic and musculoskeletal disorders. Arch Phys Med Rehabil 2010;91(3):414-420. [doi: 10.1016/j.apmr.2009.11.012] [Medline: 20298833]
24. Thomas M, Lenka A, Pal PK. Handwriting analysis in Parkinson's disease: current status and future directions. Mov Disord Clin Pract 2017;4(6):806-818 [FREE Full text] [doi: 10.1002/mdc3.12552] [Medline: 30363367]
25. Snyder P, Tummalacherla M, Perumal T, Omberg L. mhealthtools: a modular R package for extracting features from mobile and wearable sensor data. J Open Source Softw 2020;5(47):2106 [FREE Full text] [doi: 10.21105/joss.02106]
26. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
27. Petzinger GM, Fisher BE, McEwen S, Beeler JA, Walsh JP, Jakowec MW. Exercise-enhanced neuroplasticity targeting motor and cognitive circuitry in Parkinson's disease. Lancet Neurol 2013;12(7):716-726 [FREE Full text] [doi: 10.1016/S1474-4422(13)70123-6] [Medline: 23769598]
28. Nonnekes J, Ružicka E, Nieuwboer A, Hallett M, Fasano A, Bloem BR. Compensation strategies for gait impairments in Parkinson disease: a review. JAMA Neurol 2019;76(6):718-725. [doi: 10.1001/jamaneurol.2019.0033] [Medline: 30907948]
29. Helmich RC, Hallett M, Deuschl G, Toni I, Bloem BR. Cerebral causes and consequences of Parkinsonian resting tremor: a tale of two circuits? Brain 2012;135(11):3206-3226 [FREE Full text] [doi: 10.1093/brain/aws023] [Medline: 22382359]
30. Lora-Millán JS, López-Blanco R, Gallego J, Méndez-Guerrero A, de la Aleja JG, Rocon E. Mechanical vibration does not systematically reduce the tremor in essential tremor patients. Sci Rep 2019;9(1):16476 [FREE Full text] [doi: 10.1038/s41598-019-52988-8] [Medline: 31712728]
31. Seo NJ, Woodbury ML, Bonilha L, Ramakrishnan V, Kautz SA, Downey RJ, et al. TheraBracelet stimulation during task-practice therapy to improve upper extremity function after stroke: a pilot randomized controlled study. Phys Ther 2019;99(3):319-328 [FREE Full text] [doi: 10.1093/ptj/pzy143] [Medline: 30690609]
32. Chen KH, Lin PC, Yang BS, Chen YJ. The difference in visuomotor feedback velocity control during spiral drawing between Parkinson's disease and essential tremor. Neurol Sci 2018;39(6):1057-1063. [doi: 10.1007/s10072-018-3331-4] [Medline: 29572654]

Abbreviations

PD: Parkinson disease

Edited by T Leung; submitted 21.06.22; peer-reviewed by M Mostafa, PV Eswaradass, C Shah, S Thirunavukkarasu; comments to author 07.12.22; revised version received 16.03.23; accepted 10.07.23; published 23.10.23.

Please cite as:
Pacheco A, van Schaik TA, Paleyes N, Blacutt M, Vega J, Schreier AR, Zhang H, Macpherson C, Desai R, Jancke G, Quinn L
A Wearable Vibratory Device (The Emma Watch) to Address Action Tremor in Parkinson Disease: Pilot Feasibility Study
JMIR Biomed Eng 2023;8:e40433
URL: https://biomedeng.jmir.org/2023/1/e40433
doi:10.2196/40433
PMID:38875672

Original Paper

A Radar-Based Opioid Overdose Detection Device for Public Restrooms: Design, Development, and Evaluation Study

Jessica Oreskovic[1], MASc; Jaycee Kaufman[1], MSc; Anirudh Thommandram[1], MASc; Yan Fossat[1], MSc

Klick Labs, Klick Inc, Toronto, ON, Canada

Corresponding Author:
Jessica Oreskovic, MASc
Klick Labs
Klick Inc
175 Bloor Street East
North Tower #300, 3rd Floor
Toronto, ON, M4W3R8
Canada
Phone: 1 416 214 4977
Email: joreskovic@klick.com

Abstract

Background: The opioid epidemic is a growing crisis worldwide. While many interventions have been put in place to try to protect people from opioid overdoses, they typically rely on the person to take initiative in protecting themselves, requiring forethought, preparation, and action. Respiratory depression or arrest is the mechanism by which opioid overdoses become fatal, but it can be reversed with the timely administration of naloxone.

Objective: In this study, we described the development and validation of an opioid overdose detection radar (ODR), specifically designed for use in public restroom stalls. In-laboratory testing was conducted to validate the noncontact, privacy-preserving device against a respiration belt and to determine the accuracy and reliability of the device.

Methods: We used an ODR system with a high-frequency pulsed coherent radar sensor and a Raspberry Pi (Raspberry Pi Ltd), combining advanced technology with a compact and cost-effective setup to monitor respiration and detect opioid overdoses. To determine the optimal position for the ODR within the confined space of a restroom stall, iterative testing was conducted, considering the radar's bounded capture area and the limitations imposed by the stall's dimensions and layout. By adjusting the orientation of the ODR, we were able to identify the most effective placement where the device reliably tracked respiration in a number of expected positions. Experiments used a mock restroom stall setup that adhered to building code regulations, creating a controlled environment while maintaining the authenticity of a public restroom stall. By simulating different body positions commonly associated with opioid overdoses, the ODR's ability to accurately track respiration in various scenarios was assessed. To determine the accuracy of the ODR, testing was performed using a respiration belt as a reference. The radar measurements were compared with those obtained from the belt in experiments where participants were seated upright and slumped over.

Results: The results demonstrated favorable agreement between the radar and belt measurements, with an overall mean error in respiration cycle duration of 0.0072 (SD 0.54) seconds for all recorded respiration cycles (N=204). During the simulated overdose experiments where participants were slumped over, the ODR successfully tracked respiration with a mean period difference of 0.0091 (SD 0.62) seconds compared with the reference data.

Conclusions: The findings suggest that the ODR has the potential to detect significant deviations in respiration patterns that may indicate an opioid overdose event. The success of the ODR in these experiments indicates the device should be further developed and implemented to enhance safety and emergency response measures in public restrooms. However, additional validation is required for unhealthy opioid-influenced respiratory patterns to guarantee the ODR's effectiveness in real-world overdose situations.

(JMIR Biomed Eng 2023;8:e51754) doi:10.2196/51754

KEYWORDS

60 GHz radar; opioid overdose; overdose detection; overdose prevention; respiratory depression

Introduction

The opioid crisis is a growing problem worldwide, with devastating consequences. In the United States alone, fatal overdoses involving opioids claimed the lives of over 80,000 people in 2021 [1]. This problem has been escalating since the 1990s and has seen a fourfold increase from 2010 to 2021, with the rate showing no sign of abating [1]. The more recent surge in opioid-related mortality is largely a result of the rising prevalence of synthetic opioids such as fentanyl, which are substantially more potent than natural opiates [2-4]. Fentanyl and fentanyl analogs are found to be 50-10,000 times more potent than morphine [2,4,5]. These synthetic opioids are added to street drugs without the user's knowledge to enhance the effects of the drug [6]. As a result, individuals who consume drugs laced with synthetic opioids are at a significantly greater risk of overdosing, as their usual dose may have an unexpected and dangerously high potency [5,6].

The principal mechanism by which opioid overdoses become fatal is respiratory depression through suppression of the region of the brain responsible for respiration rhythm regulation [7], referred to as opioid-induced respiratory depression (OIRD). In the early stages of an overdose, individuals lose consciousness and experience slowed respiration, which can lead to dangerously low blood oxygen levels [7]. Consequently, the individual will experience hypoxia, hypercapnia, and, if left untreated, complete respiration cessation and suffocation [7,8].

Fortunately, there is a fast-acting opioid antagonist called naloxone that reverses the effects of opioid toxicity. Naloxone has a stronger affinity to bond with the same receptor sites in the brain than opioids, thereby blocking the effects of opioids and restoring proper respiration functionality [9-11]. While the effect of naloxone is temporary, lasting around 30-90 minutes, it provides sufficient time for emergency medical personnel to be contacted, arrive, and assume treatment [9-11]. To administer naloxone and save a survivor's life, the early detection of an overdose is imperative before OIRD has caused blood oxygen levels to drop significantly. When the brain is deprived of oxygen for 3 minutes, consciousness is lost, and lasting brain damage is to be expected [12]. After 10 minutes of oxygen deprivation, if the individual is still alive, they will likely be in a coma and experience permanent brain damage [12]. Therefore, it is crucial that the overdose survivor be discovered as soon as possible to administer naloxone and restore respiration and oxygen flow to the brain.

Many initiatives have been introduced with the goal of improving safety surrounding drug use, such as digital overdose monitoring services, syringe services programs, and supervised injection facilities or safe consumption sites. Digital overdose monitoring services include phone lines where a person can remain on the phone with someone informed of their location and other pertinent information while they use drugs [13]. If the person using drugs becomes unresponsive, the volunteer on the line can contact emergency services and provide the overdose survivor's location. Syringe services programs supply people who inject drugs with sterile injection supplies to prevent needle reuse and inhibit the transmission of diseases such as HIV and hepatitis [14]. Supervised injection facilities supply trained medical supervision while people use drugs, providing a safe and sterile environment equipped with overdose prevention measures such as naloxone [14,15]. While these safe injection facilities improve safety for people who use drugs and choose to take advantage of them, many people prefer to use drugs in more private settings, such as their homes, or in public facilities, such as restrooms [16]. An exploratory study was completed in New York where business managers were interviewed, and 58% of those questioned had encountered drug use in the restroom of their place of work, indicating the significance of the public restroom drug use problem [17].

The prevention of opioid overdoses in public restrooms has become an increasingly important focus of research and development in response to the opioid crisis [17]. While some businesses have used methods to deter drug use, such as the use of blue lights in restrooms to make veins less visible, restricting restroom access, and even removing stall doors, others have implemented strategies to enhance the safety of individuals who use drugs [17]. For example, supplying naloxone in restrooms and the implementation of antimotion detectors have been proposed as potential solutions [17]. However, while antimotion detectors such as the Brave Sensor or the South End Clinic Anti-Motion Sensor have shown promise in detecting potential overdoses, their lack of specificity means they are prone to triggering many false positives [18,19]. To address this issue, a more precise detection method that monitors respiration, specifically detecting OIRD, could be implemented to enhance opioid overdose identification accuracy and alert staff. This will ultimately produce a more rapid and effective response, as required to reverse overdoses [20].

Respiration monitoring is a customary practice in medical settings and is done using capnography, a medical monitoring technology that uses infrared light to measure the carbon dioxide concentration in expired breath [21]. The capnograph outputs a waveform of carbon dioxide concentration over time, which is representative of inspiration and expiration periods used to calculate respiration rate [21]. If the patient does not require assistance breathing, capnography can be measured with nasal cannulas, or if they are on a ventilator, an adaptor is added to the mouthpiece [21]. While this method has high accuracy in respiration rate monitoring because it tracks respiration waveforms, it requires specialized and expensive equipment [21]. Furthermore, it needs contact with the participant and therefore is not suitable for monitoring in a public restroom.

A simplified means of measuring respiration rate during physical activity is through the use of a respiration belt. The nonelastic belt is affixed around the participant's thorax or abdomen and uses a strain sensor to detect the corresponding expansion and contraction of the chest wall during respiration. The sensitive strain gauges in the respiration belts allow them to work well to monitor all levels of respiration if they are tightly worn, including measuring OIRD [22]. Although respiration belts do not offer insight into breath composition as provided by capnography, they provide valuable information regarding respiration rate, depth, and patterns. The noninvasive nature of respiration belts makes them a convenient and appropriate

solution for a diverse range of applications; however, the belt still requires contact with the participant to monitor respiration.

To identify OIRD in a public restroom setting, a device only needs to measure respiration rate, not breath composition. However, it must be practical, have sufficient precision and sensitivity, and most importantly, be noncontact. Some systems cover most of these requirements but cause concerns for privacy, such as computer vision technologies that can monitor respiration from pixel intensity between frames [23]. Others preserve privacy but are not noncontact, such as pulse oximetry and capnography, which track respiration with a finger monitor or nasal cannulas [23]. A novel noncontact and privacy-preserving smartphone app has been developed for monitoring respiration rates during opioid use and generating alerts in the event of dangerously low respiration rates [24]. This innovative tool leverages short-wave active sonar technology combined with frequency shifts to enhance the sensitivity of the measurement system [24]. This app allows for remote supervision of opioid use in any setting, with the user positioning their phone to face their chest and the app tracking respiration rate by measuring the distance to the user's chest. The smartphone app requires users to be in a setting where they can position their phone properly and requires a sufficiently charged smartphone. According to a study regarding people who use substances' acceptability of technological solutions, less than 50% of the people interviewed owned a cellphone [25]. Of those who had a smartphone and access to the internet, less than 70% would consider using it for monitoring their drug use [25]. Most critically, this approach requires active participation from the person using drugs. While this app has the potential to enhance safety during opioid use in any location, like safe injection facilities, it requires forethought and does not address the needs of individuals who are not actively engaged in taking responsibility for their own safety.

We developed a novel approach for detecting and preventing public opioid overdose fatalities. A stand-alone radar device is to be installed in a public restroom stall that can monitor the respiration rate of individuals and subsequently trigger an alert for bystanders to administer naloxone when overdose respiration patterns are detected. The design and development of this device involved several challenges, particularly with respect to ensuring the privacy and anonymity of the individuals being monitored. To overcome these challenges, pulsed coherent radar technology is proposed as a viable, privacy-preserving solution. This compact, low-cost, and high-precision modality allows for the monitoring of respiration rates by detecting changes in the direction of chest movement. The pulsed coherent radar sensor works by emitting a series of high-frequency pulses sweeping through a bandwidth around a set center frequency [26]. The pulses are reflected by body tissues back to the receiver, creating a signal to be analyzed by the system [26]. This technique is derived from the Doppler shift principle, which measures the change in frequency of the received signal caused by chest motion. This technology also controls the phase of the transmitted and received signals, allowing for high-accuracy measurements and extreme sensitivity to very small movements [26]. Because the sensor only records distances, it protects the participant's privacy. Another benefit of the radar sensor is that the radio waves only reflect off materials with high reflectivity, such as metal, water, and human tissue. This feature allows the radar to sense chest movement and monitor breathing through clothing, even heavy winter coats, making it a viable rescue-alerting solution for many scenarios where a person experiences opioid toxicity in a public restroom.

This paper details the development and preliminary exploratory testing of a radar-based opioid overdose detection device for public restrooms. We provide information about the design process, the respiration tracking ability of the sensor and algorithm, the device's accuracy in a confined restroom stall and its optimal position, and the device's ability to monitor respiration during a simulated opioid overdose. The aim of this study is to assess the feasibility of an overdose detection radar (ODR) for use in public restrooms.

Methods

Overview

The ODR development was informed by a thorough review of relevant literature and consultation with medical practitioners. Through interviews with emergency medical responders and emergency department physicians, we gained valuable insight regarding the appropriate respiration rate thresholds at which to trigger an opioid overdose alert. We also learned details about the expected position, respiratory condition, and sequence of events that happen to a participant experiencing an opioid overdose. The accuracy of the respiration rate measured by the ODR was assessed in this study.

Design

Minimizing the size of the device was important to consider throughout the component selection and design process, as the intended use of the device is for small public restroom stalls.

Technology Selection

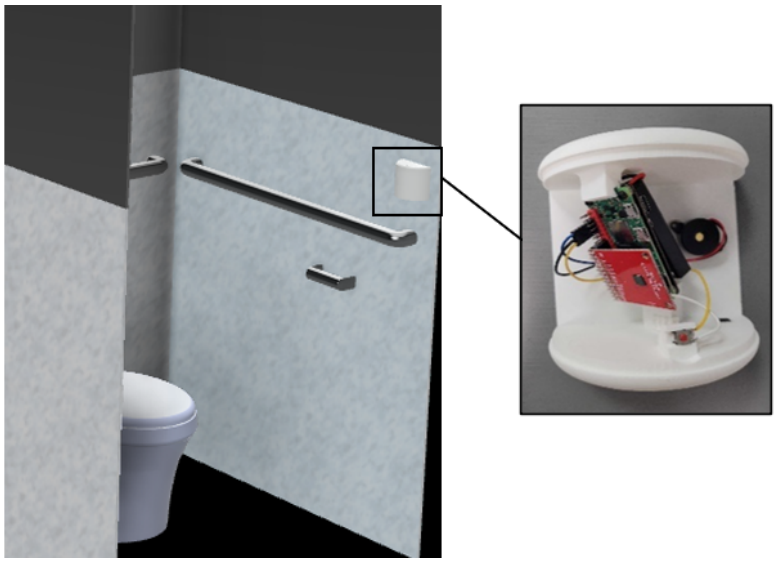
The selection of the Raspberry Pi Zero 2 W (Raspberry Pi Ltd) and Acconeer's A111 pulsed coherent radar breakout sensor (SparkFun Electronics) for monitoring respiration was based on their performance specifications and compact size. During the exploration phase, the specifications and software development kits for several radar sensors with operating frequencies of both 24 GHz and 60 GHz were compared. The Acconeer A111 60GHz pulsed coherent radar sensor was selected for this application because of its robust software development kit, configurability for the application, and high measurement accuracy, with the 60GHz allowing for the sensitivity and resolution to measure very small movements.

Form Factor and Other Considerations

The enclosure for the ODR was carefully designed to ensure it was discreet and unobtrusive, resembling common restroom fixtures such as air fresheners and smoke alarms (Figure 1). The device was intentionally crafted to not display any visible lights or lenses to avoid concern about identifying information being recorded. While the device does collect and store some data as required for the radar technology, it only records distance data to monitor respiration. The data involved are similar to that used by an automatic flush sensor or motion sensor for lights;

therefore, we expect no issues with installing such technology in public restrooms. A key consideration in the enclosure design was its suitability for public settings, requiring it to be tamper-resistant, durable, easy to clean, and securely mountable to restroom stall walls. For this, we designed a compact (10 cm × 11 cm × 6 cm) hemicylindrical enclosure. The enclosure is currently 3D-printed with polylactic acid, a durable and cost-effective material. Further iterations may progress to an injection-molded design.

Figure 1. Image of the enclosure and internal components, consisting of the Raspberry Pi Zero 2W, Acconeer's A111 sensor, and an uninterruptible power supply battery package. For prototyping, the enclosure includes different mounting angles for the electronics. Iterative testing found that the 45-degree angle gave more consistent and accurate measurements.



Position

According to the sensor specifications, the optimal position for the ODR is directly facing the participant's chest within the programmed range. However, it was found through consultation with professionals that the device must be designed with consideration that the individual will likely lose consciousness before their respiration rate drops to dangerous levels. Thus, the device must accurately measure respiration when the person is slumped in all different directions within the stall. Due to the bounded capture area of the radar, 65 degrees horizontal and 53 degrees vertical, it was difficult to monitor the entire volume of the stall. Therefore, the scope was limited to scenarios where the participant was slumped over while remaining on the toilet seat. In-laboratory experiments were conducted using a mock restroom stall setup following building standards for public restroom stalls: 890-940 mm wide, 1500 mm long, and with an 860 mm door opening [27].

The participant was seated in the restroom stall setup (Figure 2) with the ODR attached to the wall at different distances, heights, and angles from the participant's chest. The device was secured to the side wall using Velcro strips, allowing for easy repositioning. To determine an optimal position, the targeted respiration rate was held constant at 15 breaths per minute (bpm) by following audible cues to inhale and exhale. Once the device accurately detected respiration while the participant was seated upright, the participant slumped in various directions to ensure respiration could be tracked accurately in different body positions. The measured respiration rate was monitored to ensure it continued to accurately measure approximately 15 bpm while the participant slumped in various directions to approximate an overdose. If the device failed to track respiration while the participant was slumped in any direction, the device position was adjusted, and the experiment was repeated until respiration was consistently tracked in all slumped positions.

Figure 2. Mock restroom stall setup according to building code regulations (890-940 mm wide and 1500 mm long), with the toilet centered between the side walls.

Respiration

The respiration tracking algorithm implemented for opioid overdose monitoring is a customized version of Acconeer's A111 SDK breathing module. Acconeer's Python algorithm uses distance measurements to monitor the respiratory cycle by tracking the relative distance of the reflecting body [28]. An exhale is registered when the distance of the reflecting body increases, indicating that the chest surface is moving away from the sensor, while an inhale is registered when the distance decreases, indicating that the chest is moving toward the sensor. To avoid incorporating body movements into the respiration tracking, the algorithm imposes constraints on the distance of the motion used to calculate respiration, limiting it to a range between 0.75 mm and 20 mm. Each breath cycle is defined by 1 inhale and 1 exhale, which are identified based on the change in distance. The duration of a breath cycle is subsequently used to compute the current respiration rate. The algorithm accounts for variations in respiratory patterns that may lead to brief inaccuracies in the calculated respiration rate. To customize the device for the opioid overdose detection application, we added respiration rate thresholds and respiration cessation triggers for detecting breathing patterns consistent with opioid overdoses.

Respiration validation was completed for 2 positions. The first validation testing protocol for the device included a healthy participant outfitted with the Go Direct respiration belt (Vernier) and seated upright in a chair directly in front of the radar device at a distance of 1.5 m. The participant was instructed to breathe at a comfortable pace for approximately 5 seconds until both measurement devices began recording, and then to hold their breath for 3 seconds. This breath-hold event was included to allow the data sources to be synced in analysis and remove any delays caused by setup or calibration time after the algorithm was started.

The second respiration validation protocol was modified to include the radar in its final position. This was completed to ensure that respiration measurements were accurate for multiple participants. Participants were outfitted with the Go Direct respiration belt for validation purposes and followed the same protocol as previously described. Each participant recorded 2 tests with the device fixed in the mock stall in its mounting position and the participant seated upright on the mock toilet. Breathing motion captured from the radar was compared with the peaks recorded from the respiration belt, and the error between respiration periods was computed.

OIRD is a life-threatening condition characterized by a significant decrease in breathing rate caused by opioids. To establish a reliable threshold for triggering an overdose alert, consultation was sought from medical professionals. According to medical experts and literature, a sustained respiration rate of less than 8 bpm is considered dangerous and requires immediate intervention [29]. Accordingly, the overdose detection algorithm has been designed to trigger an alert when the respiration rate falls below this threshold for 15 seconds or 2 full breaths. The inclusion of 2 breaths enables the algorithm to discern a momentary low respiration rate reading caused by intentional motion from an actual overdose event, including a sustained lowered rate. Additionally, the algorithm is programmed to trigger an alert if no respiration motion is detected for 10 seconds, providing an added layer of protection if respiration drops too quickly to track.

The duration of the respiration period is crucial in computing the respiration rate. Therefore, we elected to assess the error by analyzing the mean difference between the respiration periods calculated from the respiration belt and the overdose radar. The direction of the error was included because it is important to know whether the respiration rate over several breath cycles will remain accurate, regardless of an error in a single respiration cycle duration.

Overdose Simulation

Validation testing was performed on the ODR's ability to monitor respiration rate in the mock stall while the participants simulated an overdose by slumping in all directions. The

protocol for this validation test includes outfitting the participant with the respiration belt, sitting upright in the mock stall, and breathing at a comfortable pace for 30 seconds. The participant is then signaled to simulate falling unconscious by slumping forward, then backward, left, and right, holding each position for 30 seconds and breathing at a comfortable pace. The ODR's respiration tracking algorithm and the respiration belt recorded data throughout all positions and movements, and the data was analyzed to determine the agreement between the algorithm and belt respiration periods in each position.

Ethical Considerations

Because of the challenges of running a study in a public restroom, including privacy and health concerns, to confirm the performance of the ODR, preliminary testing was conducted in-laboratory on 3 of the authors (JO, JK, and YF), who are all healthy and consented to perform validation tests and approximate an overdose scenario in a mock restroom stall. All the data recorded in the preliminary testing contained no identifying information and only recorded respiration patterns. To ensure the safety of the human participants, the ODR was designed using approved, commercially available devices, and any testing was done to help inform the development of the device. Therefore, as per Article 6.11 of the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans, we did not require approval from the Research Ethics Board for this exploratory testing [30]. Additionally, the testing procedures did not require the participants to perform any potentially dangerous respiration patterns to replicate an opioid overdose. Participants were able to breathe at a self-selected, comfortable pace throughout all experiments, and while the radar is unaffected by what type of clothing is worn, participants wore light to medium-weight long-sleeved clothing for the described tests.

Results

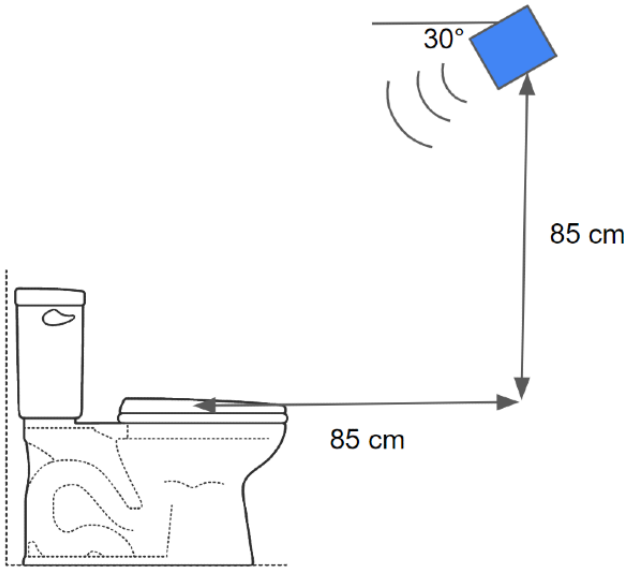
Overview

The experimental procedures outlined in this study involved the use of 2 distinct data sources: the respiration belt, which measures force in Newtons, and the radar, which measures distance in millimeters. Although the sources provide differing data signals, both can provide valuable information on breath timing and can be used to calculate the duration of each respiration cycle or period. We retained the sign of the difference instead of studying the absolute difference since a slight deviation in the peak could yield 1 positive and 1 negative difference while still resulting in the same respiration rate. The respiration period is important in the context of the overdose detection algorithm, where the respiration rate is a critical feature.

Position

The initial placement of the radar device was at chest height, tilted at an angle of 45 degrees away from the wall, directed toward the chest. The sensor-chest distance was situated within the range of 0.75-1.5 m. However, when the participant slumped forward, the radar was aimed directly at the top of the head, which did not display any motion related to respiration. To address this limitation, we iteratively adjusted the radar height and tilt angle to target the chest directly. Eventually, the optimal position was determined to be 85 cm horizontally from the front of the chest (center of seat) while seated upright, 85 cm above the top of the toilet seat, angled 30 degrees down from horizontal, and at a 45-degree angle setting within the device pointing out from the wall toward the toilet (Figure 3). The 30-degree downward angle facilitated a direct path toward 35 cm above the toilet seat, specifically targeting the chest and abdomen at a distance of approximately 1 m. This elevated position allows the chest to be monitored when the participant is upright or slumped backward and the back to be monitored if the participant collapses forward.

Figure 3. Radar positioning. The optimal position for the radar device in the mock restroom stall was found to be 85 cm in front of the center of the seat, 85 cm above the top of the seat, and angled 30 degrees down toward the toilet. Additionally, inside the casing, the sensor was angled at 45 degrees out from the wall to target the participant's chest while seated on the toilet. Through testing, this position was capable of consistently monitoring respiration with the participant in many different slumped positions.



Respiration

In the initial benchtop validation experiment, a participant was seated upright in a comfortable position while their respiration was recorded simultaneously by the respiration belt and radar. A total of 67 respiration periods were analyzed over 6 experiments (2 per participant). These tests revealed a favorable agreement in peak distance between the radar and the respiration belt, with a mean error in cycle duration of 0.0045 (SD 0.42) seconds (Figure 4; representative test displayed in Figure 5).

A secondary set of validation tests was conducted to evaluate the tracking accuracy of the radar sensor's position in detecting respiration. Participants were seated upright in the mock stall, with the radar device placed in its optimal position. Participants were instructed to breathe at a comfortable pace, and their breath periods were measured and compared using both the respiration belt and radar methods. A total of 85 respiration cycles were analyzed for the position validation testing, and good agreement was observed across all tests (Figure 4). The mean breath period error was found to be 0.012 (SD 0.42) seconds, and a representative test is presented in Figure 6.

Figure 4. Errors in measurement between the radar and the respiration belt for the duration of each respiration cycle monitored were split into each of the test protocols. The mean error for all tests is minimal, with similar SDs between tests. The SD is larger as it demonstrates that the error in cycle duration can be positive or negative, but overall, it balances out to a very small mean error, having a small effect on the respiration rate over time.

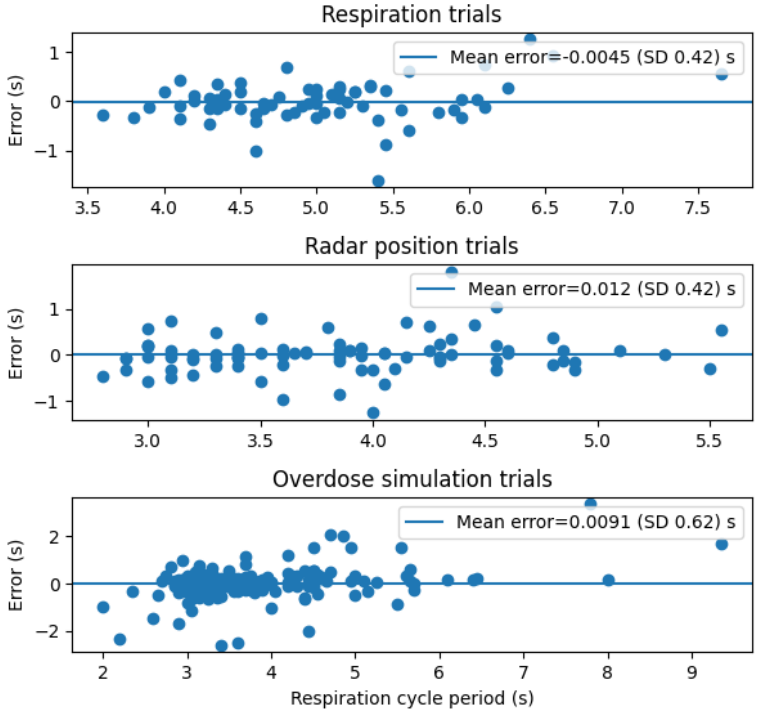


Figure 5. Representative respiration experiment: breathing motion captured by overdose radar (blue) and respiration belt (red) in the first plot. The participant was breathing comfortably while seated upright, directly 1.5 m in front of the radar sensor. The second plot shows the respiration cycle period in seconds calculated from peak-to-peak distance for both radar-measured (blue) and belt-measured (red) breathing movements. The error is shown in black, and the mean error is 0.018 (SD 0.18) seconds.

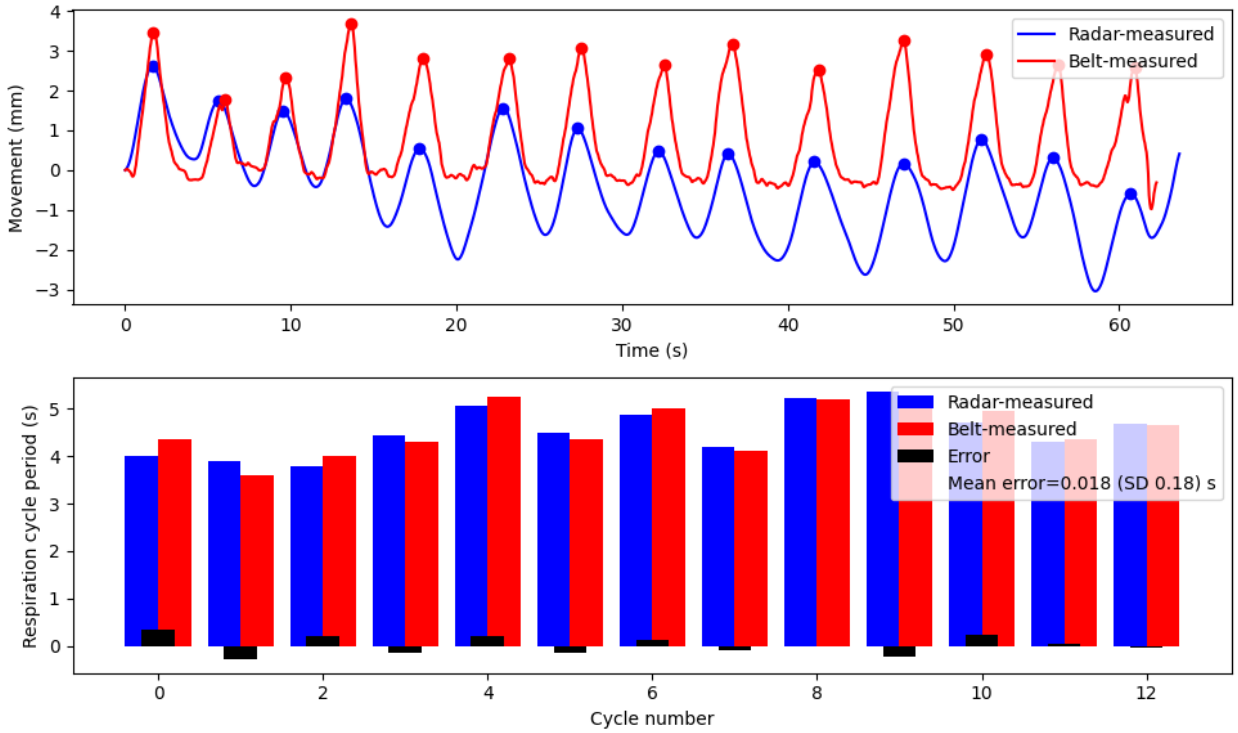
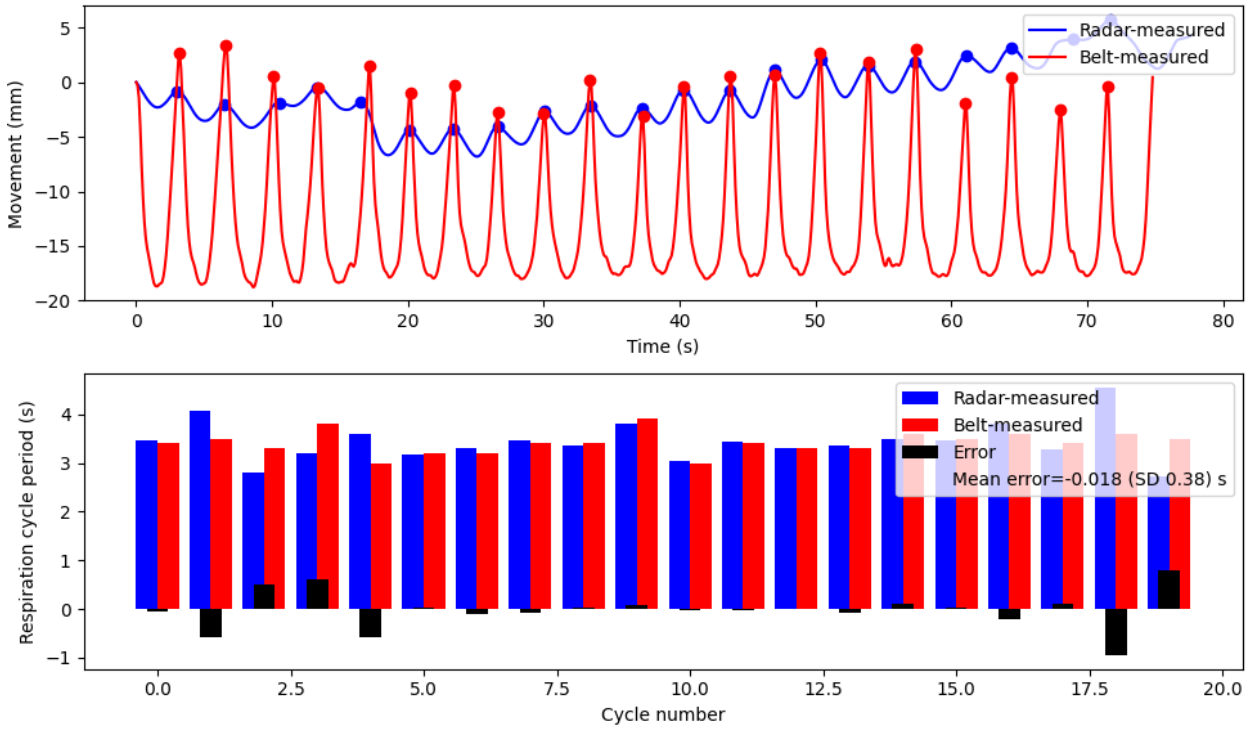


Figure 6. Representative position experiment: the top plot shows breathing motion captured by overdose radar (blue) and respiration belt (red). The participant was seated upright in the mock restroom stall with the sensor mounted at 85 cm forward, 85 cm above, and angled 30 degrees down toward the toilet seat. The second plot shows the respiration cycle period in seconds calculated from peak-to-peak distance for both radar-measured (blue) and belt-measured (red) breathing movements. The error is shown in black, and the mean error for the test is –0.018 (SD 0.38) seconds.

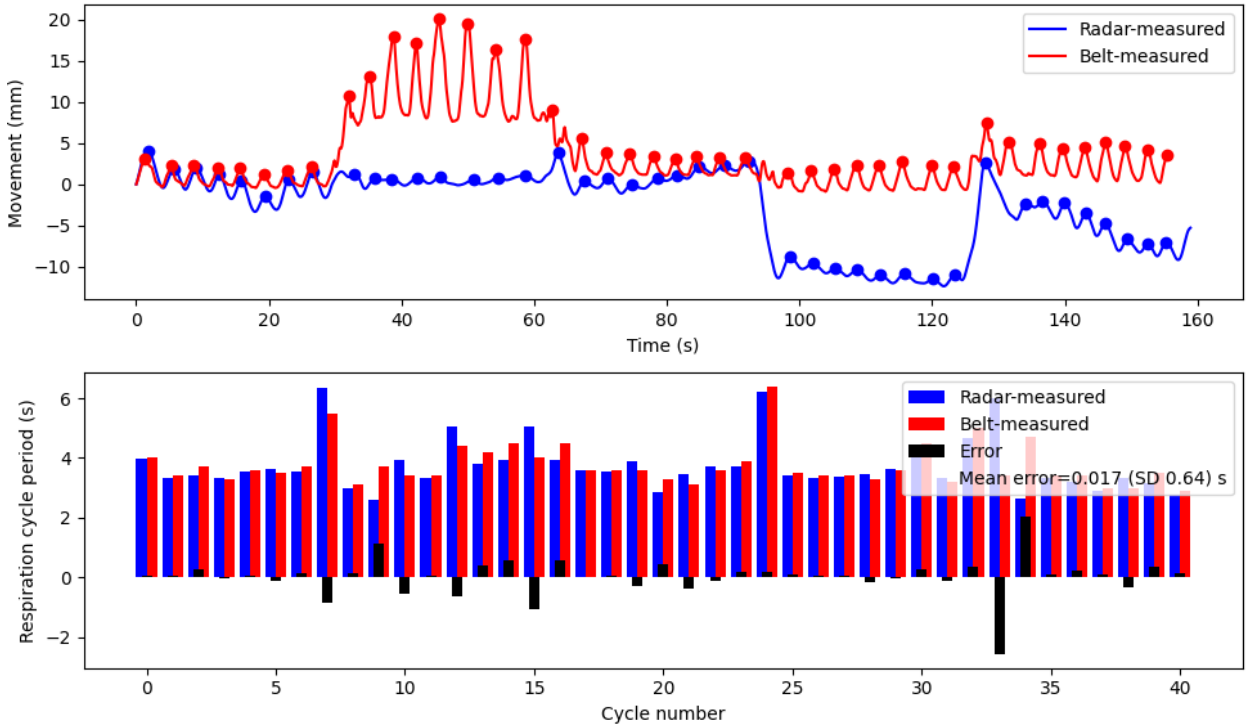


Overdose Simulation

In the final evaluation, the radar position was evaluated under simulated overdose scenarios by instructing participants to slump over in various directions while wearing the respiration belt for validation. The radar device was capable of tracking respiration even as participants transitioned into different slumped positions, causing much shallower respiration with minimal calibration delay, as demonstrated in Figure 7. Across all overdose simulation tests, which encompassed a total of 204 respiration cycles, the respiration was effectively tracked, with a mean period difference of 0.0091 (SD 0.62) seconds (Figure 4). It is worth noting that the slightly larger error can be attributed to the inherent noise present in the radar signal during positional changes, which was anticipated.

Overall, after analyzing the error for all respiration cycles, including all 3 protocols, the error was 0.0072 (SD 0.54) seconds.

Figure 7. Representative overdose simulation experiment: the top plot shows breathing motion captured by overdose radar (blue) and respiration belt (red). The participant was seated upright in the mock restroom stall for 30 seconds before slumping forward, backward, left, and right, holding each position, and breathing comfortably for 30 seconds. The bottom plot shows the respiration cycle period in seconds calculated from peak-to-peak distance for both radar-measured (blue) and belt-measured (red) breathing movements. The error is shown in black, and the mean error for the test is 0.017 (SD 0.64) seconds. This signifies that while the cycle errors can vary positively and negatively, they balance out to a small mean error of 0.017 seconds, showing that the cycle error has a minimal effect on calculated respiration rate over time.



Discussion

Overview

With further development and tuning, the ODR has the potential to detect opioid overdoses in public restrooms with accurate respiration tracking, finding a mean error of 0.0091 (SD 0.62) seconds when monitoring respiration cycle duration in overdose simulation experiments. This study has demonstrated that respiration can be accurately monitored within a restroom stall using a pulsed coherent radar sensor and a Raspberry Pi. The identification of OIRD could be used to alert staff or bystanders to the emergency and mobilize a rescue response.

Principal Findings

Following an iterative process to ensure monitoring of all slumped positions, the final device mounting position for the ODR was determined to be on the side wall of the stall. The device was mounted 85 cm above the toilet seat and 85 cm in front of the surface of the chest or center of the seat, angled down by 30 degrees. While concerns were initially raised regarding the placement of the device in locations that may not be feasible for all public restrooms, regulations for larger handicap stalls to have support bars next to the toilet on the adjacent wall alleviated these concerns as it meant that toilets would always be a standard distance from the wall [27].

The benchtop validation experiments demonstrated the capability of the radar to monitor respiration with accuracy and consistency in controlled settings. The signal processing involved in the A111 breathing algorithm might cause slight variations in the duration of the respiration cycle, as filtering the signal could lead to flattening of the peak, which may cause offset peak locations. Nevertheless, given that the respiration rate is a crucial feature for detecting overdoses, the slight shift in peak location would not significantly impact the calculated respiration rate as it would be counterbalanced by a negative shift in the adjacent period rather than accumulate drift.

The peak-to-peak distances are important, but the waveform morphology and peak magnitudes were disregarded because each method measures distinct physiological data; the radar measures the distance to the chest in millimeters, while the respiration belt gauges the external force exerted by the chest wall in Newtons. Therefore, in this preliminary testing, the peak distances measured by the radar were compared with the peak distances recorded by the respiration belt.

Exploratory validation testing was conducted to compare ODR motion-derived respiration periods with those collected using a respiration belt. Good agreement between the 2 methods was observed for both benchtop testing with the participant seated upright directly in front of the sensor as well as in a mock restroom setup with the ODR mounted to the wall. Minimal deviations in respiratory cycle duration were observed. The variance in error is due to single peaks being shifted, resulting in positive and negative errors in adjacent respiration cycles. While this error is important to identify, it does not significantly impact the overdose detection algorithm's analysis of respiration rate, as the errors are not sustained for a window long enough to affect the algorithm's functioning (15 seconds).

The principal source of error in the overdose simulation experiments, as seen in Figure 7, was related to participant movements changing to a different slumped position. It is worth noting that the device uses relative movement tracking to monitor respiration, requiring recalibration to ensure accurate tracking across different positions. The study participants encountered difficulty assuming a stable slumped position immediately after moving and often resorted to readjusting for 1-2 seconds before settling into a stable position that could be maintained for the 30-second test period. This poses a challenge in differentiating whether the radar's recovery time after each position change is delayed due to its calibration process or because the participant is still in motion. While the device's loss of respiration tracking during large movements was expected, it was able to rapidly restore accurate tracking.

Limitations and Future Work

This preliminary testing included limitations that will be considered in the design of a future pilot study conducted on a larger and more diverse sample size. First, the overdose simulation experiments were not representative of a fatal opioid overdose. Second, while the device does have a battery backup, it currently relies on being connected to a power supply. Third, the device's field of view, while tuned well for a restroom stall, would not accommodate a larger-than-standard space. Finally, we have not yet consulted with people involved in harm reduction or people who use drugs regarding the details of the implementation and alert system. We have considered how these limitations will be addressed moving forward.

The overdose simulation tests did not replicate an actual opioid overdose scenario because healthy participants found it strenuous to repeat the respiration patterns observed during fatal overdoses. Instead, the participants were instructed to breathe at a comfortable pace during the experiments. This was beneficial as they were able to breathe shallowly rather than gasping for large breaths to maintain a slow respiration rate. The shallow breathing was beneficial to include in the study because it confirmed that the radar had a high sensitivity to respiration movement, accurately measuring even very weak respiration activity. Because the preliminary testing conducted used healthy participants breathing normally when the device's primary objective is to capture irregular and unhealthy respiration patterns, a future study is planned to validate the device on unhealthy patients. In the pilot study, unhealthy respiratory patterns will be monitored by the ODR and compared against the gold standard respiratory tracking output.

The next consideration to be addressed is the power supply of the device. For these experiments, the ODR was connected to a power supply using a USB cord and included a battery backup. To facilitate the device's implementation and use, it will be modified to be battery-powered. The high-sensitivity measurements and on-board processing required to calculate respiration rates consume a considerable amount of battery power; therefore, we plan to integrate a less power-intensive presence detection component into the algorithm that triggers respiration monitoring upon detecting human entry into the restroom stall. Once an opioid overdose is detected during respiration monitoring, the device will temporarily switch back to presence detection to verify human presence before alerting. This feature optimizes battery life and reduces false positives that could occur when the device continues to track respiration after a person has left the stall.

Because this device was designed to monitor a standard public restroom stall, the position and field of view were selected accordingly. A limitation of this device's position in a larger stall is that individuals may collapse into an open space, fall to the floor, and potentially be out of the detection range. Unfortunately, this scenario cannot currently be accommodated in the device design as the detection range is limited and needs to be centered around where the participant's chest is when seated upright to capture all other positions. Importantly, the device can measure respiration in all body positions of someone using the toilet, meaning that the device is able to monitor healthy and unhealthy people regardless of their position.

While this study is limited to the development and evaluation of the ODR device, forethought has been considered about its application. We aim to gather insight from businesses and people who use drugs on how to implement this device in a way where we can maximize its life-saving potential. While it would be beneficial to notify patrons of the device's presence and functionality, it can be understood that business owners may not appreciate advertising their facilities as a safe space for drug use. Conversely, people who use drugs could be deterred by such a device if they are concerned it may result in police presence with the stigma surrounding drug use. Regardless of the mentioned concerns, both groups of people can recognize the device as a life-saving measure that can make a significant impact on the safety of the community.

The technology design is currently limited to implementation in a public restroom stall to detect opioid overdoses. However, other applications could be possible with slight design modifications to change specifications such as the detection area, device position, or battery size. These modifications could

allow the continual monitoring of other sites such as addiction treatment facility rooms, hospital rooms, and holding cells.

Conclusion

This study describes the design, development, and exploratory research of a device with promising potential to monitor respiration rate in a restroom stall setting and detect opioid overdose events. The ODR successfully monitors respiration in a restroom stall setup when the participants are in a variety of slumped positions consistent with overdoses. The development of this opioid ODR considers the physiological aspects of opioid overdoses as recommended by medical professionals, and the device is tailored to accurately monitor respiration, the main indicator of opioid overdose.

The end goal of this device concept is to add an additional layer of safety to public restrooms, enabling the timely administration of naloxone and the mobilization of rescue teams in response to overdose emergencies. Future pilot and research studies will help design the device to detect overdoses in public restroom stalls. Given the rapidly escalating opioid epidemic, it is essential to seek and develop innovative solutions that can help protect the public and facilitate prompt and effective responses to overdose events. The results of this study represent a crucial step forward in this regard, and further investigations are warranted to validate the device's performance under a variety of conditions and settings, including being evaluated on people who use drugs to ensure we can accurately track the afflicted respiration patterns.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

The device was developed by JO and AT. JO, JK, and YF were participants in the experiments. JO completed the data collection and analysis as well as the original manuscript draft. All authors participated in manuscript editing.

Conflicts of Interest

None declared.

References

1. Underlying cause of death 1999-2020. Centers for Disease Control and Prevention. 2023. URL: https://wonder.cdc.gov/wonder/help/ucd.html [accessed 2023-05-31]
2. Fentanyl. Drug fact sheet. Department of Justice/Drug Enforcement Agency. 2022. URL: https://www.dea.gov/sites/default/files/2020-06/Fentanyl-2020_0.pdf [accessed 2023-06-16]
3. Lucyk SN, Nelson LS. Novel synthetic opioids: an opioid epidemic within an opioid epidemic. Ann Emerg Med 2017;69(1):91-93 [FREE Full text] [doi: 10.1016/j.annemergmed.2016.08.445] [Medline: 27745765]
4. Salle S, Bodeau S, Dhersin A, Ferdonnet M, Goncalves R, Lenski M, et al. Novel synthetic opioids: a review of the literature. Toxicol Anal Clin 2019;31(4):298-316 [FREE Full text] [doi: 10.1016/j.toxac.2019.10.001]
5. Suzuki J, El-Haddad S. A review: fentanyl and non-pharmaceutical fentanyls. Drug Alcohol Depend 2017;171:107-116 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.11.033] [Medline: 28068563]
6. 2020 national drug threat assessment. US Department of Justice Drug Enforcement Administration. 2021. URL: https://www.justice.gov/d9/pages/attachments/2021/08/18/2020_national_drug_threat_assessment_ndta.pdf [accessed 2023-10-12]
7. Bachmutsky I, Wei XP, Kish E, Yackle K. Opioids depress breathing through two small brainstem sites. Elife 2020;9:e52694 [FREE Full text] [doi: 10.7554/eLife.52694] [Medline: 32073401]
8. Messina Z, Shapshak AH, Mills R. Anoxic Encephalopathy. Treasure Island, FL: StatPearls Publishing; 2023.
9. Sawynok J, Pinsky C, LaBella FS. On the specificity of naloxone as an opiate antagonist. Life Sci 1979;25(19):1621-1632 [FREE Full text] [doi: 10.1016/0024-3205(79)90403-x] [Medline: 229377]
10. Jordan MR, Morrisonponce D. Naloxone. Treasure Island, FL: StatPearls Publishing; 2023.
11. Naloxone for opioid overdose: life-saving science. National Institute on Drug Abuse. 2017. URL: https://nida.nih.gov/publications/naloxone-opioid-overdose-life-saving-science [accessed 2023-05-03]
12. Bernat JL. 411 - Coma, vegetative state, and brain death. In: Goldman L, Schafer AI, editors. Goldman's Cecil Medicine E-Book, Twenty Fourth Edition. Moulineaux, France: Elsevier Health Sciences; 2012:2294-2299.
13. Matskiv G, Marshall T, Krieg O, Viste D, Ghosh SM. Virtual overdose monitoring services: a novel adjunctive harm reduction approach for addressing the overdose crisis. CMAJ 2022;194(46):E1568-E1572 [FREE Full text] [doi: 10.1503/cmaj.220579] [Medline: 36442886]
14. Kennedy MC, Karamouzian M, Kerr T. Public health and public order outcomes associated with supervised drug consumption facilities: a systematic review. Curr HIV/AIDS Rep 2017;14(5):161-183 [FREE Full text] [doi: 10.1007/s11904-017-0363-y] [Medline: 28875422]

15. Milloy MJS, Kerr T, Mathias R, Zhang R, Montaner JS, Tyndall M, et al. Non-fatal overdose among a cohort of active injection drug users recruited from a supervised injection facility. Am J Drug Alcohol Abuse 2008;34(4):499-509 [FREE Full text] [doi: 10.1080/00952990802122457] [Medline: 18584579]
16. Fozouni L, Buchheit B, Walley AY, Testa M, Chatterjee A. Public restrooms and the opioid epidemic. Subst Abus 2020;41(4):432-436 [FREE Full text] [doi: 10.1080/08897077.2019.1640834] [Medline: 31368865]
17. Wolfson-Stofko B, Bennett AS, Elliott L, Curtis R. Drug use in business bathrooms: an exploratory study of manager encounters in New York City. Int J Drug Policy 2017;39:69-77 [FREE Full text] [doi: 10.1016/j.drugpo.2016.08.014] [Medline: 27768996]
18. Brave technology co-op. Brave Sensors. URL: https://www.brave.coop/sensor [accessed 2023-06-16]
19. Capelouto JD. Anti-motion detector preventing overdose-related deaths in public bathrooms. The Boston Globe. 2018. URL: https://www.bostonglobe.com/metro/2018/06/26/anti-motion-detector-aims-prevent-overdoses-public-bathrooms/xQEHDyAkcBuMQRFKhAbI4H/story.html [accessed 2023-06-10]
20. Teck JTW, Oteo A, Baldacchino A. Rapid opioid overdose response system technologies. Curr Opin Psychiatry 2023;36(4):308-315 [FREE Full text] [doi: 10.1097/YCO.0000000000000870] [Medline: 37185583]
21. Kirby RR. Ventilation Capnography and respiratory assessment outside of the operating room. In: Paulus DA, Gravenstein JS, Jaffe MB, Gravenstein N, editors. Capnography. 2nd Edition. Cambridge, United Kingdom: Cambridge University Press; 2011:11-18.
22. Chan J, Iyer V, Wang A, Lyness A, Kooner P, Sunshine J, et al. Closed-loop wearable naloxone injector system. Sci Rep 2021;11(1):22663 [FREE Full text] [doi: 10.1038/s41598-021-01990-0] [Medline: 34811425]
23. Singh NK, Sidhu GK, Gupta K. Current and future perspective of devices and diagnostics for opioid and OIRD. Biomedicines 2022;10(4):743 [FREE Full text] [doi: 10.3390/biomedicines10040743] [Medline: 35453493]
24. Nandakumar R, Gollakota S, Sunshine JE. Opioid overdose detection using smartphones. Sci Transl Med 2019;11(474):eaau8914 [FREE Full text] [doi: 10.1126/scitranslmed.aau8914] [Medline: 30626717]
25. Tsang VWL, Papamihali K, Crabtree A, Buxton JA. Acceptability of technological solutions for overdose monitoring: perspectives of people who use drugs. Subst Abus 2021;42(3):284-293 [FREE Full text] [doi: 10.1080/08897077.2019.1680479] [Medline: 31657675]
26. Li C, Lubecke VM, Boric-Lubecke O, Lin J. A review on recent advances in doppler radar sensors for noncontact healthcare monitoring. IEEE Trans Microwave Theory Techn 2013;61(5):2046-2060 [FREE Full text] [doi: 10.1109/tmtt.2013.2256924]
27. O. Reg. 368/13: Building Code. Ontario Government. 2013 Dec 27. URL: https://www.ontario.ca/laws/regulation/r13368 [accessed 2023-10-12]
28. Developer. Acconeer. 2021. URL: https://developer.acconeer.com/ [accessed 2023-02-20]
29. Hillman KM, Chen J, Jones D. Rapid response systems. Med J Aust 2014;201(9):519-521. [doi: 10.5694/mja14.01088] [Medline: 25358575]
30. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-council policy statement: ethical conduct for research involving humans. Panel on Research Ethics. 2022. URL: https://ethics.gc.ca/eng/tcps2-eptc2_2022_chapter6-chapitre6.html [accessed 2023-10-12]

Abbreviations

bpm: breaths per minute
ODR: opioid overdose detection radar
OIRD: opioid-induced respiratory depression

Edited by T Leung; submitted 11.08.23; peer-reviewed by K Gupta, J Tay Wee Teck, C Baxter; comments to author 09.09.23; revised version received 27.09.23; accepted 06.10.23; published 24.10.23.

Please cite as:
Oreskovic J, Kaufman J, Thommandram A, Fossat Y
A Radar-Based Opioid Overdose Detection Device for Public Restrooms: Design, Development, and Evaluation Study
JMIR Biomed Eng 2023;8:e51754
URL: https://biomedeng.jmir.org/2023/1/e51754
doi:10.2196/51754
PMID:38875668

Original Paper

Continuous Critical Respiratory Parameter Measurements Using a Single Low-Cost Relative Humidity Sensor: Evaluation Study

Fabrice Vaussenat[1], BSc, PhD; Abhiroop Bhattacharya[1], BSc; Julie Payette[1], BSc; Jaime A Benavides-Guerrero[1], BSc, PhD; Alexandre Perrotton[1], BSc; Luis Felipe Gerlein[1], BSc, MSc; Sylvain G Cloutier[1], BSc, MSc, PhD

Department of Electrical Engineering, École de Technologie Supérieure, Montreal, QC, Canada

Corresponding Author:
Fabrice Vaussenat, BSc, PhD
Department of Electrical Engineering
École de Technologie Supérieure
1100 Rue Notre-Dame Ouest
Montreal, QC, H3C 1K3
Canada
Phone: 1 514 451 1970
Email: fabrice.vaussenat@lacime.etsmtl.ca

Abstract

Background: Accurate and portable respiratory parameter measurements are critical for properly managing chronic obstructive pulmonary diseases (COPDs) such as asthma or sleep apnea, as well as controlling ventilation for patients in intensive care units, during surgical procedures, or when using a positive airway pressure device for sleep apnea.

Objective: The purpose of this research is to develop a new nonprescription portable measurement device that utilizes relative humidity sensors (RHS) to accurately measure key respiratory parameters at a cost that is approximately 10 times less than the industry standard.

Methods: We present the development, implementation, and assessment of a wearable respiratory measurement device using the commercial Bosch BME280 RHS. In the initial stage, the RHS was connected to the pneumotach (PNT) gold standard device via its external connector to gather breathing metrics. Data collection was facilitated using the Arduino platform with a Bluetooth Low Energy connection, and all measurements were taken in real time without any additional data processing. The device's efficacy was tested with 7 participants (5 men and 2 women), all in good health. In the subsequent phase, we specifically focused on comparing breathing cycle and respiratory rate measurements and determining the tidal volume by calculating the region between inhalation and exhalation peaks. Each participant's data were recorded over a span of 15 minutes. After the experiment, detailed statistical analysis was conducted using ANOVA and Bland-Altman to examine the accuracy and efficiency of our wearable device compared with the traditional methods.

Results: The perfused air measured with the respiratory monitor enables clinicians to evaluate the absolute value of the tidal volume during ventilation of a patient. In contrast, directly connecting our RHS device to the surgical mask facilitates continuous lung volume monitoring. The results of the 1-way ANOVA showed high *P* values of .68 for respiratory volume and .89 for respiratory rate, which indicate that the group averages with the PNT standard are equivalent to those with our RHS platform, within the error margins of a typical instrument. Furthermore, analysis utilizing the Bland-Altman statistical method revealed a small bias of 0.03 with limits of agreement (LoAs) of –0.25 and 0.33. The RR bias was 0.018, and the LoAs were –1.89 and 1.89.

Conclusions: Based on the encouraging results, we conclude that our proposed design can be a viable, low-cost wearable medical device for pulmonary parametric measurement to prevent and predict the progression of pulmonary diseases. We believe that this will encourage the research community to investigate the application of RHS for monitoring the pulmonary health of individuals.

(JMIR Biomed Eng 2023;8:e47146) doi:10.2196/47146

KEYWORDS

relative humidity sensor; design; develop; development; tidal volume; pulmonary volume; COPD; pulmonary; respiratory; sensor; sensors; wearables; humidity; medical device; development; breathing; wearable; ventilation; air

Introduction

Theory

In 2020, respiratory disorders impacted approximately 550 million individuals globally and caused 4 million annual deaths. The COVID-19 pandemic led to an increase in health care expenditure, particularly in the field of respiratory diseases [1]. The pandemic expedited the development of respiratory diseases, and despite persistent respiratory and neurological problems, many patients have been discharged [2,3]. Simultaneously, obstructive sleep apnea (OSA) [4] affects a significant proportion of adults and is related to increased morbidity and mortality worldwide.

These significant global health care issues warrant the pursuit of solutions to prevent and optimize health care pathways [5]. Lung volume measurement is especially important for patients with respiratory diseases [6,7], with OSA and under ventilation [8], or in intensive care units (ICUs) [9]. Spirometers can evaluate breathing in patients with asthma or chronic obstructive pulmonary disease (COPD) [10]. Patients with OSA using an automatic positive airway pressure (A-PAP) or continuous positive airway pressure (C-PAP) device must be awake to execute deep breathing (inhale and exhale) to be tested with a spirometer [11]. Wearable devices that assess respiratory rate (RR) using validated methods have been developed recently [12-15], but they do not measure lung volume [16]. Thus, a wearable device that gives comprehensive lung volume data to improve quality of life, monitor remotely, and avoid respiratory disease progression would be highly desirable [17,18].

In this work, we present a simple and inexpensive sensor platform that can be used to quantify pulmonary inspiration, expiration, and lung volumes. Our device uses a relative humidity sensor (RHS) to detect breathing and calculate tidal volumes (TVs), expiratory reserve volumes (ERVs), and inspiratory reserve volumes (IRVs) [19]. To the best of our knowledge, the innovative aspects of this study are the direct measurement of respiratory cycles and the exact derivation of TD, ERV, IRV, and vital capacity (VC) data from the calculation of the breathing surface [16,20].

Prior Work

There are numerous techniques for assessing respiratory function and detecting lung disorders such as COPD and asthma [1]. Pulmonary function tests need accurate breathing volume and flow measurements using a basic spirometer [21], requiring the patient to inhale deeply and then expel as forcefully as possible via the mouthpiece over a period of time. This is an inexpensive, noninvasive test that can be administered in a medical facility or at home. However, the requirements of this test can still be hard for some patients who cannot fully empty their lungs during the procedure. This limits the usefulness of this well-known diagnostic tool [22]. Additionally, laboratory blood tests can be used to evaluate respiratory health. However, because this is an intrusive procedure, it cannot be used to indefinitely monitor patients outside of hospitals.

The number of rib cage movements per minute is another crucial indicator that indicates respiratory and heart health, via RR [23,24]. RR monitoring can be accomplished using ICU-specific equipment [25]. Mathematical correlations of photoplethysmography (PPG) and electrocardiography (ECG) data yield accurate RR values [26-28]. Numerical methods estimate the RR from PPG and ECG using the following 3 physiological modulations of breathing: amplitude modulation, frequency modulation, and baseline wander [29,30]. The noise in the PPG and ECG signals affects the accuracy of these RR measures [31], and considerable signal processing is needed to extract meaningful information from the noise and improve the measurements [32,33].

Measuring RR alone cannot assess lung capacity, a critical indication of COPD status and development [34]. Respiratory depression can be detected by lung volume measurements such as TV, ERV, IRV, and VC. Pulmonary function tests measure lung volume, capacity, flow rates, and gas exchange. Spirometry has low accuracy and significant latency and cannot be utilized during sleep [35]. Plethysmography, which measures intrathoracic gas during airflow obstruction, is used to calculate lung volume [36]. Modern wearable devices like CO_2 gas sensors analyze CO_2 or O_2 fluctuations during inspiration and expiration to estimate lung capacities [37]. Computed tomography radiography, which is invasive and time-consuming, is another option [38,39].

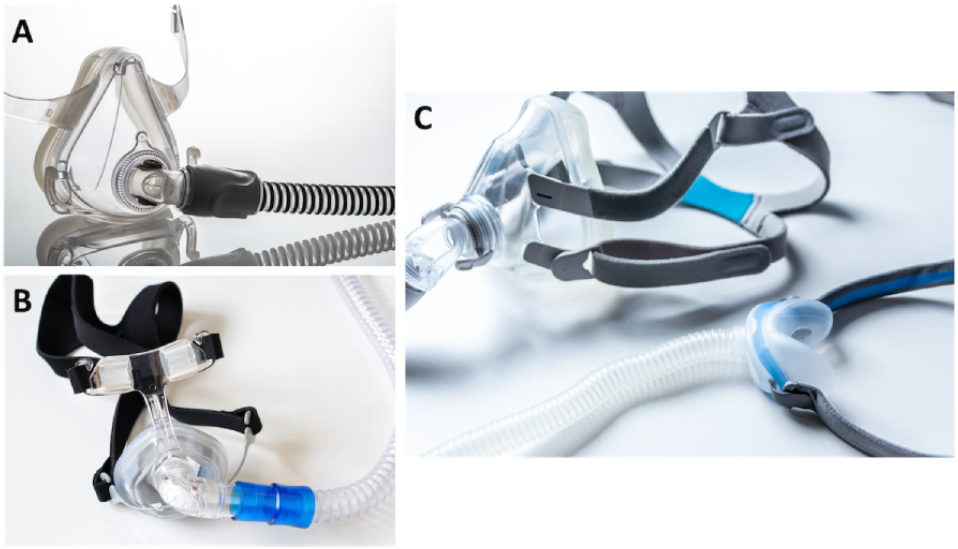
A much easier method measures the RR directly from the moisture content of the breath [20,40,41]. Patients are typically attached to various monitoring devices in ICUs and during surgery to continually monitor their pulse, blood pressure, breathing rate, and oxygen saturation. Typically, humidity sensors are included in the tube adaptor entry of respirator face masks (Figure 1) [42].

Continuously measuring the humidity of the exhaled air provides an accurate measurement of the patient's RR. In addition, it was recently shown that TV might potentially be effectively estimated from surface measurement during the normal breathing cycle [43]. Exhaled air has a relative humidity (RH) of 100% and is saturated with water [44]. This exhaled humidity is a function of pulmonary capacity and is proportional to RR and lung volume [20,45]. Most crucially, atmospheric pressure, external temperature, and sex-dependent fluctuations alter the signal amplitude's maximum RH content [46].

A huge network of internet-connected objects, including Bluetooth Low Energy devices, sensors, and global positioning systems, is the goal of emerging Internet of Things (IoT) paradigms [40,41]. Medical IoT devices that use cloud compute power could improve chronic respiratory illness detection and therapy. In medical IoT devices, embedded electronics like accelerometers [17,42-44] and temperature sensors [45,46] are used in new ways. Using a chest-mounted belt, an accelerometer may record rib cage movement to determine breathing rate. The inhalation-exhalation temperature differential can also be monitored [47]. We can estimate the key pulmonary indicators from studying the relationship among temperature, pressure, and humidity of a person [48]. The only IoT-compatible methods for measuring lung capacity are somewhat sophisticated spirometry [49], capnometry [50], and impedance pneumography devices [51,52]. In this new paradigm, data

collection occurs simultaneously in real time and over an extended period of time, enabling a move from a reactive treatment strategy to an early warning and detection mode that maximizes results while minimizing the related human and financial costs [53].

Figure 1. Commercial face masks typically used (A) in intensive care units with the (B) ventilation connector including respiratory rate, humidity, and pressure measurement sensors and (C) at home by patients with obstructive sleep apnea. All products are manufactured by ResMed Ltd.



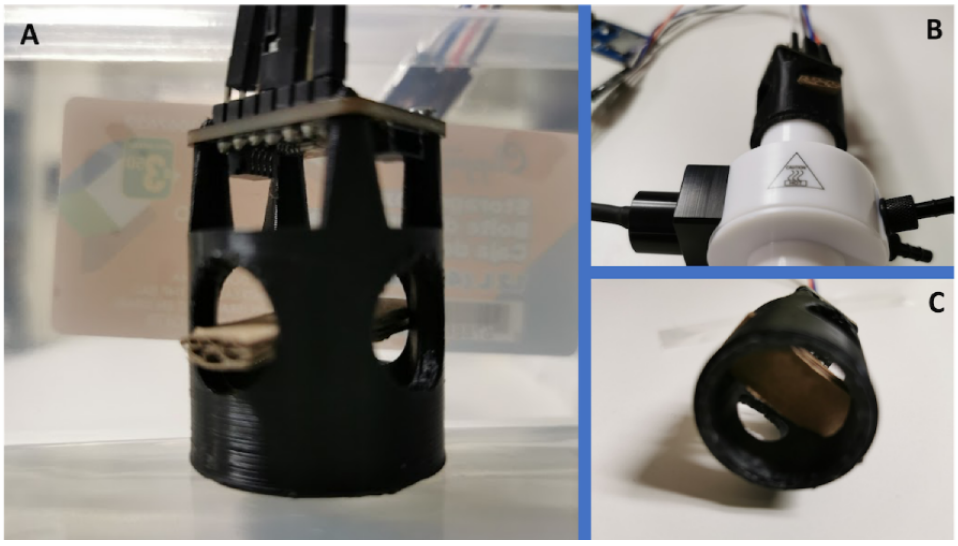
Hypothesis

The key hypothesis of our work was that the changes in RH during breathing provide an indication of pulmonary disorders in a patient. Moreover, we also believed that the change in RH during deep inhale-exhale cycles of breathing can be used to measure pulmonary volume, especially TV, ERV, and IRV. The combination of measurements provides the total pulmonary VC. To prove our hypothesis, we designed a low-cost wearable device that uses a single RHS to provide accurate and sophisticated lung volume (TV, ERV, IRV, and VC) and RR measurements. To validate our strategy, we incorporated the electronic prototype into an OSA face mask with our own flow adapter (Figure 2), which is designed to prevent moisture retention and keep the sensor close to the mouth and nose.

The gap between the mask and humidity sensor was defined to avoid saturation of the humidity sensor. In fact, we tested different tube sizes to support the sensor, so as to avoid phenomena saturation during deep exhalations. We based our adapter on the venturi effect [54]. Various flow measurements were carried out to create this adapter, from which the sensor is attached and onto which the mask is fitted. This method is also used to easily change the mask from one person to another.

We used statistical analysis methods including Bland Altman, 1-way ANOVA, and box plots to validate our results. The preliminary results of the experiments with multiple participants showed that our hypothesis was correct, and the same was corroborated by the statistical tests. With more participants and sophisticated models, we will be able to classify pulmonary disorders based on changes in RH.

Figure 2. Humidity sensor adapter connected to the obstructive sleep apnea face mask to reduce the amount of trapped moisture: (A) outside view of the airflow adapter with the reduction and humidity sensor, (B) inside view of the airflow adapter with the reduction and humidity sensor, and (C) airflow adapter with the reduction and humidity sensor and the heater connected.



Methods

Experimental Setup

We used the humidity sensor to estimate lung volumes, a key respiratory function, by monitoring inspiratory and expiratory humidity [47]. Figure 3 shows the electronic prototype inserted into an OSA face mask to keep the sensor close to the mouth and nose. The BME280 humidity sensor was protected from face mask moisture by a 3D-printed airflow adapter. Bluetooth data transmission and SD card data recording were enabled by an Arduino Nano Bluetooth Low Energy interfaced to the sensor. Figure 3A provides a schematic overview, and Figure 3B depicts our prototype. We developed the device using a proven health care IoT architecture strategy that combines data acquisition, low energy, and embedded systems [55]. Figure 3A shows the medical-compatible BME280 sensor measuring RH, ambient temperature, and barometric pressure [56,57]. Additionally, it has quick start-up and recovery times, with a 63% recovery occurring in roughly 1 second [58]. BME280 is a widely used, commercial humidity sensor that retains a long-term stability of ±0.5% RH per year and a low signal-to-noise ratio of 0.02% RH, as provided in the sensor's data sheet [58]. The range of the RH measurement at an overall level is 89 (SD 7.8).

The sensor's embedded data preprocessing is enabled by an NRF52832 microcontroller unit used in other medical devices [59]. The microcontroller unit stores data on the SD card for later retrieval when a Bluetooth connection is unavailable. The smartphone's wireless connection and real-time data recording app were developed in Android Studio. The IEEE 754 standard requires converting the data transmitted via Bluetooth from HEX to float [60]. Data cleaning, processing, and automated parameter calculation were done via Python scripting.

We worked on the raw data to highlight all possible anomalies and artifacts. As the humidity sensor is placed on the front of the mask, separated by a support to prevent saturation of the sensor during deep exhalation, no movement artifacts were detected. The only artefacts measured were those associated with breathing, which can sometimes present saccades that highlight the absence of ventilatory recovery.

A standard commercial breathing monitoring device (pneumotach [PNT]; Hans Rudolph) was used to independently calculate the lungs' volume and RR for comparison with our sensing platform's results. Figure 4A shows the PNT controller and heater [61]. Figure 4B shows the Hans Rudolph PA-1 PNT Amplifier [62]. Standard PNT devices monitor respiratory parameters at 1 kHz [62,63].

For best results, our sensor should be placed behind the PNT heater using an adapter to lower face mask humidity. This adaptor in the face mask allowed precise humidity measurement during respiration, enhancing sensor sensitivity and limiting saturation. This design integrates our humidity measuring device into the surgical face mask and connects via the PNT, as shown in Figures 4 and 5. All participants were instructed to breathe deeply every 30 seconds to determine TV, IRV, and ERV, then they were instructed to breathe normally for the rest of the 15-minute test. The VC is the sum of the IRV and ERV parameters. The RH sensor sampled data at 5 kHz, and the PNT sampled data at 1 kHz during the tests.

Figure 3. (A) Schematics of the proposed health care Internet of Things architecture and (B) our prototype. BLE: Bluetooth Low Energy; OSA: obstructive sleep apnea.

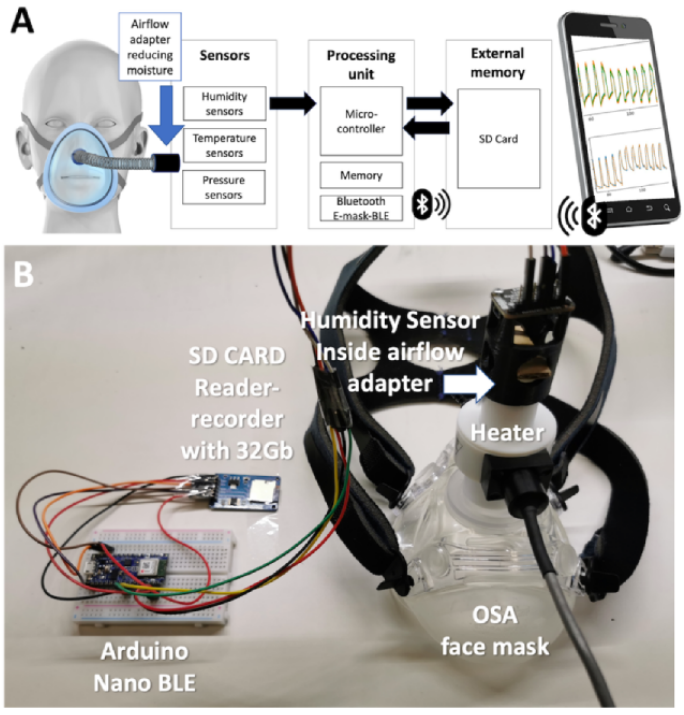


Figure 4. Commercial pneumotach (PNT) used for reference baseline measurements: (A) PNT controller and heater and (B) PA-1 PNT amplifier. The photos were taken from Hans Rudolph website [61].

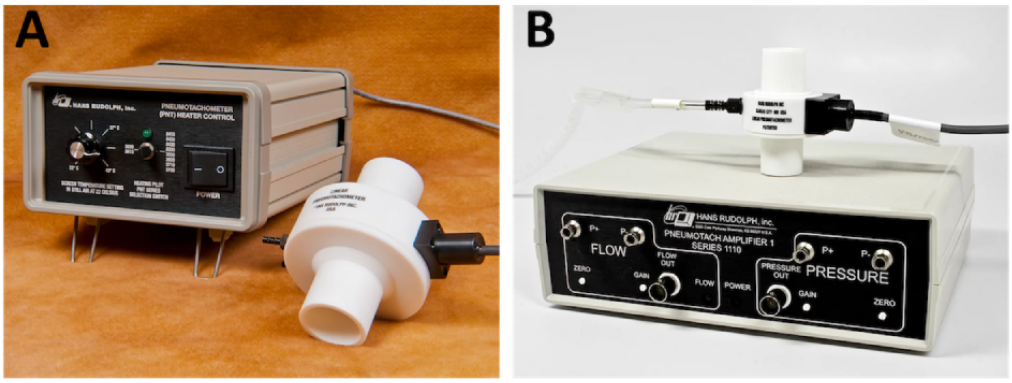
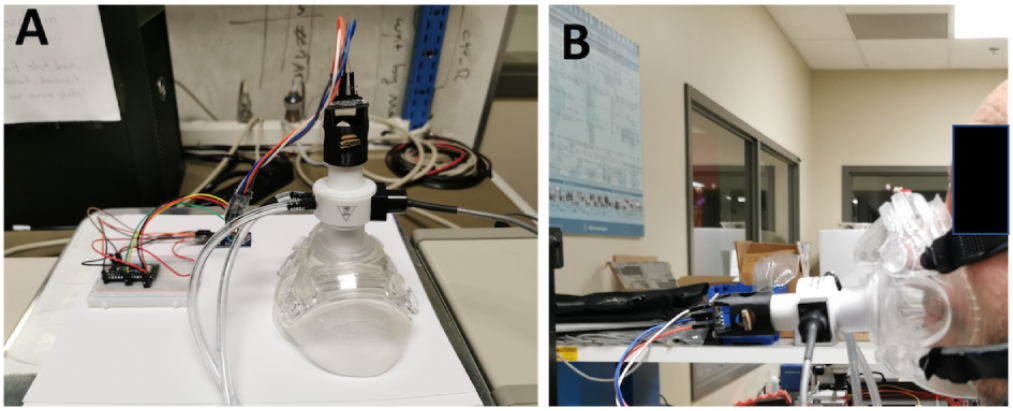


Figure 5. Testing procedures, including the (A) global measurement setup and (B) face mask connected to the heater and relative humidity sensor adapter.



Recruitment

All study participants, comprising 5 men and 2 women, were healthy and had no preexisting respiratory conditions. Their mean height was 171 (SD 10) cm, and their mean weight was 75 (SD 20) kg. The mean BMI was 25 (SD 4) kg/m^2. The mean age of this population was 38 (SD 8) years. All participants were tested under identical conditions under medical supervision.

The experiment was conducted according to a specific protocol that ensured "identical" conditions for all participants involved. First, implementation of the mask was dealt with in 2 parts. Each participant was given a dedicated, single-use mask. The sensor part of the mask, affixed to the mount, was reusable and did not need to be cleaned. To monitor its operation, a data acquisition system was used to confirm that the sensor maintained the same humidity saturation as the room environment.

All measurements were taken at the same RH, in the same location, and under the same automated climatization conditions. These conditions included temperature, humidity, and atmospheric pressure. To ensure this uniformity, all measurements were conducted on the same day.

Regarding the selection of participants, they were chosen based on specific criteria to ensure their homogeneity. Participants had no respiratory history and were in good physical condition. They all used the same seated position and the same chair and were exposed to the same environmental conditions as other participants.

In terms of equipment setup, the equipment was placed uniformly for all participants. A preliminary step in the protocol confirmed that the baseline conditions (temperature and humidity in the room) were the same for all participants before starting the experiment. A new mask was used for each measurement to eliminate any risk of cross-contamination or interference.

Physical parameters such as height, weight, and blood pressure were measured before data collection, and all participants were allowed 5 minutes of relaxation before starting.

Ethics Approval

The experiments presented in this paper were approved by the research ethics committee at École de Technologie Supérieure (approval number: H20230603).

Statistical Analysis

We compared the means and IQRs from the commercial PNT and our prototype using a box plot. We also compared the deep breathing area (DBA) and RR series using a 1-way ANOVA [64]. The hypothesis was tested with an optimal *P* value of .05 in the ANOVA analysis. DBA and RR Bland-Altman charts were plotted against the PNT for all participants.

Methodology

A cleaning protocol was used for the mask before each measurement to remove contamination risks. The following method derives respiratory parameters from sensor-measured RH and PNT: According to the experimental section, the RH was measured throughout the 15-minute test. Counting maxima over a time segment gives the RR. We calculated TV by calculating the area under the curve between 2 consecutive minima using normal breathing data. All lung volumetric parameters were calculated using the rectangle method, which is presented in mathematical form in equation 1 [65].

where *i* indicates the sample number, *x* indicates the time duration of measurement in seconds, *f* indicates the function to compute the RH value at time *x*, and *m* indicates the total number of samples.

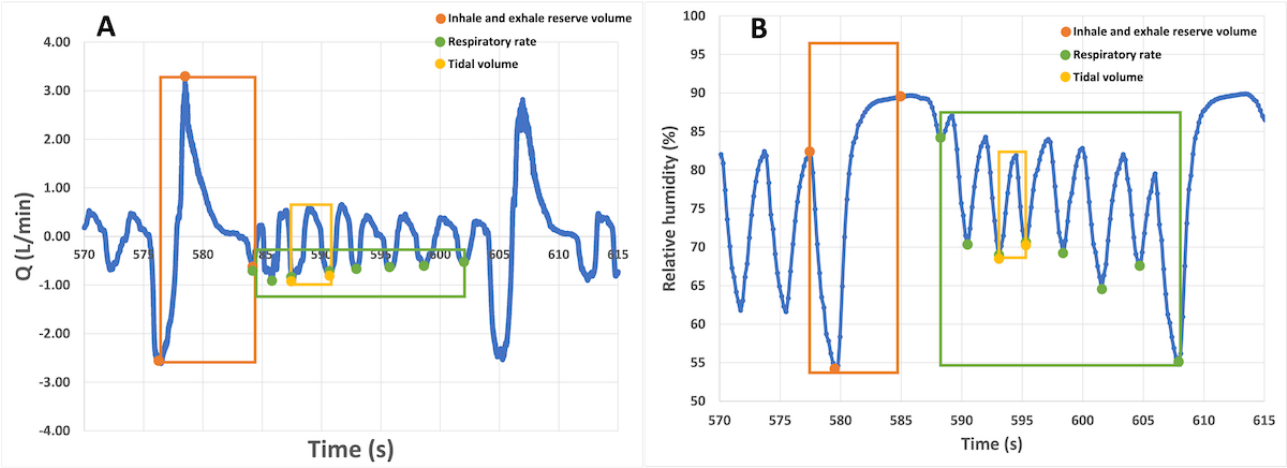
Deep inhalation and exhalation areas determine IRV and ERV parameters, and the VC is the sum of both. The beginning and end of each deep inspiration and expiration cycle were indicated by a sign change in the signal's second derivative [5]. Counting minima in beats per minute over a chosen time can also determine the RR, as shown in equation 2.

where *x* indicates the time duration of thr measurement in seconds.

Indeed, RR and TV have been used to validate respiratory parameters. As shown in Figure 6, lung volumes and RR were calculated by measuring deep inhalation and exhalation breathing and the respiratory cycle over 1 minute. The room temperature remained at 21.5 °C. External humidity control in the test room was turned off because the mask uses a heater to control humidity.

Our prototype's reset button synchronized data collection timers with the PNT. A computer saved the PNT's benchmark respiratory parameters. Volume capacity is indicated by the areas of the orange rectangle in Figure 6A, which represents the closed DBA signal that begins with inhalation and ends with deep exhalation. RH and PNT had the best correlation (R=0.84) when calculating the area using the rectangle method. The triangle method had a lower correlation coefficient (R=0.40).

Figure 6. Respiratory parameters calculations from the (A) commercial pneumotach (PNT) recordings and (B) relative humidity sensor–based prototype.

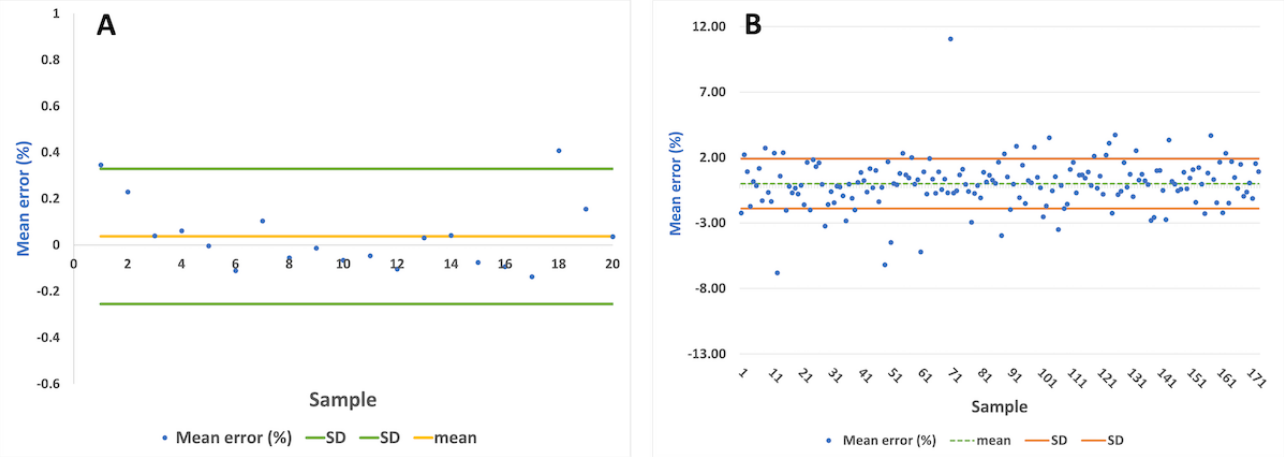


Results

Bland-Altman Analysis

Population-based Bland-Altman analysis was used for the participant data [66]. We used the mean difference and limits of agreement (LoAs) to quantify the humidity sensor-PNT correlation. Bland-Altman graph analysis is a simple way to assess the bias between the average differences and estimate an interval of agreement in which 95% of the sensor data differences fall relative to the PNT data. Both data sets were analyzed using unit difference and percent difference charts. Figure 7 shows the Bland-Altman analysis for the DBA-normalized plots to help estimate the within-subject and between-subject variability. Bias and variability terms were fixed to zero when they did not increase the objective function by more than 2 points, and bias and variance variability were assumed to be normal and log-normally distributed [66]. The DBA's overall analysis yielded a bias of 0.03 with LoAs of –0.25 and 0.33. The RR bias was 0.018, and the LoAs were –1.89 and 1.89.

Figure 7. The Bland-Altman plot and calculations comparing the values from the relative humidity sensor with those from the pneumotach for the (A) deep breathing area and (B) respiratory rate. The blue line indicates the bias, and the dotted lines indicate the limits of agreement.



Statistical Analysis

We compared the PNT's experimental DBA and RR data with our RHS-based IoT prototype. Figure 8 shows box plots comparing the medians and IQRs from the commercial PNT with those from our prototype. The average DBA values from the PNT and our sensor were 0.56 (SD 0.29) and 0.59 (SD 0.28), respectively, showing similar values. The average RR with PNT was 17.61 (SD 1.73), and the average RR with our prototype was 17.58 (SD 1.71). This shows that our IoT sensor can accurately measure the participants' respiratory parameters.

Furthermore, the 1-way ANOVA analysis was useful to compare the similarities between the DBA and RR data sets, as shown in Table 1 [64]. After reviewing the ANOVA results, we still needed to understand subgroup differences among the different experimental and control groups. For the ANOVA, we used an optimal *P* value of .05 to test the hypothesis. The results showed values of $F_{1,39}$=0.016 and $F_{1,347}$=0.01, indicating that the differences between the group averages were negligible [67]. We could not reject the null hypothesis because the corresponding *P* values of .64 for DBA and .89 for RR were greater than .05 [67]. Thus, we concluded that there were no statistically significant differences between the mean DBA and RR measurements taken by the commercial PNT and our IoT sensor prototype.

Finally, Figure 9 shows the root mean square error (RMSE) for the DBA and RR measurements, when comparing the commercial PNT with our RHS prototype. The trend includes the highest and lowest RR values. Abnormal breathing patterns during DBA and RR measurements cause data anomalies. The PNT and RHS prototype anomalies are shown in Figure 10. Errors in the DBA and RR can also be caused by variations in the participant's breathing pattern during the test [65]. This pattern appeared only once during our measurement, represented by the outlier points outside of the RMSE and DBA regions present in Figure 10. Most devices are very sensitive to changes in breathing patterns, especially during flow measurements because the face mask makes it hard to breathe normally. It is possible to ameliorate this error by averaging the volume and RR measurements over a longer period. Future generations could leverage sensor fusion, multitenancy (sequential usage of different sensors), or deep learning predictive value structures to continuously monitor patients' vital signs [68]. After examining the volumes (ERV, IRV, TV, VC) and RR measurements, we concluded that our IoT RHS-based device offers a unique way to properly measure essential respiratory parameters using a low-cost sensor and without heavy-duty medical devices.

Figure 8. Box plots comparing the calculated (A) deep breathing area (DBA) and (B) respiratory rate values from the relative humidity sensor prototype with those from the pneumotach for all participants in the test data set.

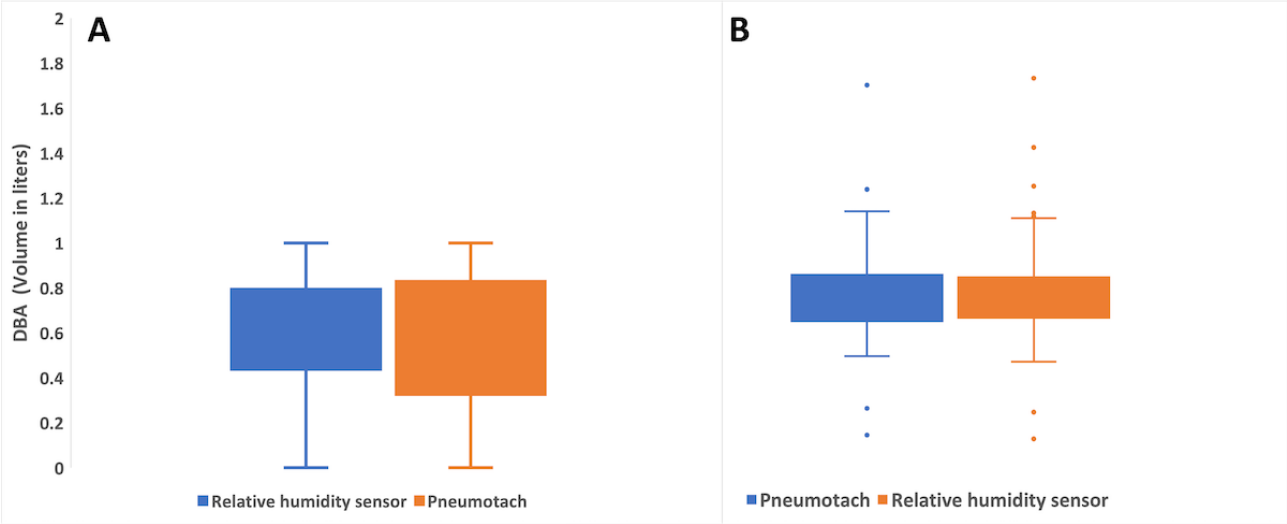


Table 1. Results from the 1-way ANOVAs of respiratory volume and rate measurements.

Source of variation	SS[a]	*df*	MS[b]	*F*	*P* value
Respiratory volume (L)					
Between groups	0.02	1	0.014	0.016	.68
Within groups	3.13	38	0.08	__[c]	—
Total	3.15	39	—	—	—
Respiratory rate (beats per minute)					
Between groups	0.002	1	0.001	0.01	.89
Within groups	39.09	346	0.11	—	—
Total	39.09	347	—	—	—

[a]SS: sum of squares.

[b]MS: mean squares.

[c]Not applicable.

Figure 9. The plot shows the root mean square error (RMSE) of the (A) deep breathing area (DBA) and (B) respiratory rate (RR) values, comparing our relative humidity sensor (RHS) prototype against the commercial pneumotach (PNT).

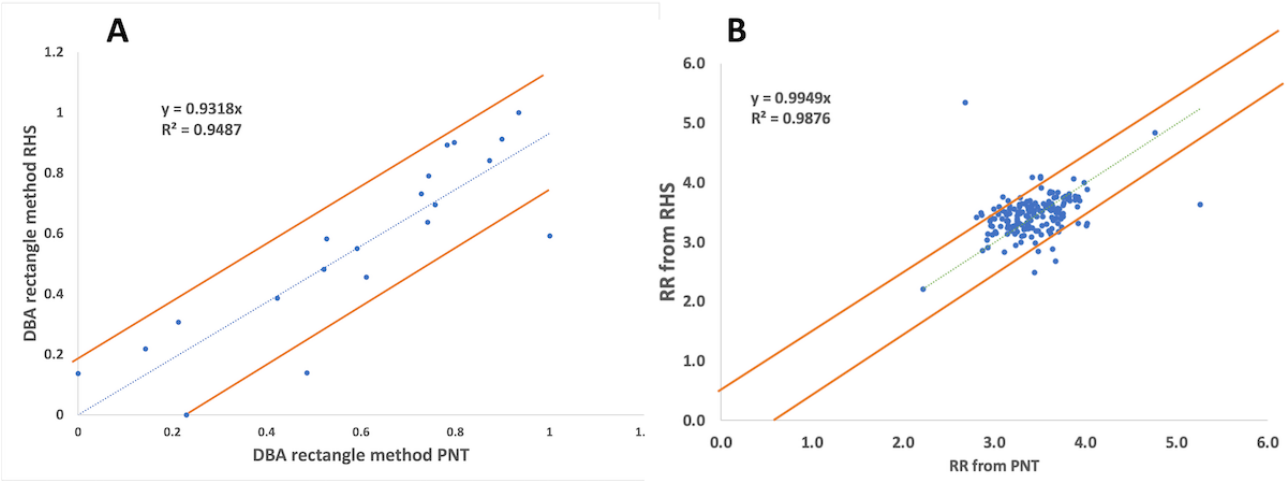
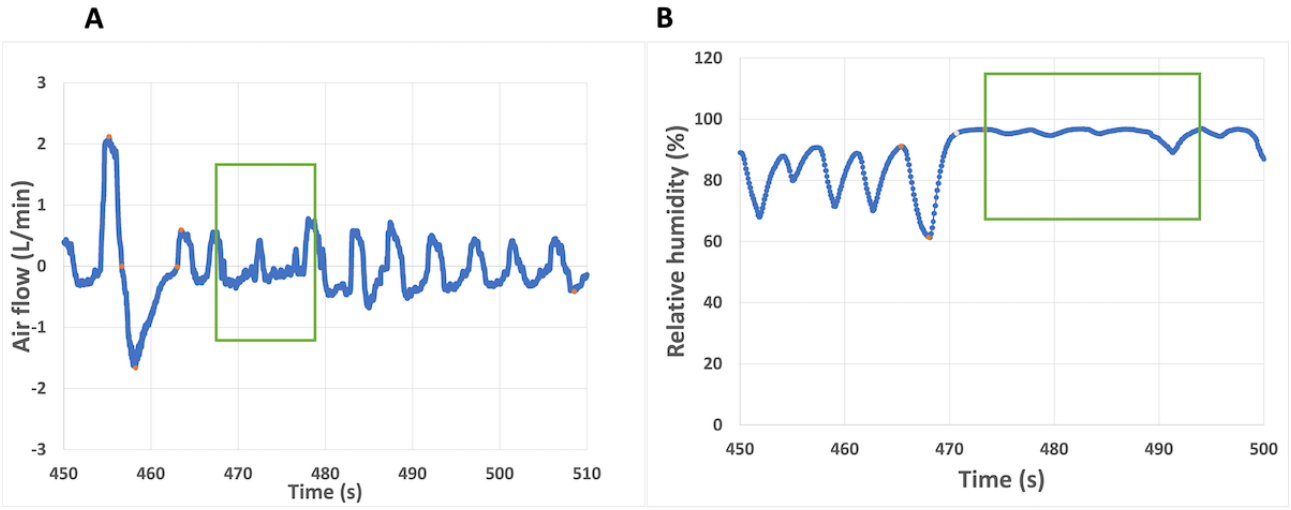


Figure 10. Two examples of breathing pattern anomalies during the deep breathing area and respiratory rate measurements using (A) the commercial pneumotach device and (B) our own relative humidity–based sensing device. The maximum point shows the value of the inspiratory reserve volume, while the minimum represents the expiratory reserve volume during deep breathing measurements; during a normal breathing pattern, the graph shows the tidal volume.



Discussion

Principal Findings

First, we showed that we can measure respiratory parameters using the proposed RHS-based device. To calculate the TV, we used the area under the RH curve using the rectangle method, as described. Second, we compared the measurements from the proposed device with those from the gold standard PNT for all participants. The Bland-Altman analysis showed that the measurements were within the LoAs and the bias was very low. Furthermore, the results from the ANOVA indicated that the *P* values were above the threshold value of $P<.05$. Both statistical experiments confirmed that the group average measurements from the PNT and our proposed device were similar within statistical limits (ie, we can use our proposed device to measure respiratory parameters with an accuracy that is close to that of the gold standard). This is further corroborated by the box plots of the DBA and RR values. Most importantly, our device is 100 times [61,62] less expensive than the PNT device.

Comparison With Prior Work

The literature suggests that continuous and precise lung volume and RR monitoring is difficult. Lung volume, a critical indication of COPD status and development, cannot be assessed by RR alone [34]. TV, ERV, IRV, and VC can indicate respiratory depression. Researchers have measured lung volume with a spirometer, but it is inaccurate, has a large latency, and cannot be used while the patient is asleep [35]. Our method improves continuous lung volume monitoring by not requiring the patient to be supine. It also measures lung volume without a costly sensor or imaging tests.

Strengths and Limitations

Our device is a low-cost medical device. It provides an opportunity to measure respiratory parameters in real time. The device uses a widely used and tested commercial humidity sensor that is stable and has a high signal-to-noise ratio. We tested the device with a variety of participants with different demographic characteristics and of different sexes to ensure that the results are reproducible. Despite all the strengths, there are certain limitations to our device. First, the framework of our device is not suitable for commercial applications as there is a large number of ad hoc components. There is scope to improve the design of the device by incorporating the sensors in the fabrication of the mask. Second, detailed study of the device with a larger number of participants and varying environmental conditions is required for further testing and calibration.

Future Work

Due to the restrictions inherent to variations in breathing patterns and device sensitivity, we foresee the development of our own array of humidity sensors, coupled with deep learning data processing, with the goal of resolving any problems associated with breathing pattern deviations. This will enable extrapolation of the unusable measurements to ensure the same level of precision as linear results. This array of humidity sensors printed on a flexible base would permit the incorporation into materials for surgical masks without affecting the usability. Limitations may include a high-humidity environment, which may increase the risk of saturating the humidity sensor. A combination of pressure and temperature measurements would limit this bias. Ideally, the sensor should be integrated into the mask so that it can be used in real-life conditions, without obstructing the ventilatory connection. Increasing the sample size will not validate the use of the sensor, but the inclusion of profiles with respiratory disorders would make it possible to obtain measurements with less obvious variations in amplitude and thus include situations in which respiratory disorders could be detected [12].

Conclusions

Wearable IoT medical technologies are developing as viable options not just to monitor patients at home after hospitalization but also to boost the affordability and accessibility of quality health care. In fact, the development of more effective and less expensive wearable medical devices could allow patients to monitor their health at home. In recent years, the number of

wearable medical devices for chronic disease monitoring has expanded. Last, improved medically certified sensors will facilitate the development of better and less expensive medical IoT devices. In our investigation, we utilized a single sensor to simultaneously detect pressure, temperature, and humidity. The literature indicates that temperature is frequently used to measure RR [45]. The results indicate that our RHS can measure RR within statistically acceptable control limits. The results of the 1-way ANOVA indicate that the group means of the PNT are equivalent to our RHS within the standard margin of error of the instrument. This is further supported by the Bland-Altman analysis, which revealed low values of 0.03 and 0.018 for the bias and RR, respectively. The data analysis revealed that the evolution of RRs over time for the PNT and our low-cost RHS follows a similar pattern. This is, as far as we are aware, the first study to investigate the use of RHS for reliably monitoring respiratory volumes on a medical IoT platform. We plan to examine the applicability of the RHS sensor to detect more complex respiratory disorders using deep learning in a future phase of development.

In terms of utility, it is evident that continuous and precise monitoring of lung capacity and RR represents a significant basic obstacle. Nonetheless, it presents a tremendous opportunity to monitor patients with OSA in intensive care or during surgery. Lung volumes, namely TV, can be utilized to manage respiratory pauses or identify the breathing pattern in patients with OSA who utilize C-PAP or A-PAP devices, which, when combined with artificial intelligence, can detect apnea occurrences and enhance A-PAP performance. Regarding ventilated patients, assessing TV will allow for better control of ventilatory weaning.

Acknowledgments

SGC thanks the Canada Research Chair and the Natural Sciences and Engineering Research Council (NSERC) Discovery programs for their support. We are also thankful to Dr. Ali Bahloul and Clothilde Brochot from Quebec's Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) for fruitful discussions.

Data Availability

Data will be provided upon reasonable request.

Authors' Contributions

FV takes responsibility for the integrity of the complete work. SGC made the final decision to submit the manuscript. FV participated in the study concept and design as well as the design for participant tests. JP acquired the data. FV, AB, and JP processed and analyzed the data. FV, AB, JP, JABG, and LFG drafted the manuscript. All authors interpreted the data and critically reviewed the data and manuscript.

Conflicts of Interest

None declared.

References

1. Halpin DMG, Criner GJ, Papi A, Singh D, Anzueto A, Martinez FJ, et al. Global Initiative for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease. The 2020 GOLD Science Committee report on COVID-19 and chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2021 Jan 01;203(1):24-36 [FREE Full text] [doi: 10.1164/rccm.202009-3533SO] [Medline: 33146552]
2. Wang F, Kream RM, Stefano GB. Long-term respiratory and neurological sequelae of COVID-19. Med Sci Monit 2020 Nov 01;26:e928996 [FREE Full text] [doi: 10.12659/MSM.928996] [Medline: 33177481]
3. Williams GW, Berg NK, Reskallah A, Yuan X, Eltzschig HK. Acute respiratory distress syndrome. Anesthesiology 2021 Feb 01;134(2):270-282 [FREE Full text] [doi: 10.1097/ALN.0000000000003571] [Medline: 33016981]
4. Lyons MM, Bhatt NY, Pack AI, Magalang UJ. Global burden of sleep-disordered breathing and its implications. Respirology 2020 Jul;25(7):690-702 [FREE Full text] [doi: 10.1111/resp.13838] [Medline: 32436658]
5. Fazleen A, Wilkinson T. Early COPD: current evidence for diagnosis and management. Ther Adv Respir Dis 2020;14:1753466620942128 [FREE Full text] [doi: 10.1177/1753466620942128] [Medline: 32664818]
6. Chuang M, Lin I. Investigating the relationships among lung function variables in chronic obstructive pulmonary disease in men. PeerJ 2019;7:e7829 [FREE Full text] [doi: 10.7717/peerj.7829] [Medline: 31592356]
7. Mannée DC, de Jongh F, van Helvoort H. Telemonitoring techniques for lung volume measurement: accuracy, artifacts and effort. Front Digit Health 2020 Sep 17;2:559483 [FREE Full text] [doi: 10.3389/fdgth.2020.559483] [Medline: 34713036]
8. Kwon OE, Shin KH, Dang TH, Ahn HJ, Rhie EH, Jang GY, et al. Tidal volume and stroke volume changes caused by respiratory events during sleep and their relationship with OSA severity: a pilot study. Sleep Breath 2021 Dec;25(4):2025-2038. [doi: 10.1007/s11325-021-02334-y] [Medline: 33683548]
9. Brochard L, Martin GS, Blanch L, Pelosi P, Belda FJ, Jubran A, et al. Clinical review: Respiratory monitoring in the ICU - a consensus of 16. Crit Care 2012 Dec 12;16(2):219 [FREE Full text] [doi: 10.1186/cc11146] [Medline: 22546221]

10. Johns DP, Walters JAE, Walters EH. Diagnosis and early detection of COPD using spirometry. J Thorac Dis 2014 Nov;6(11):1557-1569 [FREE Full text] [doi: 10.3978/j.issn.2072-1439.2014.08.18] [Medline: 25478197]
11. Rouatbi S, Mezghani S, Khalfallah M, Benzarti M. [Early detection of COPD in smoking cessation outpatients]. Tunis Med 2015 Jul;93(7):458-464 [FREE Full text] [Medline: 26757504]
12. Tipparaju VV, Xian X, Bridgeman D, Wang D, Tsow F, Forzani E, et al. Reliable breathing tracking with wearable mask device. IEEE Sens J 2020 May 15;20(10):5510-5518 [FREE Full text] [doi: 10.1109/jsen.2020.2969635] [Medline: 33746622]
13. 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;2:8 [FREE Full text] [doi: 10.1038/s41746-019-0083-3] [Medline: 31304358]
14. Gargiulo GD, Gunawardana U, O'Loughlin A, Sadozai M, Varaki ES, Breen PP. A wearable contactless sensor suitable for continuous simultaneous monitoring of respiration and cardiac activity. Journal of Sensors 2015;2015:1-6 [FREE Full text] [doi: 10.1155/2015/151859]
15. Monaco V, Stefanini C. Assessing the tidal volume through wearables: a scoping review. Sensors (Basel) 2021 Jun 16;21(12):1 [FREE Full text] [doi: 10.3390/s21124124] [Medline: 34208468]
16. Lutfi MF. The physiological basis and clinical significance of lung volume measurements. Multidiscip Respir Med 2017;12:3 [FREE Full text] [doi: 10.1186/s40248-017-0084-5] [Medline: 28194273]
17. Fekr AR, Radecka K, Zilic Z. Tidal volume variability and respiration rate estimation using a wearable accelerometer sensor. 2014 Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH); November 3-5, 2014; Athens, Greece. [doi: 10.4108/icst.mobihealth.2014.257528]
18. Scholz R, Bracio BR, Brutscheck M, Trommler P. Non-invasive respiratory rate detection in spontaneous respiration by humidity measurement. 2017 Presented at: 28th Irish Signals and Systems Conference (ISSC); June 20-21, 2017; Killarney, Ireland. [doi: 10.1109/issc.2017.7983620]
19. da Costa TD, Vara MFF, Cristino CS, Zanella TZ, Neto GNN, Nohama P. Breathing Monitoring and Pattern Recognition with Wearable Sensors. In: Nasiri N, editor. Wearable Devices. London, England: IntechOpen; 2019.
20. Kano S, Jarulertwathana N, Mohd-Noor S, Hyun JK, Asahara R, Mekaru H. Respiratory monitoring by ultrafast humidity sensors with nanomaterials: a review. Sensors (Basel) 2022 Feb 07;22(3):1 [FREE Full text] [doi: 10.3390/s22031251] [Medline: 35161997]
21. Rivero-Yeverino D. [Spirometry: basic concepts]. Rev Alerg Mex 2019;66(1):76-84. [doi: 10.29262/ram.v66i1.536] [Medline: 31013409]
22. Liou TG, Kanner RE. Spirometry. Clin Rev Allergy Immunol 2009 Dec;37(3):137-152. [doi: 10.1007/s12016-009-8128-z] [Medline: 19347610]
23. Nicolò A, Massaroni C, Schena E, Sacchetti M. The importance of respiratory rate monitoring: from healthcare to sport and exercise. Sensors (Basel) 2020 Nov 09;20(21):1 [FREE Full text] [doi: 10.3390/s20216396] [Medline: 33182463]
24. Rolfe S. The importance of respiratory rate monitoring. Br J Nurs 2019 Apr 25;28(8):504-508. [doi: 10.12968/bjon.2019.28.8.504] [Medline: 31002547]
25. Hodgson CL, Stiller K, Needham DM, Tipping CJ, Harrold M, Baldwin CE, et al. Expert consensus and recommendations on safety criteria for active mobilization of mechanically ventilated critically ill adults. Crit Care 2014 Dec 04;18(6):658 [FREE Full text] [doi: 10.1186/s13054-014-0658-y] [Medline: 25475522]
26. Tamura T. Current progress of photoplethysmography and SPO for health monitoring. Biomed Eng Lett 2019 Feb;9(1):21-36 [FREE Full text] [doi: 10.1007/s13534-019-00097-w] [Medline: 30956878]
27. Lu G, Yang F, Taylor JA, Stein JF. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. J Med Eng Technol 2009;33(8):634-641. [doi: 10.3109/03091900903150998] [Medline: 19848857]
28. Alian AA, Shelley KH. Photoplethysmography. Best Pract Res Clin Anaesthesiol 2014 Dec;28(4):395-406. [doi: 10.1016/j.bpa.2014.08.006] [Medline: 25480769]
29. Birrenkott DA, Pimentel MAF, Watkinson PJ, Clifton DA. Robust estimation of respiratory rate via ECG- and PPG-derived respiratory quality indices. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:676-679. [doi: 10.1109/EMBC.2016.7590792] [Medline: 28268418]
30. Felblinger J, Boesch C. Amplitude demodulation of the electrocardiogram signal (ECG) for respiration monitoring and compensation during MR examinations. Magn Reson Med 1997 Jul;38(1):129-136. [doi: 10.1002/mrm.1910380118] [Medline: 9211388]
31. Lei R, Ling BW, Feng P, Chen J. Estimation of heart rate and respiratory rate from PPG signal using complementary ensemble empirical mode Decomposition with both independent component analysis and non-negative matrix factorization. Sensors (Basel) 2020 Jun 06;20(11):1 [FREE Full text] [doi: 10.3390/s20113238] [Medline: 32517226]
32. Giardino ND, Lehrer PM, Edelberg R. Comparison of finger plethysmograph to ECG in the measurement of heart rate variability. Psychophysiology 2002 Mar;39(2):246-253. [doi: 10.1111/1469-8986.3920246]
33. Hernando A, Peláez-Coca MD, Lozano MT, Lázaro J, Gil E. Finger and forehead PPG signal comparison for respiratory rate estimation. Physiol Meas 2019 Sep 30;40(9):095007. [doi: 10.1088/1361-6579/ab3be0] [Medline: 31422948]

34. Braun SR. Respiratory Rate and Pattern. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations, 3rd edition. Boston, MA: Butterworths; 1990.
35. Rebuck DA, Hanania NA, D'Urzo AD, Chapman KR. The accuracy of a handheld portable spirometer. Chest 1996 Jan;109(1):152-157. [doi: 10.1378/chest.109.1.152] [Medline: 8549178]
36. Garfield JL, Marchetti N, Gaughan JP, Steiner RM, Criner GJ. Total lung capacity by plethysmography and high-resolution computed tomography in COPD. Int J Chron Obstruct Pulmon Dis 2012;7:119-126 [FREE Full text] [doi: 10.2147/COPD.S26419] [Medline: 22399851]
37. Escobedo P, Fernández-Ramos MD, López-Ruiz N, Moyano-Rodríguez O, Martínez-Olmos A, Pérez de Vargas-Sansalvador IM, et al. Smart facemask for wireless CO monitoring. Nat Commun 2022 Jan 10;13(1):72 [FREE Full text] [doi: 10.1038/s41467-021-27733-3] [Medline: 35013232]
38. Delgado BJ, Bajaj T. Physiology, Lung Capacity. Treasure Island, FL: StatPearls Publishing; 2023.
39. Tantucci C, Bottone D, Borghesi A, Guerini M, Quadri F, Pini L. Methods for measuring lung volumes: is there a better one? Respiration 2016;91(4):273-280 [FREE Full text] [doi: 10.1159/000444418] [Medline: 26982496]
40. Kelly JT, Campbell KL, Gong E, Scuffham P. The Internet of Things: impact and implications for health care delivery. J Med Internet Res 2020 Nov 10;22(11):e20135 [FREE Full text] [doi: 10.2196/20135] [Medline: 33170132]
41. Subahi AF. Edge-based IoT medical record system: requirements, recommendations and conceptual design. IEEE Access 2019;7:94150-94159. [doi: 10.1109/access.2019.2927958]
42. Hung PD. Estimating respiration rate using an accelerometer sensor. CSBio '17: Proceedings of the 8th International Conference on Computational Systems-Biology and Bioinformatics 2017:11-14. [doi: 10.1145/3156346.3156349]
43. Lee J, Yoo SK. Respiration rate estimation based on independent component analysis of accelerometer data: pilot single-arm intervention study. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e17803 [FREE Full text] [doi: 10.2196/17803] [Medline: 32773384]
44. Leube J, Zschocke J, Kluge M, Pelikan L, Graf A, Glos M, et al. Reconstruction of the respiratory signal through ECG and wrist accelerometer data. Sci Rep 2020 Sep 03;10(1):14530 [FREE Full text] [doi: 10.1038/s41598-020-71539-0] [Medline: 32884062]
45. Basra A, Mukhopadhayay B, Kar S. Temperature sensor based ultra low cost respiration monitoring system. 2017 Presented at: 9th International Conference on Communication Systems and Networks (COMSNETS); January 4-8, 2017; Bengaluru, India. [doi: 10.1109/comsnets.2017.7945448]
46. Raji A, Devi PK, Jeyaseeli PG, Balaganesh N. Respiratory monitoring system for asthma patients based on IoT. 2017 Presented at: Online International Conference on Green Engineering and Technologies (IC-GET); November 19, 2016; Coimbatore, India. [doi: 10.1109/get.2016.7916737]
47. Shevchenko GV, Glubokov NA, Yupashevsky AV, Kazmina AS. Air Flow Sensor Based on Environmental Sensor BME280. 2020 Presented at: 21st International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM); June 29-July 3, 2020; Chemal, Russia. [doi: 10.1109/edm49804.2020.9153474]
48. Mansour E, Vishinkin R, Rihet S, Saliba W, Fish F, Sarfati P, et al. Measurement of temperature and relative humidity in exhaled breath. Sensors and Actuators B: Chemical 2020 Feb;304:127371. [doi: 10.1016/j.snb.2019.127371]
49. Paul A, Chandran DV, Ramesh R. A Novel IoT-Based Solution for Respiratory Flow Diagnosis. In: Nandan Mohanty S, Chatterjee JM, Satpathy S, editors. Internet of Things and Its Applications. EAI/Springer Innovations in Communication and Computing. Cham, Switzerland: Springer International Publishing; 2022:101-115.
50. Hill B, Stapley R, Nesar MSB, Whitaker BM. Touchless Respiratory Monitor Preliminary Data and Results. 2021 Presented at: IEEE Aerospace Conference (50100); March 6-13, 2021; Big Sky, MT. [doi: 10.1109/aero50100.2021.9438154]
51. Ashutosh K, Gilbert R, Auchincloss JH, Erlebacher J, Peppi D. Impedance pneumograph and magnetometer methods for monitoring tidal volume. J Appl Physiol 1974 Dec;37(6):964-966. [doi: 10.1152/jappl.1974.37.6.964] [Medline: 4436235]
52. Mannée D, de Jongh F, van Helvoort H. The accuracy of tidal volume measured with a smart shirt during tasks of daily living in healthy subjects: cross-sectional study. JMIR Form Res 2021 Oct 18;5(10):e30916 [FREE Full text] [doi: 10.2196/30916] [Medline: 34661546]
53. Liu H, Allen J, Zheng D, Chen F. Recent development of respiratory rate measurement technologies. Physiol Meas 2019 Aug 02;40(7):07TR01. [doi: 10.1088/1361-6579/ab299e] [Medline: 31195383]
54. Miyoshi E, Fujino Y, Uchiyama A, Mashimo T, Nishimura M. Effects of gas leak on triggering function, humidification, and inspiratory oxygen fraction during noninvasive positive airway pressure ventilation. Chest 2005 Nov;128(5):3691-3698. [doi: 10.1378/chest.128.5.3691] [Medline: 16304335]
55. Xiang G, Zhu X, Ma L, Huang H, Wu X, Zhang W, et al. Clinical guidelines on the application of Internet of Things (IOT) medical technology in the rehabilitation of chronic obstructive pulmonary disease. J Thorac Dis 2021 Aug;13(8):4629-4637 [FREE Full text] [doi: 10.21037/jtd-21-670] [Medline: 34527304]
56. Erdenechimeg D, Enkhzul D, Munkhtamir O, Enkhbat B. Wireless monitoring method for diabetic foot temperature. 2017 Presented at: 19th International Conference on Advanced Communication Technology (ICACT); February 19-22, 2017; PyeongChang, South Korea. [doi: 10.23919/icact.2017.7890053]

57. Pratama NB, Nasution SM, Nugrahaeni RA. Clip On Wearable Device Design For Body Condition Monitoring. 2019 Presented at: International Conference on Information Technology Systems and Innovation (ICITSI); October 22-26, 2018; Bandung, Indonesia. [doi: 10.1109/icitsi.2018.8695965]
58. Humidity sensor BME280. Bosch Sensortec. URL: https://www.bosch-sensortec.com/products/environmental-sensors/humidity-sensors-bme280/ [accessed 2022-12-15]
59. Faraone A, Delgado-Gonzalo R. Convolutional-Recurrent Neural Networks on Low-Power Wearable Platforms for Cardiac Arrhythmia Detection. 2020 Presented at: 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS); August 31-September 2, 2020; Genova, Italy. [doi: 10.1109/aicas48895.2020.9073950]
60. Kupriianova O, Lauter C, Muller JM. Radix conversion for IEEE754-2008 mixed radix floating-point arithmetic. 2014 Presented at: Asilomar Conference on Signals, Systems and Computers; November 3-6, 2013; Pacific Grove, CA. [doi: 10.1109/acssc.2013.6810471]
61. Hans Rudolph. URL: https://www.rudolphkc.com/product-page/linear-pneumotachs-heater-controllers [accessed 2022-12-15]
62. PA-1 Pneumotach Amplifiers. Hans Rudolph. URL: https://www.rudolphkc.com/product-page/pa-1-pnemotach-amplifiers [accessed 2022-12-15]
63. Xu XK, Harvey BP, Lutchen KR, Gelbman BD, Monfre SL, Coifman RE, et al. Comparison of a micro-electro-mechanical system airflow sensor with the pneumotach in the forced oscillation technique. Med Devices (Auckl) 2018;11:419-426 [FREE Full text] [doi: 10.2147/MDER.S181258] [Medline: 30588132]
64. McHugh ML. Multiple comparison analysis testing in ANOVA. Biochem Med (Zagreb) 2011;21(3):203-209 [FREE Full text] [doi: 10.11613/bm.2011.029] [Medline: 22420233]
65. Honal M, Leupold J, Huff S, Baumann T, Ludwig U. Compensation of breathing motion artifacts for MRI with continuously moving table. Magn Reson Med 2010 Mar;63(3):701-712 [FREE Full text] [doi: 10.1002/mrm.22162] [Medline: 20187180]
66. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
67. Wilcox RR. Introduction to Robust Estimation and Hypothesis Testing. Cambridge, MA: Academic Press; 2011.
68. Rumiński J. Analysis of the parameters of respiration patterns extracted from thermal image sequences. Biocybern Biomed Eng 2016;36(4):731-741 [FREE Full text] [doi: 10.1016/j.bbe.2016.07.006] [Medline: 32287710]

Abbreviations

A-PAP: automatic positive airway pressure
COPD: chronic obstructive pulmonary disease
C-PAP: continuous positive airway pressure
DBA: deep breathing area
ECG: electrocardiography
ERV: expiratory reserve volume
ICU: intensive care unit
IoT: Internet of Things
IRV: inspiratory reserve volume
LoA: limit of agreement
OSA: obstructive sleep apnea
PNT: pneumotach
PPG: photoplethysmography
RH: relative humidity
RHS: relative humidity sensor
RMSE: root mean square error
RR: respiratory rate
TV: tidal volume
VC: vital capacity

Edited by T Leung; submitted 10.03.23; peer-reviewed by MS Arefin, O Sverdlov; comments to author 31.07.23; revised version received 22.08.23; accepted 07.09.23; published 25.10.23.

Please cite as:
Vaussenat F, Bhattacharya A, Payette J, Benavides-Guerrero JA, Perrotton A, Gerlein LF, Cloutier SG
Continuous Critical Respiratory Parameter Measurements Using a Single Low-Cost Relative Humidity Sensor: Evaluation Study
JMIR Biomed Eng 2023;8:e47146
URL: https://biomedeng.jmir.org/2023/1/e47146
doi:10.2196/47146
PMID:38875670

Original Paper

Severity Classification Using Dynamic Time Warping–Based Voice Biomarkers for Patients With COVID-19: Feasibility Cross-Sectional Study

Teruhisa Watase[1], BSc, MPH; Yasuhiro Omiya[2], ME, PhD; Shinichi Tokuno[1], MD, PhD

[1]Gradutate School of Health Innovation, Kanagawa University of Human Service, Kawasaki, Kanagawa, Japan

[2]Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Teruhisa Watase, BSc, MPH
Gradutate School of Health Innovation
Kanagawa University of Human Service
RGB 2A, 3-25-10 Tonomachi
Kawasaki-ku
Kawasaki, Kanagawa, 2100821
Japan
Phone: 81 44 589 8100
Fax: 81 44 589 8188
Email: terrywatase@gmail.com

Abstract

Background: In Japan, individuals with mild COVID-19 illness previously required to be monitored in designated areas and were hospitalized only if their condition worsened to moderate illness or worse. Daily monitoring using a pulse oximeter was a crucial indicator for hospitalization. However, a drastic increase in the number of patients resulted in a shortage of pulse oximeters for monitoring. Therefore, an alternative and cost-effective method for monitoring patients with mild illness was required. Previous studies have shown that voice biomarkers for Parkinson disease or Alzheimer disease are useful for classifying or monitoring symptoms; thus, we tried to adapt voice biomarkers for classifying the severity of COVID-19 using a dynamic time warping (DTW) algorithm where voice wavelets can be treated as 2D features; the differences between wavelet features are calculated as scores.

Objective: This feasibility study aimed to test whether DTW-based indices can generate voice biomarkers for a binary classification model using COVID-19 patients' voices to distinguish moderate illness from mild illness at a significant level.

Methods: We conducted a cross-sectional study using voice samples of COVID-19 patients. Three kinds of long vowels were processed into 10-cycle waveforms with standardized power and time axes. The DTW-based indices were generated by all pairs of waveforms and tested with the Mann-Whitney *U* test (α<.01) and verified with a linear discrimination analysis and confusion matrix to determine which indices were better for binary classification of disease severity. A binary classification model was generated based on a generalized linear model (GLM) using the most promising indices as predictors. The receiver operating characteristic curve/area under the curve (ROC/AUC) validated the model performance, and the confusion matrix calculated the model accuracy.

Results: Participants in this study (n=295) were infected with COVID-19 between June 2021 and March 2022, were aged 20 years or older, and recuperated in Kanagawa prefecture. Voice samples (n=110) were selected from the participants' attribution matrix based on age group, sex, time of infection, and whether they had mild illness (n=61) or moderate illness (n=49). The DTW-based variance indices were found to be significant (*P*<.001, except for 1 of 6 indices), with a balanced accuracy in the range between 79% and 88.6% for the /a/, /e/, and /u/ vowel sounds. The GLM achieved a high balance accuracy of 86.3% (for /a/), 80.2% (for /e/), and 88% (for /u/) and ROC/AUC of 94.8% (95% CI 90.6%-94.8%) for /a/, 86.5% (95% CI 79.8%-86.5%) for /e/, and 95.6% (95% CI 92.1%-95.6%) for /u/.

Conclusions: The proposed model can be a voice biomarker for an alternative and cost-effective method of monitoring the progress of COVID-19 patients in care.

(JMIR Biomed Eng 2023;8:e50924) doi:10.2196/50924

KEYWORDS

voice biomarker; dynamic time warping; COVID-19; smartphone; severity classification; biomarker; feasibility study; illness; monitoring; respiratory disease; accuracy; logistic model; tool; model

Introduction

Background

COVID-19 originated in Wuhan, China, in December 2019 and turned into a worldwide pandemic. As of December 2022, the number of people infected with this disease reached approximately 650 million, of whom more than 6.64 million had lost their lives. Although the number of new infections appeared to have abated in the spring of 2023, the past explosion of infections strained medical care systems in several countries. To cope with this pressure, these countries have changed their responses toward infected patients based on the severity of their illness. In Japan, as defined in Table 1, which shows the Ministry of Health, Labour and Welfare guidelines on the severity of COVID-19 [1], responses were divided into 4 categories of severity, ranging from mild to serious illness. Mild illness is defined as "an oxygen saturation of 96% or more, or a clinical condition of no respiratory symptoms or coughing without shortness of breath (SoB) but no evidence of pneumonia in either case," and moderate illness I is defined as "an oxygen saturation of greater than 93% but less than 96%, or a clinical condition of shortness of breath or pneumonia." Moderate illness II is defined as "oxygen saturation of 93% or less, or oxygen administration is required." The target population for this study was individuals who were recovering at home or in recuperation facilities. Therefore, they were theoretically patients with mild illness. Still, due to worsening conditions or shortcomings of medical services, this population included patients with moderate illness I who should have been treated in a hospital. Therefore, accurately classifying these 2 adjacent severity categories (mild illness and moderate illness I) is essential in determining appropriate measures, such as early hospitalization, by detecting worsening conditions in patients with mild illness.

Table 1. Definitions of the severity of COVID-19 infections.

Severity	Oxygen saturation, %	Clinical condition
Mild illness	≥96	Absence of respiratory symptoms or presence of coughing without shortness of breath, but no evidence of pneumonia in either case
Moderate illness I	93-96	Shortness of breath and pneumonia are evident
Moderate illness II	≤93	Oxygen administration is required
Serious illness	N/A	Admission to intensive care unit or requirement of a ventilator

Oxygen saturation (SpO_2) measurements using a pulse oximeter were crucial for assessing the severity of illness. Daily measurements of SpO_2 and body temperature, along with the assessment of physical conditions, were essential for monitoring disease progression from mild illness to moderate illness I over approximately 1 week or more during the recuperation period. However, the explosive increase in the number of COVID-19 patients made it difficult to distribute pulse oximeters to all patients with mild illness, especially young patients who were forced to recuperate at their homes rather than in health care facilities. This unexpected shortage of pulse oximeters has motivated us to devise alternative and cost-effective ways to monitor for worsening medical condition in persons exhibiting mild illness.

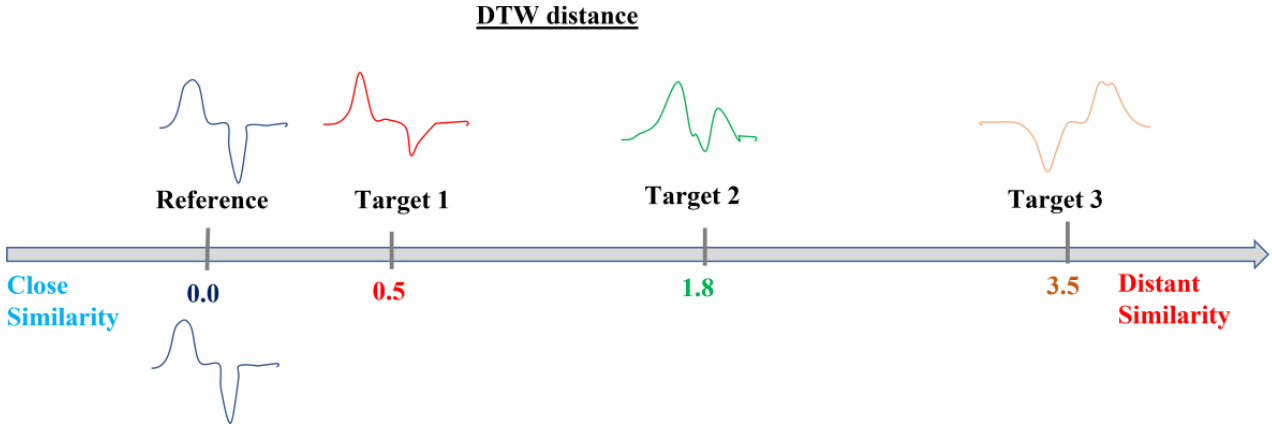
Voice Biomarkers

Previous research on Parkinson disease, Alzheimer disease, depression, and other psychiatric disorders such as stress [2-7] has shown that voice biomarkers can be leveraged to noninvasively and cost-effectively identify the presence or absence of diseases, classify symptoms, and monitor conditions. Voice biomarkers could also be an alternative method to detect changes in disease severity from mild illness to moderate illness I in COVID-19, which is a respiratory disease and has been reported to cause acoustic changes in the voice due to inflammation of the pharynx in the vocal tract, vocal cords, or both, as well as a lower expiratory volume due to pneumonia [8]. Moreover, significant differences in jitter (fluctuation of the fundamental frequency on the time axis), shimmer (fluctuation of the amplitude on the power axis), and harmonic-to-noise ratio (HNR) were reported between healthy subjects and those with COVID-19 [8-11]. There are also reports that COVID-19 can be detected from acoustic data obtained from a patient's cough [12].

Dynamic Time Warping

Dynamic time warping (DTW) is an effective algorithm for measuring the similarity between 2 patterns. The DTW distance, which is a computational result obtained by the DTW algorithm using 2 waveform features, progressively approaches zero as the features become more similar, whereas it increases as the features become less similar (Figure 1).

Figure 1. Illustration of the DTW distance, with 3 target example waves compared to a reference wave. DTW: dynamic time warping.



This metric has been widely used since the 1980s in fields such as motion recognition, speech recognition, and time-series data analysis [13-15]. For example, it was reported that DTW could differentiate between healthy people and people with walking disabilities with high accuracy by processing differences in the gait patterns acquired by accelerometer sensors on smartphones [13]. The effectiveness of an automated scoring system applied in conjunction with the DTW algorithm for evaluating the progress of speech audiometric rehabilitation was reported to be similar to that of conventional manual scoring methods [14]. It has also been reported that complementing the mel-frequency cepstral coefficient (MFCC) algorithm with the DTW algorithm improved voice recognition performance. The DTW algorithm has been introduced as a feature-matching technique for voice recognition [15]. The feature-matching performance of the DTW algorithm (ie, the scoring method for 2D feature similarity) may function effectively for the desired classification of long vowel samples.

Goal of This Study

This feasibility study aimed to test whether DTW-based voice biomarkers can be used to achieve a binary classification of mild illness and moderate illness I for COVID-19 at a significant level.

Methods

Study Design

We conducted a cross-sectional study using the voice samples of COVID-19 patients.

Participants

This study recruited participants through a brochure that was distributed exclusively to COVID-19 patients who were aged 20 years and older, were positive for SARS-CoV-2 in PCR testing, and recuperated at designated facilities or at home in Kanagawa prefecture, Japan, between June 2021 and March 2022. Patients who consented to the study's objectives were requested to register for participation using the QR code on the brochure through their smartphones. The participants were asked to provide their daily vital signs data, including voice recordings during the recuperation period, but were also given the option to withdraw from the study (opt out) at any time of their own accord. A ¥1000 (US $6.68) Amazon gift card was given to participants as compensation. Because this study only included patients with mild illness or moderate illness I, who did not require hospitalization, patients with moderate illness II were not included. The participants were divided into 2 groups, the mild group and moderate I group, according to the definitions given in Table 1.

Data Collection

Those who agreed to participate in the study were asked to install a smartphone app and enter their basic information, vital signs data (temperature and SpO_2), symptom scores, and voice recordings on the first day of recuperation. From the second day onward, the participants were required to enter their vital signs data, symptom scores, and voice recordings daily until the last day of recuperation. Voice data were stored together with text data indicating the symptoms and vital signs on a dedicated server with high security. The voice recordings were in the WAV format with a sampling rate of 48 kHz and a bit depth of 16 bits using the 3 long vowels /a/, /e/, and /u/. Participants were asked to explain their reasoning if they wished to withdraw from the study. Table 2 shows the timing and data entry items of the participants.

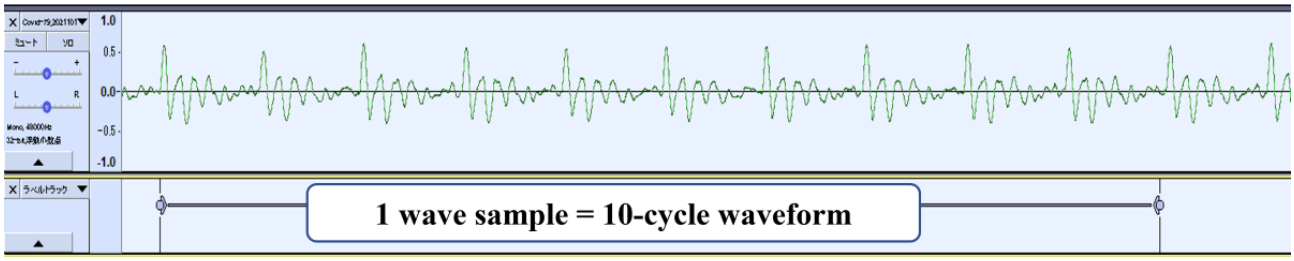
Table 2. Timing and items for the input data and information from participants.

Timing and items	Description
Baseline	
Orientation	• The researchers explained the purpose of the study and obtained participant consent
Basic information	• Sex • Age • Symptom onset date • Diagnosis confirmation date • Treatment start date
On a daily basis during recuperation	
Vital sign data	• Body temperature • Blood oxygen saturation
Questionnaire	• Change in symptoms • Symptomatic or not • Respiratory distress • Taste or olfactory disorders • Cough or sputum • Chest pain • Runny nose or nasal congestion • Sore throat • Nausea or vomiting • Diarrhea • Appetite • Fatigue • Headache • Joint pain • Rash • Red eyes
Voice recording	• Three long vowels: /a/, /e/, /u/
Dropout during recuperation	
Dropout	• Confirm the reason for dropping out from the research

Waveform Sample Cutout and Standardization

To calculate the DTW distance, a 10-cycle waveform sample was extracted for each date from the participants' long-vowel recordings in the WAV format using Audacity (version 3.1.3; Audacity Team) at a sampling rate of 48 kHz (Figure 2). Then, standardization was achieved along the power axis within the range of –1 to +1 as the maximum amplitude, and the time axis involved 1000 data points that were multiplied by 1/48,000 seconds, considering the length of a 10-cycle waveform. To read and standardize the WAV data, R (version 4.4.2; R Core Team) with the *tuneR* package (version 1.4.0) was used.

Figure 2. Screenshot showing 10-cycle waveform data extracted for each date from each patient's voice recording of vowels.



Calculation of the DTW Distance for 2 Groups

After standardizing the power and time of the 110 waveform samples, the DTW distance was calculated for each sample paired with those of the remaining 109 samples. The DTW distances that were obtained were divided into 2 categories based on 2 kinds of labels for the 109 waveform samples. Therefore, each sample was assigned 2 variables for DTW distance. In the mild group, these were 61 or 60 DTW distances, and in the moderate I group, these were 48 or 49 DTW distances. The average and variance of the DTW distances were calculated for each group. For the mild group, the average index (ie, the mild-group filtering [MiF] average) and variance index (ie, MiF variance) of the DTW distance were obtained, whereas for the

moderate I group, the average index (ie, moderate-group filtering [MoF] average) and variance index (MoF variance) of the DTW distance were obtained. These 4 indices were obtained from a single waveform sample. The indices for the 3 vowels, /a/, /e/, and /u/, were prepared and are represented as shown in Table 3. Thus, 12 indices were used in the subsequent analyses. The average values and variances of the DTW distances for the 2 groups were statistically investigated to determine whether they exhibited significant values for the 2-group classification scheme.

Table 3. Twelve indices generated from 3 vowels and 4 indices.

Indices	Vowels		
	/a/	/e/	/u/
MiF[a] average	/a/-MiF average	/e/-MiF average	/u/-MiF average
MoF[b] average	/a/-MoF average	/e/-MoF average	/u/-MoF average
MiF variance	/a/-MiF variance	/e/-MiF variance	/u/-MiF variance
MoF variance	/a/-MoF variance	/e/-MoF variance	/u/-MoF variance

[a]MiF: mild-group filtering.

[b]MoF: moderate-group filtering.

Data Analysis

Linear Discriminant Analysis Considering the Average and Variance Indices of the DTW Distance

The Mann-Whitney *U* test was used to determine whether there was any statistical significance between the mild and moderate I groups. This test was performed on 12 indices that measured the average and variance of the DTW distance for the 3 vowels /a/, /e/, and /u/. A significance level of 1% was established, with the null hypothesis of no statistical significance between the 2 groups. Box plots and linear discriminant analysis (LDA) were used to determine the indicators of the 3 vowels most effective for determining statistical significance between the 2 groups. The confusion matrix obtained from the LDA results was displayed with a specific index for the true positive rate (TPR), true negative rate (TNR), and balanced accuracy (BA). The boxplot function was calculated and plotted using the R *ggplot* package (version 3.4.0), and the LDA function was calculated using the R *MASS* package (version 7.3).

Generalized Linear Model With the DTW distance

The significant indices from the 4 categories of DTW distance were used to distinguish between severity levels (mild or moderate I). These indices were then used as explanatory variables to create generalized linear models (GLMs) for each vowel. A 5-fold cross-validation method with 110 waveform samples was used to train the model for each vowel, which was then used to predict the severity classification. R was used for GLM modeling and label prediction. The *pROC* package (version 1.18.0) for R was used to obtain the receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC), whereas confusion matrices were generated using the *Caret* package (version 6.0) for R.

Ethics Approval

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Kanagawa University of Human Services (SHI 3-001, dated May 27, 2021, and SHI 26, dated November 25, 2021). Informed consent was obtained from all participants involved in the study.

Results

Participants

In June 2021 and March 2022, approximately 540,000 COVID-19 patients were recorded in Kanagawa prefecture. After requesting approximately 10,000 people to participate in the study, our study recruited 295 participants in the same period, of whom 291 were eligible to participate because participants who did not meet the inclusion criteria, such as minors, those with invalid data registration, or those who withdrew midway through evaluation, were excluded.

Seventy-four participants who reported no symptoms of coughing, throat pain, chest pain, or SoB during recuperation were assigned to the mild group. Of the 217 participants who reported any symptoms, 68 of them with symptoms of SoB were assigned to the moderate I group. Of the 149 participants who reported symptoms other than SoB, 6 of them with SpO_2 values less than 96% were assigned to the moderate I group, and 143 participants who reported SpO_2 values of no less than 96% were assigned to the mild group. The 291 participants were classified into 2 groups: 217 as mild and 74 as moderate I. Figure 3 shows a flowchart of study participation.

The primary periods of infection in Japan were during the Delta period, from July to December 2021, and the Omicron period, from January to June 2022. According to previous reports [16,17], COVID-19 exhibits varying levels of infectivity, severity, and symptoms, depending on the type of mutant strain present. We identified the time of infection in Japan and carefully matched the 291 study participants who had already been labeled into 2 groups. Table 4 shows the attribution matrix for the participants by the time of infection, sex, and severity. Table 5 shows the attribution matrix for the same sample based on the time of infection, sex, and age group. Finally, 110 participants (61 with mild illness and 49 with moderate illness I) were included in the study.

Figure 3. Flowchart showing classification of 291 participants into the mild illness and moderate illness I groups. “Four symptoms” refers to the 4 major symptoms: coughing, throat pain, chest pain, and shortness of breath. SpO2: oxygen saturation.

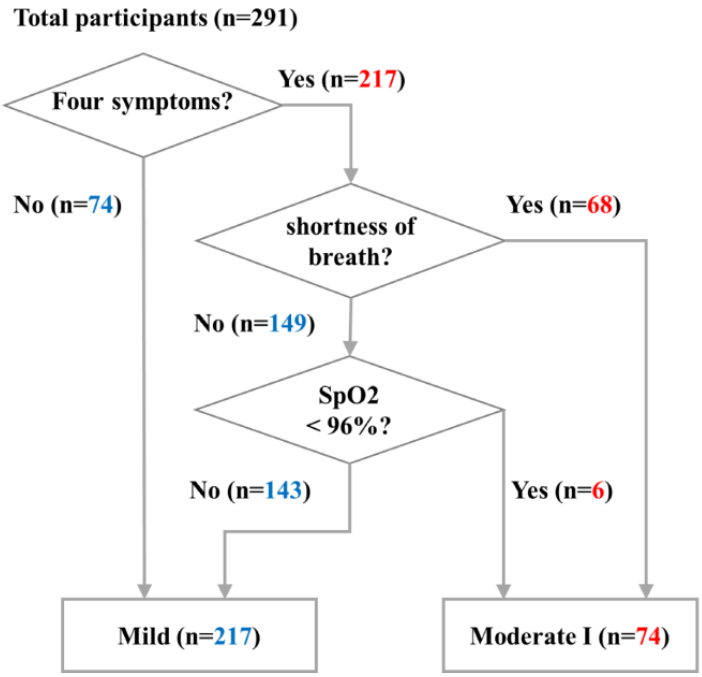


Table 4. Participants’ attribution matrix by the time of infection, sex, and severity (n=110).

Severity	Delta period[a] (n=46), participants, n	Omicron period[b] (n=64), participants, n
Mild illness (n= 61)		
Male	14	16
Female	13	18
Total	27	34
Moderate illness I (n=49)		
Male	10	16
Female	9	14
Total	19	30

[a]Delta period: July 2021 to December 2021.

[b]Omicron period: January 2022 to June 2022.

Table 5. Participants' attribution matrix by the time of infection, sex, and age group (n=110).

Age group (years)	Delta period[a] (n=46), participants, n	Omicron period[b] (n=64), participants, n
20-29		
Male	9	10
Female	16	15
Total	25	25
30-39		
Male	4	9
Female	1	9
Total	5	18
40-49		
Male	7	6
Female	2	5
Total	9	11
50-59		
Male	3	4
Female	3	3
Total	6	7
60-69		
Male	0	2
Female	0	0
Total	0	2
≥70		
Male	1	1
Female	0	0
Total	1	1

[a]Delta period: July 2021 to December 2021.

[b]Omicron period: January 2022 to June 2022.

Linear Discriminant Analysis Considering the Average and Variance Indices of the DTW Distance

Distribution of the Average and Variance Indices of the DTW Distance

Table 6 displays the Mann-Whitney *U* test results for the 2 groups based on 3 vowels and 4 indicators. Of the 12 indices, 6 were found to be significant; they included /u/-MiF average, /a/-MiF variance, /e/-MiF variance, /a/-MoF variance, /e/-MoF variance, and /u/-MoF variance. The only index that was significant among the average indices was /u/-MiF average, while /u/-MiF variance was the only insignificant index among the variance indices. This indicates that the variance indices were more significant overall.

Figures 4 and 5 illustrate the distributions of MiF average and MoF average, as well as MiF variance and MoF variance, for the 2 groups.

Table 6. Results of the Mann-Whitney U test for mild illness and moderate illness I group classification.

Indices	Vowels, *P* values		
	/a/	/e/	/u/
MiF[a] average	.03	.71	<.001
MoF[b] average	.60	.43	.03
MiF variance	<.001	<.001	.45
MoF variance	<.001	<.001	<.001

[a]MiF: mild-group filtering.

[b]MoF: moderate-group filtering.

Figure 4. Distribution in the 2 groups of the average index for the dynamic time warping distance (MiF average and MoF average). MiF: mild-group filtering; MoF: moderate-group filtering.

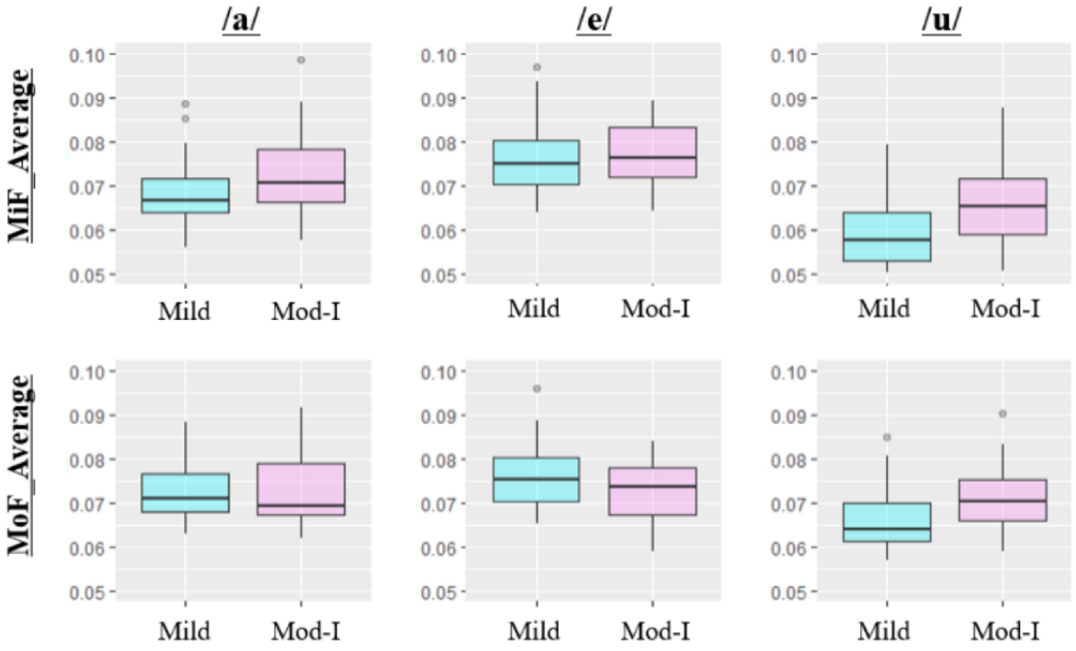
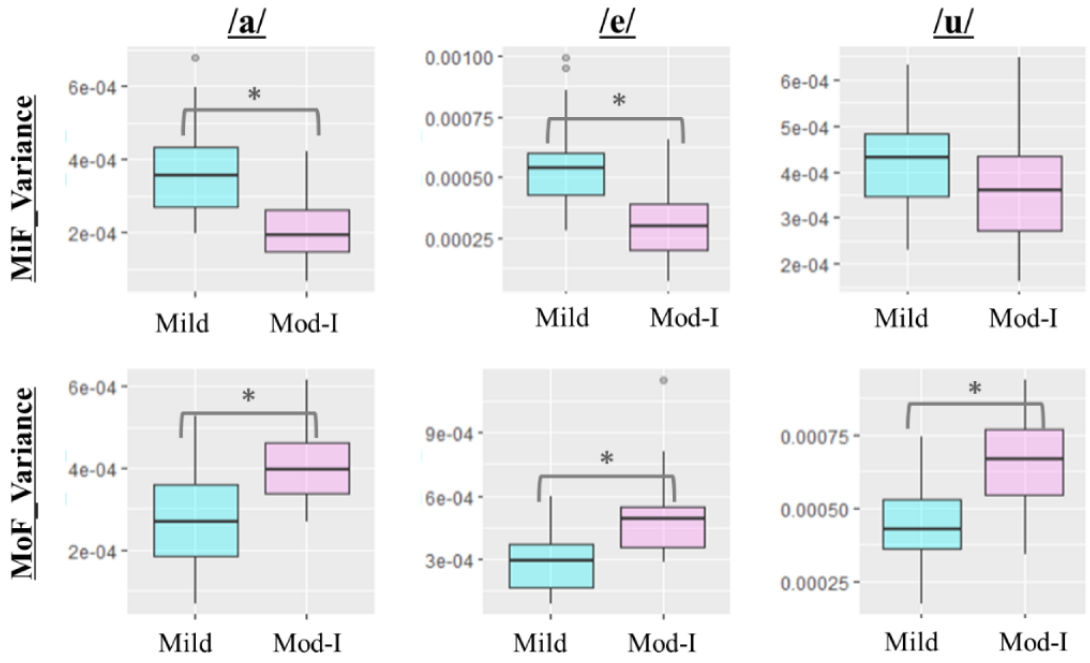


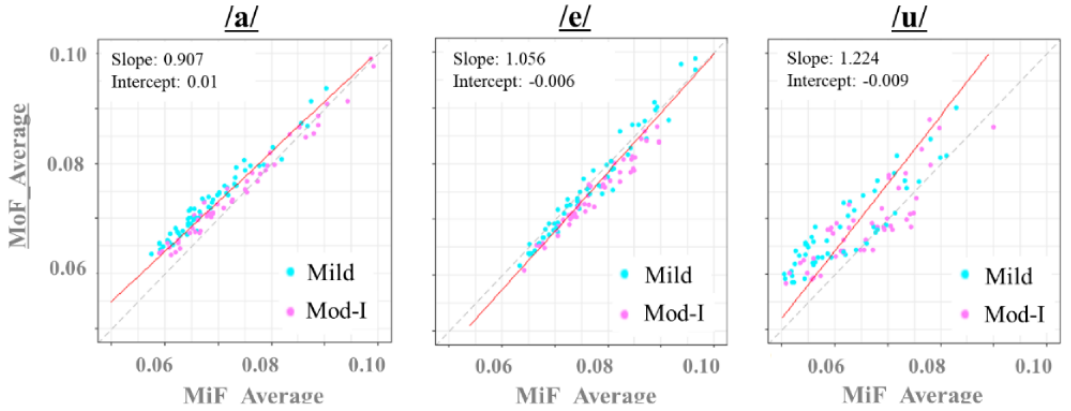
Figure 5. Distribution in the 2 groups of the variance indices for the dynamic time warping distance (MiF variance and MoF variance) *P<.001. MiF: mild-group filtering; MoF: moderate-group filtering.



Results for LDA

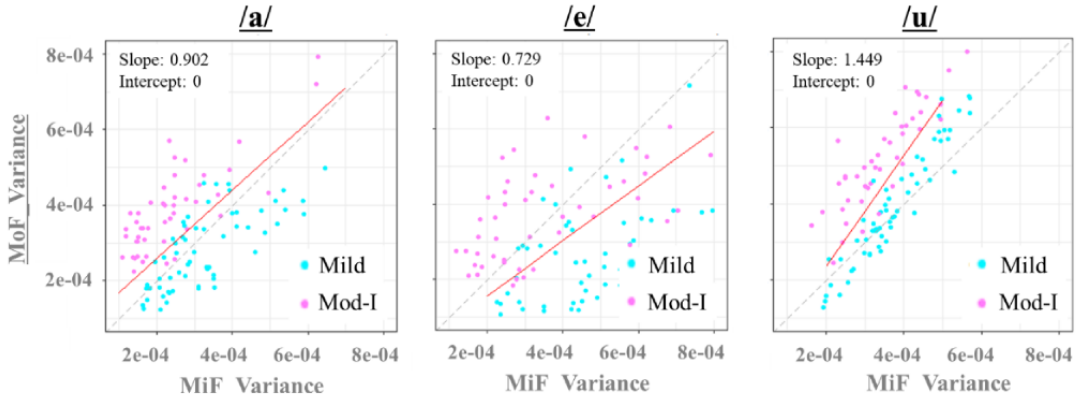
Figure 6 shows a scatter plot of the average indices, and Figure 7 shows a scatter plot of the variance indices of the DTW distance, along with the confusion matrix, TPR, TNR, and BA values of the LDA. The straight line represents the discriminant line obtained using LDA. The variance indices of the DTW distance provided overall superior results for the classification indicators, including TPR, TNR, and BA, of the confusion matrix compared to the average indices. The ease of classification can be visually verified by observing the plots achieved via LDA.

Figure 6. Linear discriminant analysis results and confusion matrix of the MiF average and MoF average indices. BA: balanced accuracy; MiF: mild-group filtering; MoF: moderate-group filtering; TPR: true positive rate; TNR: true negative rate.



	Mild labeled	Mod-I labeled	Mild labeled	Mod-I labeled	Mild labeled	Mod-I labeled
Mild predicted	54	12	48	17	44	23
Mod-I predicted	7	37	13	32	17	26
TPR	75.5		65.3		53.1	
TNR	88.5		78.7		72.1	
BA	82.0		72.0		62.6	

Figure 7. Linear discriminant analysis results and confusion matrix of the MiF-variance and MoF-variance indices. BA: balanced accuracy; MiF: mild-group filtering; MoF: moderate-group filtering; TPR: true positive rate; TNR: true negative rate.



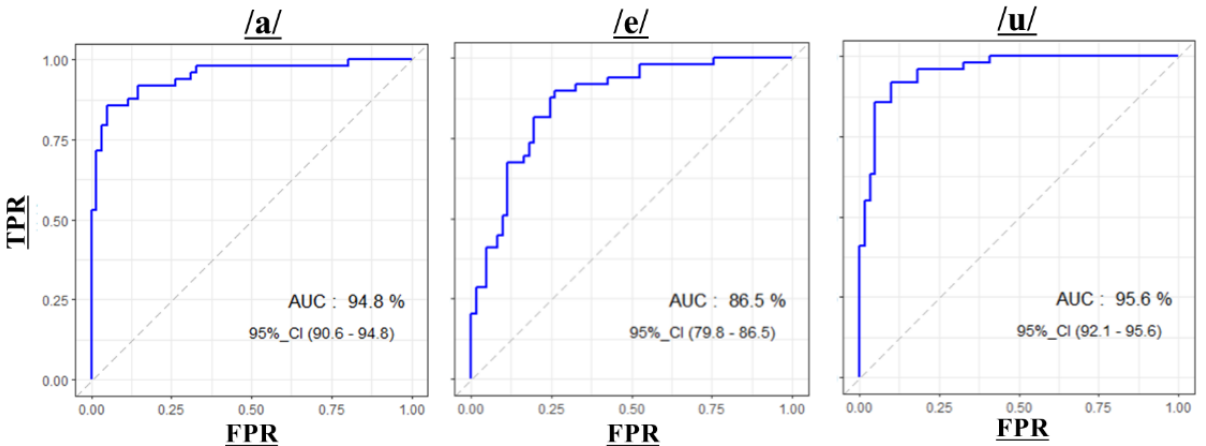
	Mild labeled	Mod-I labeled	Mild labeled	Mod-I labeled	Mild labeled	Mod-I labeled
Mild predicted	53	5	49	11	57	8
Mod-I predicted	8	44	12	38	4	41
TPR	89.8		77.6		83.7	
TNR	86.9		80.3		93.4	
BA	88.4		79.0		88.6	

GLM With the Variance Index of the DTW Distance

We used the variance indices of the DTW distance as predictors of the GLM model because they achieved better classification performances than the average indices. Figure 8 shows the ROC and AUC values of the GLM model for each vowel with the confusion matrix, including the TPR, TNR, and BA data. The models of all 3 vowels provided high model performance in

terms of the AUC and mid-to-high accuracy for TPR, TNR, and BA.

Figure 8. ROC/AUC results of the generalized linear model with confusion matrices for the mild-group filtering variance and moderate-group filtering variance indices. AUC: area under the curve; BA: balanced accuracy; FPR: false positive rate; TPR: true positive rate; TNR: true negative rate; ROC: receiver operating characteristics curve.



	Mild labeled	Mod-I labeled
Mild predicted	53	7
Mod-I predicted	8	42
TPR	85.7	
TNR	86.9	
BA	86.3	

	Mild labeled	Mod-I labeled
Mild predicted	43	5
Mod-I predicted	18	44
TPR	89.8	
TNR	70.5	
BA	80.2	

	Mild labeled	Mod-I labeled
Mild predicted	55	7
Mod-I predicted	6	42
TPR	85.7	
TNR	90.2	
BA	88.0	

Discussion

Principal Results

This feasibility study demonstrated that the DTW distance–based voice biomarkers generated by the GLM had a balanced accuracy ranging from 80.2% to 88% and high model performance, indicated by the AUC ranging from 86.5% to 96.5%, for 3 vowels when classifying between mild illness and moderate illness I in COVID-19 patients.

Comparison With Prior Work

Key 1D Features of Acoustic Parameters

Sondhi et al [10] and Pah et al [18] found that the classification of subjects with and without COVID-19 was possible using jitter, shimmer, and HNR indices; however, they did not demonstrate classification using models incorporating these parameters. Pah et al [18] stated, “The statistical analysis and SVM classification indicated that the voice features of sustained phoneme corresponding to vocal tract modulation (Mel Frequency Cepstral Coefficient (MFCC), Formants, Vocal-tract Length, and Intensity-SD) could potentially be adopted as a COVID-19 biomarker compared to the features of vocal fold vibration (jitter, shimmer, pitch, HNR, and NHR)” [18]. This suggests that a simple model using only jitter, shimmer, and HNR is not able to differentiate between subjects with and without disease. Therefore, we believe that approaches such as machine learning and deep learning are essential for performing pathophysiological analyses such as classification, presence or absence determination, and monitoring using key 1D features of acoustic parameters, such as MFCC and formants.

2D Feature Matching of DTW algorithms

In this study, the variance indices of the DTW distance were significantly different from the average indices. It appears that diseases have a wide-ranging impact on the voices of patients, making it challenging to assess and categorize voices at the desired level. To address this issue, machine learning–based voice analysis systems that focus on learning 1D parameters such as jitter and shimmer have been used, and tuning using single-dimensional key features such as spectral and prosodic speech features may be conducted [18-21]. Conversely, our 2D feature matching of the waveform using DTW algorithms may provide a more direct and practical method in the domain of binary classifications.

Advantages of the Standardization of Waveform Samples

By standardizing the time and power axes of the waveform samples in the DTW algorithm before computing, the fundamental frequency (F0) and volume were consequently transformed as parts of the elements forming a 10-cycle waveform in the unit envelope. Confounding factors derived from fundamental frequencies that vary by sex and age can be avoided as much as possible in advance, allowing the direct evaluation of the classification results by the DTW distance [8,19,22,23]. For this study, we examined 110 wavelet samples to determine the coefficient of variation (CV) of F0 estimates before and after standardization. Our goal was to analyze the variety of F0 distribution based on sex and age groups. Our

results indicated that the CV of the wavelets without standardization ranged from 10.31% to 11.4% for the 3 long vowels /a/, /e/, and /u/, which is considered a significant confounding factor. However, after standardization, the CV ranged from 0.81% to 2.4%, which was significantly reduced and, therefore, effective in minimizing confounding factors. Multimedia Appendix 1 provides more details and comparison figures.

Why Select a 10-Cycle Waveform as a Unit Sample?

Our previous pilot study using the DTW algorithm to differentiate between the voices of subjects with and without COVID-19 investigated the use of different waveforms with 1, 3, 5, 10, 20, 30, and 50 cycles. Among these waveform cycles, 10 and 20 cycles resulted in reasonable discrimination between the subjects with and without COVID-19 for all 3 subjects tested. (A summary of the test results is shown in Multimedia Appendix 2.) We selected a 10-cycle waveform, rather than a 20-cycle waveform, in order to reduce as much as possible the total computing cost incurred when using the DTW algorithm (see the section “Computing Cost”).

Robustness to Noise

While recent smartphones have improved recording features that can aid in voice analysis [24], it should be noted that the environmental noises present during the recording process were not completely controllable because the voice samples used in this study were self-recorded by the participants with their own smartphones. As Qi [25] reported, “DTW was evaluated using both synthetic and natural voices, and significant reductions in noise were achieved.” Because the DTW distance is considered to be robust to noise, it may be more practical when used for classification purposes with voices obtained from the real world compared to acoustic parameters such as jitter, shimmer, and HNR, which are considered to be sensitive to environmental noise [8,18].

Sample Size Consideration

There are known statistical correlations between significance level, power, effect size, and sample size [26]. We tested the validity of the sample size used in this study of 110 participants (61 in the mild group and 49 in the moderate I group). We calculated an effect size of 0.666 with the significance level of 1% used in this study and a power of 0.8, which generally meets requirements from a statistical point of view [27]. For the calculation, the *pwr* package (version 1.3.0) of R was used. A Cohen *d* score of 0.5 is regarded as a medium effect size, and 0.8 is regarded as a large effect size [28]. Therefore, we believe that the sample size used in this study was appropriate because the effect size of 0.666 is between medium and large. (These validation processes are disclosed in Multimedia Appendix 3.)

Future Expectations

DTW distance–based voice biomarkers may effectively supplement pulse oximeters as an objective indicator when distinguishing moderate illness I from mild illness among patients during recuperation. Even if pulse oximeters are scarce, this biomarker can be accessed through a patient’s smartphone. If persons with disease recuperating at home can detect a worsening of symptoms to moderate illness I based on changes in their voice, they will be able to determine whether they should seek medical care. In addition, this system may allow health care providers to use voice biomarkers in addition to body temperature and pulse oximeter readings as objective and quantitative indicators to properly diagnose worsening symptoms and expedite inpatient treatment.

Limitations

Computing Cost

In this study, our approach involved standardizing waveform samples, comprehensively computing the DTW distance for each sample, and subsequently using the resulting indices to determine the severity of illness using a GLM. However, this approach is considered computationally expensive, making it unsuitable for integration into standalone smartphone apps. Nevertheless, it can be used in the cloud. Despite current limitations, significant advancements in network transmission speed and information technology suggest that it may soon be practically applicable. Another potential solution for reducing computation time is to measure the DTW distance from a greater variety of representative samples as the number of cases increases. However, this remains a topic for future study.

Patient Bias

Although the definition of moderate illness I that was used in this study is based on pulse oximeter readings (SpO_2 93%-96%) or subjective reports from patients of their clinical condition (ie, SoB, as shown in Table 1), errors during labeling of the voice samples due to patient bias may have occurred as different individuals may have varying methods of describing the sensation of SoB. Unfortunately, this issue is difficult to overcome considering the use of patient-reported data. However, the inclusion of objective indicators such as voice biomarkers during the diagnosis-making process may allow for more objective labeling of data in the future.

Conclusions

Medical treatments for COVID-19 vary depending on the severity of the illness. Patients with mild illness may need only to recuperate at home or a designated facility, whereas patients with moderate illness I may need to be hospitalized. In this study, the DTW distance–based voice biomarker was tested for distinguishing between mild and moderate illness I. A balanced accuracy ranging from 80.2% to 88% was achieved, and the model performance indicated by the AUC ranged from 86.5% to 96.5% for the vowels /a/, /e/, and /u/. This voice biomarker system can be used in case of an unexpected shortage of pulse oximeters as an alternative and cost-effective method for monitoring worsening medical conditions in patients with mild illness that are recuperating at home or a medical facility.

Acknowledgments

We would like to thank Editage for English language editing. We would like to express our gratitude to Mindy Fang for introducing us to the pleasures of statistical analysis programming in R. We are deeply grateful to the Kanagawa prefectural government officials for their cooperation in this study despite their busy work on COVID-19 matters.

Data Availability

The data sets generated during and/or analyzed during this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions

TW and ST contributed to conceptualization and review and editing. TW contributed to methodology, software, validation, formal analysis, original draft preparation, and visualization. YO and ST contributed to investigation and resources. YO contributed to data curation. ST contributed to supervision, project administration, and funding acquisition. All authors have read and agreed to the manuscript.

Conflicts of Interest

YO was employed by PST Inc. This study was conducted in collaboration between PST and Kanagawa University of Human Services, but no funding for this study was received from PST.

Multimedia Appendix 1
Standardization and confounding.
[PDF File (Adobe PDF File), 291 KB - biomedeng_v8i1e50924_app1.pdf]

Multimedia Appendix 2
Validation for waveform cycles.
[PDF File (Adobe PDF File), 249 KB - biomedeng_v8i1e50924_app2.pdf]

Multimedia Appendix 3
Sample size consideration.
[PDF File (Adobe PDF File), 167 KB - biomedeng_v8i1e50924_app3.pdf]

References

1. Japan A Guide to Medical Care of COVID-19, Version 8. Ministry of Health, Labour and Welfare of Japan. URL: https://www.mhlw.go.jp/content/000936655.pdf [accessed 2023-10-12]
2. Ding H, Mandapati A, Karjadi C, Ang TFA, Lu S, Miao X, et al. Association between acoustic features and neuropsychological test performance in the Framingham heart study: Observational study. J Med Internet Res 2022 Dec 22;24(12):e42886 [FREE Full text] [doi: 10.2196/42886] [Medline: 36548029]
3. Ma A, Lau KK, Thyagarajan D. Voice changes in Parkinson's disease: What are they telling us? J Clin Neurosci 2020 Feb;72:1-7. [doi: 10.1016/j.jocn.2019.12.029] [Medline: 31952969]
4. Kappen M, van der Donckt J, Vanhollebeke G, Allaert J, Degraeve G, Madhu N, et al. Acoustic speech features in social comparison: how stress impacts the way you sound. Sci Rep 2022 Dec 20;12(1):22022 [FREE Full text] [doi: 10.1038/s41598-022-26375-9] [Medline: 36539505]
5. Higuchi M, Nakamura M, Shinohara S, Omiya Y, Takano T, Mitsuyoshi S, et al. Effectiveness of a voice-based mental health evaluation system for mobile devices: Prospective study. JMIR Form Res 2020 Jul 20;4(7):e16455 [FREE Full text] [doi: 10.2196/16455] [Medline: 32554367]
6. Shinohara S, Nakamura M, Omiya Y, Higuchi M, Hagiwara N, Mitsuyoshi S, et al. Depressive mood assessment method based on emotion level derived from voice: Comparison of voice features of individuals with major depressive disorders and healthy controls. Int J Environ Res Public Health 2021 May 19;18(10):5435 [FREE Full text] [doi: 10.3390/ijerph18105435] [Medline: 34069609]
7. Hajjar I, Okafor M, Choi JD, Moore E, Abrol A, Calhoun VD, et al. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. Alzheimers Dement (Amst) 2023;15(1):e12393 [FREE Full text] [doi: 10.1002/dad2.12393] [Medline: 36777093]
8. Asiaee M, Vahedian-Azimi A, Atashi SS, Keramatfar A, Nourbakhsh M. Voice quality evaluation in patients with COVID-19: An acoustic analysis. J Voice 2022 Nov;36(6):879.e13-879.e19 [FREE Full text] [doi: 10.1016/j.jvoice.2020.09.024] [Medline: 33051108]

9. Kaur S, Larsen E, Harper J, Purandare B, Uluer A, Hasdianda MA, et al. Development and validation of a respiratory-responsive vocal biomarker-based tool for generalizable detection of respiratory impairment: Independent case-control studies in multiple respiratory conditions including asthma, chronic obstructive pulmonary disease, and COVID-19. J Med Internet Res 2023 Apr 14;25:e44410 [FREE Full text] [doi: 10.2196/44410] [Medline: 36881540]
10. Sondhi S, Salhan A, Santoso CA, Doucoure M, Dharmawan DM, Sureka A, et al. Voice processing for COVID-19 scanning and prognostic indicator. Heliyon 2021 Oct;7(10):e08134 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e08134] [Medline: 34632133]
11. Tohidast SA, Mansuri B, Memarian M, Ghobakhloo AH, Scherer RC. Voice quality and vocal tract discomfort symptoms in patients with COVID-19. J Voice 2021 Oct 13:S0892 [FREE Full text] [doi: 10.1016/j.jvoice.2021.09.039] [Medline: 34776316]
12. Erdoğan YE, Narin A. COVID-19 detection with traditional and deep features on cough acoustic signals. Comput Biol Med 2021 Sep;136:104765 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104765] [Medline: 34416571]
13. Adhikary S, Ghosh A. Dynamic time warping approach for optimized locomotor impairment detection using biomedical signal processing. Biomed Signal Process Control 2022 Feb;72:103321. [doi: 10.1016/j.bspc.2021.103321]
14. Venail F, Legris E, Vaerenberg B, Puel J, Govaerts P, Ceccato J. Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry. Eur Ann Otorhinolaryngol Head Neck Dis 2016 Apr;133(2):101-106 [FREE Full text] [doi: 10.1016/j.anorl.2016.01.001] [Medline: 26879579]
15. Muda L, Begam M, Elamvazuthi I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv. Preprint posted online Mar 22, 2010 [FREE Full text]
16. Ribeiro Xavier C, Sachetto Oliveira R, da Fonseca Vieira V, Lobosco M, Weber dos Santos R. Characterisation of Omicron Variant during COVID-19 Pandemic and the Impact of Vaccination, Transmission Rate, Mortality, and Reinfection in South Africa, Germany, and Brazil. BioTech 2022 Apr 26;11(2):12. [doi: 10.3390/biotech11020012]
17. Skarbinski J, Wood MS, Chervo TC, Schapiro JM, Elkin EP, Valice E, et al. Risk of severe clinical outcomes among persons with SARS-CoV-2 infection with differing levels of vaccination during widespread Omicron (B.1.1.529) and Delta (B.1.617.2) variant circulation in Northern California: A retrospective cohort study. Lancet Reg Health Am 2022 Aug;12:100297 [FREE Full text] [doi: 10.1016/j.lana.2022.100297] [Medline: 35756977]
18. Pah ND, Indrawati V, Kumar DK. Voice features of sustained phoneme as COVID-19 biomarker. IEEE J Transl Eng Health Med 2022;10:4901309 [FREE Full text] [doi: 10.1109/JTEHM.2022.3208057] [Medline: 36304844]
19. Omiya Y, Mizuguchi D, Tokuno S. Distinguish the severity of illness associated with novel coronavirus (COVID-19) infection via sustained vowel speech features. Int J Environ Res Public Health 2023 Feb 15;20(4):3415 [FREE Full text] [doi: 10.3390/ijerph20043415] [Medline: 36834110]
20. Hu H, Chang S, Wang C, Li K, Cho H, Chen Y, et al. Deep learning application for vocal fold disease prediction through voice recognition: Preliminary development study. J Med Internet Res 2021 Jun 08;23(6):e25247 [FREE Full text] [doi: 10.2196/25247] [Medline: 34100770]
21. Costantini G, Dr VC, Robotti C, Benazzo M, Pietrantonio F, Di Girolamo S, et al. Deep learning and machine learning-based voice analysis for the detection of COVID-19: A proposal and comparison of architectures. Knowl Based Syst 2022 Oct 11;253:109539 [FREE Full text] [doi: 10.1016/j.knosys.2022.109539] [Medline: 35915642]
22. Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. J Voice 1999 Sep;13(3):424-446. [doi: 10.1016/s0892-1997(99)80048-4] [Medline: 10498059]
23. Berti LC, Spazzapan EA, Queiroz M, Pereira PL, Fernandes-Svartman FR, Medeiros BRD, et al. Fundamental frequency related parameters in Brazilians with COVID-19. J Acoust Soc Am 2023 Jan;153(1):576. [doi: 10.1121/10.0016848] [Medline: 36732219]
24. Uloza V, Ulozaitė-Stanienė N, Petrauskas T, Kregždytė R. Accuracy of acoustic voice quality index captured with a smartphone - measurements with added ambient noise. J Voice 2023 May;37(3):465.e19-465.e26. [doi: 10.1016/j.jvoice.2021.01.025] [Medline: 33676807]
25. Qi Y. Time normalization in voice analysis. J Acoust Soc Am 1992 Nov;92(5):2569-2576. [doi: 10.1121/1.404429] [Medline: 1479120]
26. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb) 2021 Feb 15;31(1):010502 [FREE Full text] [doi: 10.11613/BM.2021.010502] [Medline: 33380887]
27. Dumas-Mallet E, Button KS, Boraud T, Gonon F, Munafò MR. Low statistical power in biomedical science: a review of three human research domains. R Soc Open Sci 2017 Feb;4(2):160254 [FREE Full text] [doi: 10.1098/rsos.160254] [Medline: 28386409]
28. Calin-Jageman RJ. The new statistics for neuroscience majors: Thinking in effect sizes. J Undergrad Neurosci Educ 2018;16(2):E21-E25 [FREE Full text] [Medline: 30057503]

Abbreviations

AUC: area under the curve

BA: balanced accuracy
EPV10: events per variable 10
DTW: dynamic time warping
GLM: generalized linear model
HNR: harmonic-to-noise ratio
LDA: linear discriminant analysis
MiF: mild-group filtering
MFCC: mel-frequency cepstral coefficient
MoF: moderate-group filtering
ROC: receiving operator curve
SoB: shortness of breath
SpO2: oxygen saturation
TPR: true positive rate
TNR: true negative rate

Edited by T Leung; submitted 29.07.23; peer-reviewed by M Koga; comments to author 01.09.23; revised version received 08.09.23; accepted 06.10.23; published 06.11.23.

Please cite as:
Watase T, Omiya Y, Tokuno S
Severity Classification Using Dynamic Time Warping–Based Voice Biomarkers for Patients With COVID-19: Feasibility Cross-Sectional Study
JMIR Biomed Eng 2023;8:e50924
URL: https://biomedeng.jmir.org/2023/1/e50924
doi:10.2196/50924
PMID:37982072

Original Paper

Measuring Heart Rate Accurately in Patients With Parkinson Disease During Intense Exercise: Usability Study of Fitbit Charge 4

Giulia Colonna[1*]; Jocelyn Hoye[2*], PhD; Bart de Laat[1,2*], PhD; Gelsina Stanley[1*], BS; Alaaddin Ibrahimy[1*], MS; Sule Tinaz[3*], MD, PhD; Evan D Morris[1,2,4*], PhD

[1]Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, United States

[2]Department of Psychiatry, Yale University, New Haven, CT, United States

[3]Department of Neurology, Yale University, New Haven, CT, United States

[4]Department of Biomedical Engineering, Yale University, New Haven, CT, United States

[*]all authors contributed equally

Corresponding Author:
Giulia Colonna
Department of Radiology and Biomedical Imaging
Yale University
40 Temple St
New Haven, CT, 06520
United States
Phone: 1 (203) 737
Email: colonna.1844724@studenti.uniroma1.it

Abstract

Background: Parkinson disease (PD) is the second most common neurodegenerative disease, affecting approximately 1% of the world's population. Increasing evidence suggests that aerobic physical exercise can be beneficial in mitigating both motor and nonmotor symptoms of the disease. In a recent pilot study of the role of exercise on PD, we sought to confirm exercise intensity by monitoring heart rate (HR). For this purpose, we asked participants to wear a chest strap HR monitor (Polar Electro Oy) and the Fitbit Charge 4 (Fitbit Inc) wrist-worn HR monitor as a potential proxy due to its convenience. Polar H10 has been shown to provide highly accurate R-R interval measurements. Therefore, we treated it as the gold standard in this study. It has been shown that Fitbit Charge 4 has comparable accuracy to Polar H10 in healthy participants. It has yet to be determined if the Fitbit is as accurate as Polar H10 in patients with PD during rest and exercise.

Objective: This study aimed to compare Fitbit Charge 4 to Polar H10 for monitoring HR in patients with PD at rest and during an intensive exercise program.

Methods: A total of 596 exercise sessions from 11 (6 male and 5 female) participants were collected simultaneously with both devices. Patients with early-stage PD (Hoehn and Yahr ≤2) were enrolled in a 6-month exercise program designed for patients with PD. They participated in 3 one-hour exercise sessions per week. They wore both Fitbit and Polar H10 during each session. Sessions included rest, warm-up, intense exercise, and cool-down periods. We calculated the bias in the HR of the Fitbit Charge 4 at rest (5 min) and during intense exercise (20 min) by comparing the mean HR during each of the periods to the respective means measured by Polar H10 (HRFitbit – HRPolar). We also measured the sensitivity and specificity of Fitbit Charge 4 to detect average HRs that exceed the threshold for intensive exercise, defined as 70% of an individual's theoretical maximum HR. Different types of correlations between the 2 devices were investigated.

Results: The mean bias was 1.68 beats per minute (bpm) at rest and 6.29 bpm during high-intensity exercise, with an overestimation by Fitbit Charge 4 in both conditions. The mean bias of the Fitbit across both rest and intensive exercise periods was 3.98 bpm. The device's sensitivity in identifying high-intensity exercise sessions was 97.14%. The correlation between the 2 devices was nonlinear, suggesting Fitbit's tendency to saturate at high values of HR.

Conclusions: The performance of Fitbit Charge 4 is comparable to Polar H10 for assessing exercise intensity in a cohort of patients with PD (mean bias 3.98 bpm). The device could be considered a reasonable surrogate for more cumbersome chest-worn devices in future studies of clinical cohorts.

(JMIR Biomed Eng 2023;8:e51515) doi:10.2196/51515

KEYWORDS

Fitbit; heart rate measurements; Parkinson disease; exercise; accuracy; intensity; heart rate; wearable; neurodegenerative disease; aerobic exercise; physical exercise; program; device

Introduction

Background

Parkinson disease (PD) is the second most common neurodegenerative disease and affects approximately 1% of the world's population [1]. The main symptoms characterizing this disease are bradykinesia, rigidity, tremor, and postural instability as well as nonmotor symptoms, such as anxiety, depression, sleep disturbance, and fatigue. Evidence suggests that aerobic physical exercise can be beneficial in mitigating motor symptoms and slowing the progression of the disease [1-3].

The extent of benefits observed differs depending on the exercise type, intensity, and duration. Various recent clinical trials have concluded that moderate- to high-intensity exercise several times per week, when maintained over extended periods, is associated with slower deterioration of motor symptoms in PD [4,5].

Since different types of exercise interventions at varying intensity levels are used in clinical trials for PD and other clinical populations, there is a need for objective methods to monitor the intensity of physical activity. The popularity of wearable devices has grown, as they have become more affordable, useful, and less intrusive [6]. However, it is still necessary to establish the reliability of these devices in tracking physiological parameters during both clinical trials and personal use.

Wearables can measure many different parameters, such as heart rate (HR), number of steps, calories expended, and quality of sleep. HR is considered an essential indicator of physiological adjustment and intensity of effort [7]. HR is correlated linearly with moderate- and vigorous-intensity physical exercise and is a valuable option to monitor the intensity of activities (eg, cycling, swimming, and activities that are not ambulatory) that may not be easily measured with other methods, such as accelerometry [8]. Following the American Heart Association guidelines, vigorous exercise intensity can be defined as 70% to 85% of the maximum HR [9]. Many different tools can be used to assess HR, such as electrocardiogram (ECG) monitors; chest, shoulder and arm straps; and wrist watches.

In this study, we compared Polar H10 (Polar Electro Oy) and Fitbit Charge 4 (Fitbit Inc), two types of wearable devices that are commonly used as activity trackers and HR monitors.

The Polar Heart rate (referred to as Polar H10 in this paper) is a chest strap that uses ECG technology to measure the R-R interval. The Polar H10 has been reported to be highly correlated to 3-lead ECG Holter monitor (r=0.997) and is now considered the gold standard for assessment of R-R intervals in sports settings [10] as well as maintaining a certain accuracy in older adults affected by cardiac disease [11]. Despite its accuracy, the Polar H10 monitor is often perceived as too cumbersome to use, and it may cause discomfort, especially for older people [12]. Since it needs to be strapped across the sternum, it may be difficult to tolerate over extended periods [12]. On the other hand, as a wrist-worn tracker, Fitbit Charge 4 (referred to as Fitbit in this paper) is more convenient and comfortable to wear, and it promotes patient compliance in studies requiring prolonged measurements [13]. According to Düking et al [14], wrist-worn wearables, being able to provide direct biofeedback, have the potential to increase participation in exercise.

Fitbit Charge 4 is a recent model of the Fitbit Charge HR series, released in March 2020. It is a wrist-worn device that detects HR by measuring the volume changes in blood vessels via a photoplethysmography (PPG) optical HR sensor [7]. Originally designed to motivate people to exercise, Fitbits are increasingly used as measurement devices in physical activity and health promotion research; they are also used for guiding patient-health professional interactions [15].

Fitbits are commonly used for research purposes [16], but there is no consensus in the scientific literature regarding their accuracy for quantifying HR and confirming high intensity. Some authors have concluded that the device provides values of HR comparable with criterion field-based measures, while others have found that Fitbit does not satisfy the validation criteria, especially during higher exercise intensities [17]. There is even less information on the accuracy of the device in older individuals affected by chronic diseases [13]. Further evaluation is needed.

Ensuring the accuracy of exercise session intensity assessment is crucial in clinical studies involving sports activities and clinical populations. It has yet to be determined if Fitbit has comparable accuracy to Polar H10 in selecting high-intensity sessions in patients with PD or in clinical populations, generally.

Objective

This paper aims to compare Fitbit Charge 4 to Polar H10 for monitoring HR, confirming high-intensity exercises in patients with PD engaging in an intense exercise program and supporting its potential utility as an activity tracker for use in large clinical trials with similar cohorts.

Methods

Population and Study Design

The data for this paper were acquired as part of a larger study to evaluate the role of physical exercise in PD, in which we sought to confirm exercise intensity by monitoring HR. In brief, a total of 11 participants, 6 of whom were male, aged 58-68 years, all with early-stage PD (defined according to the Movement Disorder Society criteria [18]) were recruited. Participants were excluded based on the criteria of the larger study, as follows: (1) heavy drinking or illicit drug use, (2) neurologic or psychiatric disorders other than PD, (3) diseases

interfering with one's ability to exercise, (4) contraindication to positron emission tomography or magnetic resonance imaging scans, (5) severe motor symptoms (tremor and dyskinesia) likely to introduce motion artifacts in imaging data, (6) unsafe to come off dopaminergic medication, (7) BMI>30 (practical issues with the neuroimaging equipment), (8) extreme exercisers, and (9) Hoehn and Yahr disease stage>2 (stage 2 corresponds to mild bilateral disease with intact balance [19]). None of the participants had a history of arrhythmias or any other cardiac conditions that could potentially affect the measurements of the devices. Additionally, none of them were under medication, such as AV nodal blockade therapy, which might have altered the HR detection capabilities of Polar H10 and Fitbit Charge 4.

Each participant engaged in exercise for a period of 6 months, with at least 3 Beat Parkinson's Today (BPT) exercise sessions per week. The BPT program is an established exercise program that combines those aspects that have been shown to be the most effective in achieving symptom improvement in PD [20], such as high-intensity interval training and boxing [21,22]. Each session included a mix of these 2 activities, which could be adapted to any fitness level. Functional interval training circuits were designed specifically to improve explosiveness, gait, and strength. Trainers continuously encouraged participants to work at their own personal level of maximum intensity while attempting to reach a target HR.

To compare the performances of devices, participants were equipped with a Fitbit and a Polar H10. The Fitbit was worn on the wrist and positioned a finger's width above the wrist bone, as recommended by the company. The wrist-worn tracker was situated on the side less affected by PD. The Polar HR sensor was placed over the sternum and held in place by a chest strap. The exercise sessions lasted 60 minutes, including warm-up, at least 20 minutes of high-intensity exercise, boxing, and cool-down. When unable to attend classes, participants were encouraged to exercise on their own and monitor their HR using both wearables.

Ethical Considerations

All procedures with human subjects were approved by the Human Investigations Committee of Yale University (approval number 2000028563).

HR Data Processing From Wearables

The data from Fitbit Charge 4 were collected by synchronizing each watch with an anonymized web-based account for each participant and downloaded via the mobile app Pulse Watch [23]. The data from Polar H10 were exported using the mobile app Elite HRV. The Fitbit data were sampled every minute by the Pulse Watch app.

The data from Elite HRV were converted from R-R intervals to beats per minute (bpm) and were filtered using a Python function called Butterworth filter to remove high-frequency artifacts. The order was set to 5, and the filter was applied at a frequency of 0.1 Hz. The resulting data set was in units of bpm collected per 10 seconds and was then sampled every minute. The Fitbit data were sampled every minute by the Pulse Watch app.

The validity of Fitbit was compared with Polar H10 in terms of averages between single data points. For each session, HR averages (HRμ) from both the first 5 minutes, generally coinciding with the rest period before the start of the exercises, and from the 20 minutes of the highest-intensity exercise were calculated. The 20 minutes of the highest-intensity exercise were extracted from the data by calculating the HR average for consecutive 20-minute intervals, starting from the initial interval, then shifting forward by 1 minute at a time, and then picking the highest average.

The session averages were then collected, and the values from the 2 different devices were paired. Data alignment, filtering, and calculations were performed with Python (Jupyter Notebook). The precise timestamps for recordings from both devices were available and were used for data alignment. Out of a possible 792 paired sessions, a total of 596 paired sessions were obtained. Data were lost due to multiple factors—nonattendance of the participants, misplacement of the devices, and injuries. The data obtained were contributed roughly equally by all the participants, with a mean of 54 sessions per participant (Table 1).

To ensure the capability of Fitbit, compared to Polar H10, in evaluating the intensity of an exercise session, every HR average (HRμ) was normalized (HR_N) by different percentages of each participant's own theoretical maximum HR ($HR_{th/max}$) using the following formula:

where the term $HR_{th/max}$ is given by the following: $HR_{th/max}$ = 220 – age

A session was considered positive if the ratio was >1 and negative if the ratio was <1. Measurements with the Polar H10 were considered to be the gold standard. Therefore, a session was a "true positive" if the ratio was >1 for both devices, "true negative" if the ratio was <1 for both devices, "false positive" if the ratio was >1 for Fitbit and <1 for Polar H10, and "false negative" if the ratio was <1 for Fitbit and >1 for Polar H10.

Table 1. Mean difference and limits of agreement (LoA) calculated individually by participant.

Participant	Age (years)	Recorded sessions, n	Mean difference at baseline conditions (bpm[a])	LoA (bpm)	Mean difference at high-intensity exercise conditions (bpm)	LoA (bpm)
Participant 1	58	81	2.77	–20.4 to 14.8	–2.52	–17.3 to 18.7
Participant 2	73	43	–1.25	–19.4 to 16.9	–11.26	–13.9 to 35.7
Participant 3	60	55	1.93	–26.4 to 30.1	0.42	–31.6 to 30.8
Participant 4	63	38	–1.35	–16.2 to 13.5	–5.93	–4.65 to 16.5
Participant 5	63	56	0.05	–12 to 11.9	–16.37	–7.46 to 37.6
Participant 6	76	98	0.61	–14.1 to 15.3	–15.94	–9.50 to 41.3
Participant 7	56.3	61	–6.94	–31.3 to 17.4	9.44	–36.2 to 17.3
Participant 8	68	28	–8.01	–20.4 to 4.40	–12.14	–10.2 to 25.2
Participant 9	66.6	43	–1.25	–25.4 to 22.9	–5.65	–6.18 to 17.5
Participant 10	66.6	56	–2.71	–21.9 to 16.1	–20.10	–12.9 to 49.3
Participant 11	68	36	0.82	–18.3 to 19.9	5.91	–27.4 to 15.5

[a]Bpm: beats per minute.

HR Data Comparisons Between Wearables

All statistical analyses were conducted using Microsoft Excel 16 and MatLab (Mathworks, 2018b). To guarantee consistency of the results and the calculations of HR averages at baseline and high-intensity conditions, exercise sessions lasting less than 20 minutes were excluded from the analysis.

A Bland-Altman plot was used to evaluate the agreement between the two methods of measurement, with the limits of agreement (LoA) defined as the mean difference plus or minus 1.96 SD of the difference. The mean difference in HR between the Fitbit and Polar H10 was calculated for the cohort and every participant, both at baseline (rest period) and during high-intensity conditions. A final average of the two mean differences was assessed and considered to be the mean bias. Evaluations were conducted for both intrasubject and intersubject variability (Table 1).

The relationship between both devices at baseline and during high-intensity conditions was determined. The quality of the linear fit was assessed with the R^2 value, considering the data from baseline and high-intensity conditions separately. Subsequently, the entire data set of session HR averages was fitted with linear, logarithmic, negative exponential, and sigmoid model functions to explore different types of relationships between the two devices. The Akaike information criterion (AIC) value was used to assess the relative quality of the fits.

Sensitivity and Specificity

To determine Fitbit's sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) in identifying high-intensity exercise sessions, European and American Guidelines cutoffs were applied [24]. Polar H10 was considered to be the gold standard.

To illustrate the sensitivity and specificity of the Fitbit, a receiver operating characteristic (ROC) curve was created for different levels of target HR. The area under the curve (AUC) was used as an indicator of Fitbit's capability for distinguishing between high-intensity and low-intensity exercise sessions.

Results

The Bland-Altman plots revealed that the mean bias between the Fitbit and Polar H10 was 1.68 bpm (LoA –21.52 bpm to 18.8 bpm) at baseline conditions and 6.29 bpm (LoA –22.02 bpm to 36.2 bpm) under high-intensity exercise (Figure 1). Overall, the mean bias of the Fitbit was 3.98 bpm.

When data from baseline and high-intensity conditions were taken separately, the linear correlations were, respectively, as follows: R^2=0.45 (baseline); R^2=0.23 (high-intensity condition; Figure 2).

The fit of the combined high-intensity and baseline data to a sigmoid model resulted in the lowest AIC value (AIC=6.03e+03; Figure 3).

When the mean differences were calculated individually by participant, there was evidence of intersubject and intrasubject variability (Table 1).

With 70% of maximum HR as the tailored threshold indicating high intensity and considering Polar H10 as the gold standard, the sensitivity, specificity, PPV, and NPV of the Fitbit were 97%, 11%, 89%, and 35%, respectively. With 85% of the maximum HR as the threshold, the sensitivity, specificity, PPV, and NPV of the Fitbit were 78%, 56%, 62%, and 73%, respectively (Table 2).

These indicators of performance at the 2 different thresholds can be visualized graphically in Figure 4. As expected, when the threshold was set higher, the number of true positives decreased, and the sensitivity of the Fitbit decreased.

The ROC curve (Figure 5) depicts the performance of Fitbit Charge 4 for varying HR thresholds. The AUC was 0.71.

Figure 1. Bland-Altman plots for the difference in heart rate average by every session. The solid line represents the mean bias. The dashed lines represent the limits of agreement. Dots of different colors represent different participants. (A) Bland-Altman plots at baseline conditions and (B) at high-intensity conditions. Bpm: Beats per minute.

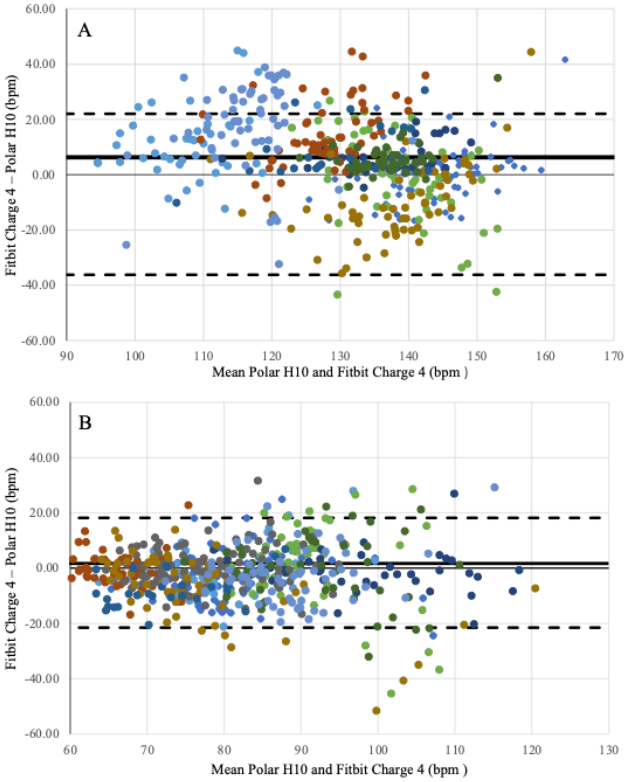


Figure 2. Linear correlations between heart rate measurements of Fitbit Charge 4 and Polar H10. The dots represent each exercise session, the solid line represents the ideal correlation (X=Y), and the dashed line is the observed correlation. Dots of different colors represent different participants. (A) Linear correlation plot at baseline and (B) during high-intensity exercise. Bpm: Beats per minute.

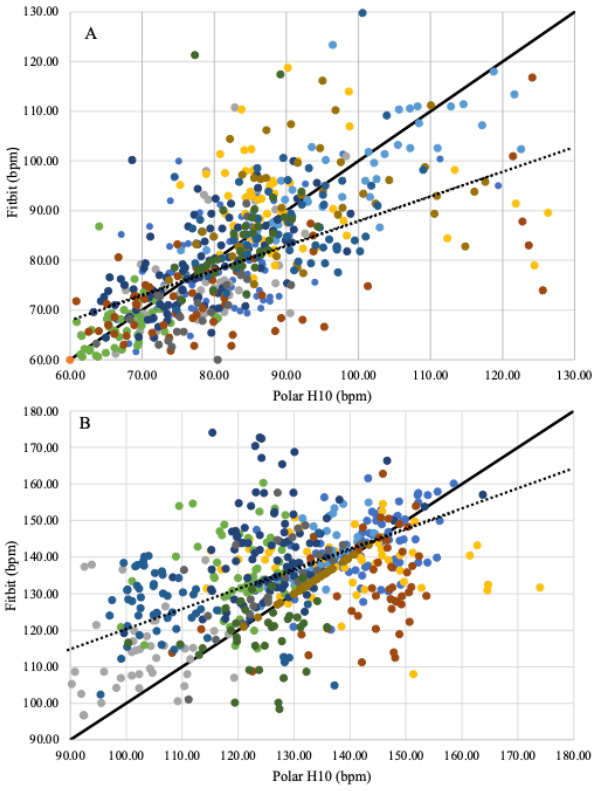


Figure 3. Model fits of Fitbit Charge 4 versus Polar H10. (A) linear, (B) logarithmic, (C) A-exponential, and (D) sigmoid. The blue dots represent heart rate averages from every session; the red solid lines represent the fitted curves. Akaike information criterion (AIC) values and fitting equations are shown as well. Bpm: Beats per minute.

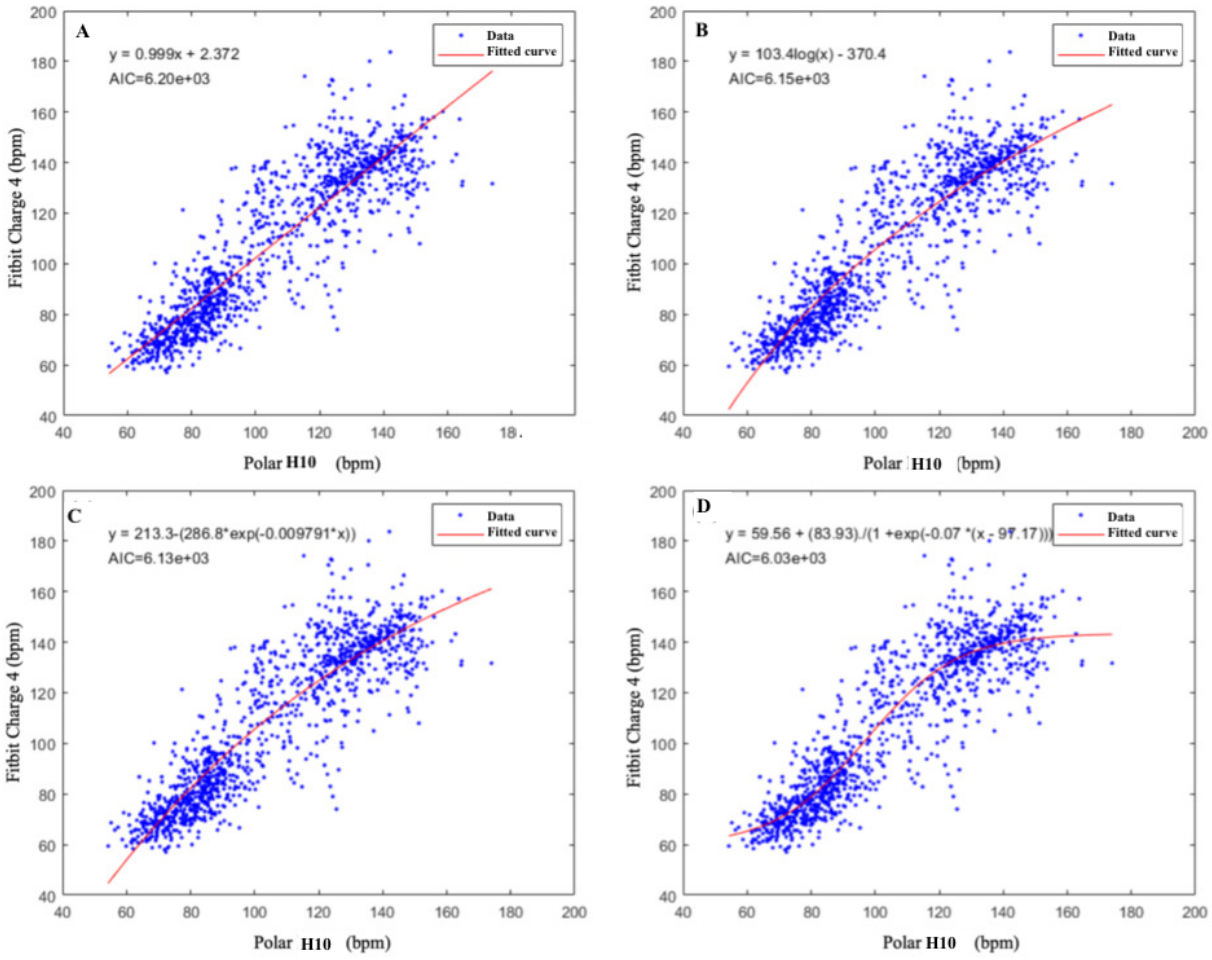


Table 2. Sensitivity and specificity of Fitbit Charge 4 given 70% and 85% of the maximum heart rate (HR) as thresholds.

Threshold	True positives, n	False positives, n	True negatives, n	False negatives, n	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)
70% of the maximum HR	509	63	8	15	97.1	11	89	35
85% of the maximum HR	223	134	172	64	78	56	62	73

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

Figure 4. (A) The number of true positives, false positives, true negatives, and false negatives based on a target of 70% of maximum heart rate (HR) and (B) based on a target of 85% of the maximum HR. The y-axis and x-axis are normalized; the dots represent the normalized value of each exercise session. The blue dots represent the true positives (>1 for both devices); the red dots represent the false positives (>1 for Fitbit Charge 4 and <1 for Polar H10); the yellow dots represent the true negatives (>1 for both devices); and the green dots represent the false negatives (<1 for Fitbit Charge 4 and >1 for Polar H10).

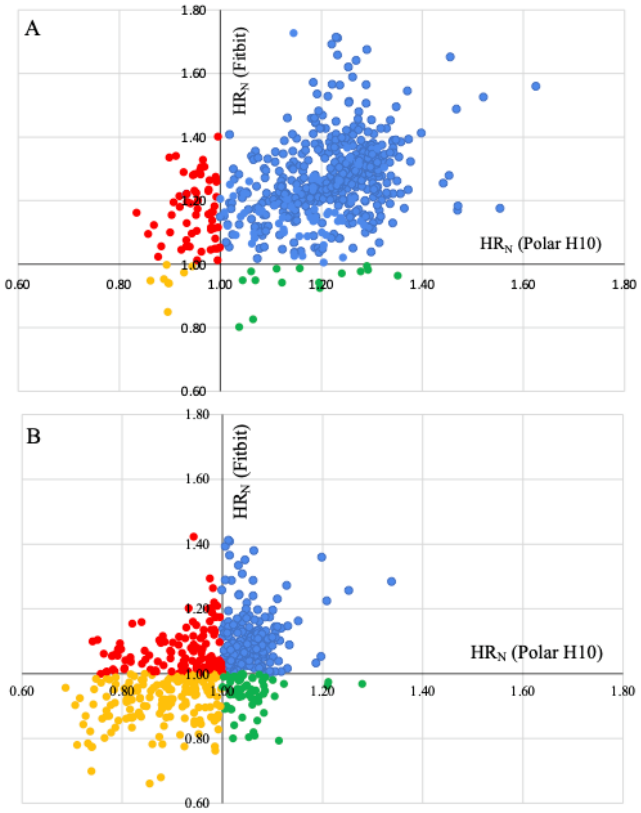
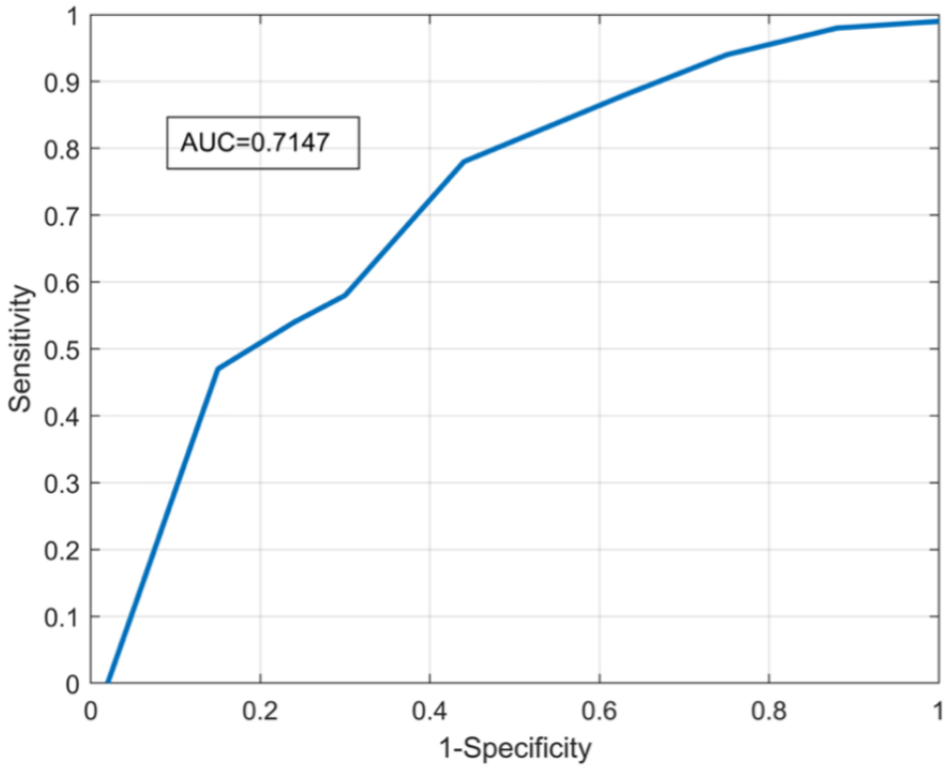


Figure 5. Receiver operating characteristic curve of Fitbit Charge 4 taking Polar H10 as the gold standard. AUC: area under the curve.



Discussion

Principal Results

To our knowledge, this is the first evaluation of Fitbit Charge 4 with a population of individuals with PD. We found the bias to be approximately 3.98 bpm during intense exercise. The magnitude of bias in the Fitbit is consistent with a report from a meta-analysis published in 2022 [16] (mean bias 3.39 bpm; LoA –24.3 bpm to 17.53 bpm). Thus, we conclude that the bias of the device, while remaining statistically significant in comparison to more precise devices, like Polar H10, is not influenced by PD and its associated symptoms, such as tremors and rigidity. Rather, it may be a limitation of the manufacturer's software process for processing data from the PPG technology. The algorithm adopted by the company to estimate HR from the PPG measurements is confidential, but some authors suggested that it may rely on the P-P intervals of the PQRST wave (atrial contractions). Not all the P peaks are consistently present and detectable when the frequency increases. ECGs, on the other hand, register the full PQRST wave and quantify the final HR by using the interval between R peaks, making it less prone to these artifacts. R peaks (ventricular contractions) are the best detectable peaks in the PQRS wave. Consequently, the calculation of the HR by PPG can be influenced by inaccurate sampling and recording of the P peaks [7]. Moreover, the performance of Fitbit, as with all wrist-based devices, is dependent on correct placement on the wrist. Therefore, when wrist movements are greater, measurement accuracy may be compromised. Devices like Polar H10, which are placed on the chest, may be less prone to movement artifacts [25].

Bland-Altman analysis showed that Fitbit tended to overestimate the values of HR compared to Polar H10 in high-intensity conditions. This result contradicts most of the previous reports [7,15,16]. The overestimation could be due to the peculiar characteristics of this study population (older adults affected by PD). These 2 conditions are known to potentially increase the heterogeneity of Fitbit accuracy results [16]. In previous studies, Fitbit overestimated time spent on moderate to vigorous activity in clinical populations with functional limitations, compared to the criterion devices [26]. Even though in our cohort, the disease appeared not to have an impact on Fitbit's magnitude of error, it could have altered the sign of the error. There are some additional conditions of our study to consider. The maximum HR values reached by our participants cannot be compared with those reached by a cohort of young, healthy individuals. If HR values had been higher, we might have observed an underestimation by Fitbit. Another important factor to consider is the particular Fitbit model. The only Fitbit Charge 4 validation study was conducted in 2022 [17], which evaluated the device on 23 young participants (average age 24.2 years) without any underlying health conditions.

The linear correlation between the two devices was poor, especially in high-intensity exercise conditions (baseline: R^2=0.45; high-intensity conditions: R^2=0.23; Figure 2). From our statistical analysis, the sigmoid fit, which resulted in the lowest AIC, best described the relationship between Fitbit and Polar H10 (Figure 3). The tendency of Fitbit to saturate at the highest HR values suggests a diminishing ability of Fitbit to resolve high HR values. This finding is in agreement with the existing literature [27]. The reduced precision of Fitbit in measuring high HR values may be attributed to motion artifacts due to physical movement, particularly those involving arm movements, as well as potential misalignment between the skin and the optical sensor [27]. Another hypothesis suggests that wrist-worn devices may not be as sensitive to sudden changes in exercise intensity [25], which occur frequently in high-intensity interval training, as used in our study. Peripheral resistance is lower at the wrist, which reduces pulse pressure changes and alters blood pulse detection. [25]. Although the sigmoid function was the best fit for the data acquired in this cohort, we caution against using the sigmoid model to extrapolate the relationship between the Fitbit and chest strap HRs beyond the range of HR values acquired in this study. In other words, if future studies in patients with PD seek to acquire Fitbit data only (no chest strap data) and want to use this model to predict the chest strap HR, the authors recommend only applying the model to data with HR in the range of 60 bpm to 160 bpm.

We also examined the ability of Fitbit to discriminate HR during high-intensity sessions, via the ROC curve. An AUC of 0.5 generally indicates no discrimination; an AUC of 0.7 to 0.8 indicates acceptable discrimination; an AUC of 0.8 to 0.9 indicates excellent discrimination; and an AUC of more than 0.9 indicates exceptional discrimination [28]. The Fitbit's measurement can be considered acceptable (AUC 0.71; Figure 5). Consequently, this device is acceptable in identifying correctly high-intensity exercise sessions and could be used with caution in large clinical trials in patients with PD.

Limitations

Our study is not without limitations. First, the number of sessions is not equally distributed between participants. Thus, some participants may have exerted a greater impact on the total mean difference between the devices than others, as shown in Table 1.

We considered the first 5 minutes of every exercise session as the baseline, during which participants were instructed to sit and breathe. However, there were instances of participants arriving late or forgetting to activate the device at the start of the session, potentially confounding the baseline measurements. Consequently, the values of HR recorded during baseline conditions may have been artifactually high.

When participants were unable to attend classes, they were encouraged to exercise independently while monitoring HR using both devices. However, during these unsupervised sessions, we were unable to ensure the proper fit of both devices, potentially affecting the accuracy of the measurements obtained.

The data processing involved multiple stages of averaging, ranging from a subsecond level to a per-minute level and ultimately to an exercise-session level. Although this averaging approach allowed us to accomplish the study objectives, it may have potentially compromised the precision and reliability of our comparisons.

Due to the inclusion and exclusion criteria of the previous study, only 11 participants, affected by mild PD were taken into account. It is possible that the outcomes would have been different with the inclusion of participants with severe PD, affected by motor symptoms likely to introduce motor artifacts in wearables data. Given the small number of participants and the peculiar characteristics of the cohort, our findings may not be applicable to all patients with PD. For future studies, it may be crucial to involve participants at more advanced stages of the disease to effectively assess Fitbit's performance under these conditions.

Lastly, the cohort in our study demonstrated significant intrasubject and intersubject variability, attributed to factors such as age, sex, and physical condition. Intrasubject variability is represented by each participant's LoA and intersubject variability is depicted as each participant's HR mean difference (Table 1). The wide LoA observed in the mean bias of Fitbit Charge 4 emphasizes some reasons for cautious interpretations of the results.

Conclusions

The magnitude of bias and the LoA for Fitbit were consistent with those of previous studies, and the performance of Fitbit fell within the range of 4 bpm, compared to Polar H10 for assessing intense exercise in a cohort of patients with PD. A wrist-worn device, Fitbit, offers clear advantages in terms of wearability and practicality. In future studies involving clinical populations, the device could be considered as a reasonable alternative to the more intrusive chest strap technology.

Acknowledgments

We would like to acknowledge the staff of Beat Parkinson's Today for training the participants and Jose Key in computer programming. The authors EDM and ST, as well as the research conducted for the larger previous study based on Parkinson disease and physical exercise, were supported by grant R21NS118764. GC was supported by a fellowship from the ONAOSI-YALE Undergraduate Research Experience in Instrumentation.

Data Availability

The data sets generated and analyzed during the study are available from the corresponding author on a reasonable request.

Authors' Contributions

GC was in charge of the study conceptualization, methodology, software, and writing of the original draft; JH contributed to the methodology, conceptualization, and formal analysis, as well as writing, reviewing, and editing the manuscript. BdL was responsible for resources and data curation. GS also contributed to providing resources. AI managed the software. ST was in charge of project administration. EDM helped in the writing, reviewing, and editing of the manuscript; supervision; project administration; and funding acquisition.

Conflicts of Interest

None declared.

References

1. Mak MKY, Wong-Yu ISK. Exercise for Parkinson's disease. Int Rev Neurobiol 2019;147:1-44. [doi: 10.1016/bs.irn.2019.06.001] [Medline: 31607351]
2. Ahlskog JE. Aerobic exercise: evidence for a direct brain effect to slow Parkinson disease progression. Mayo Clin Proc 2018 Mar;93(3):360-372. [doi: 10.1016/j.mayocp.2017.12.015] [Medline: 29502566]
3. Li X, Gao Z, Yu H, Gu Y, Yang G. Effect of long-term exercise therapy on motor symptoms in Parkinson disease patients. Am J Phys Med Rehabil 2022 Jun 11;101(10):905-912. [doi: 10.1097/phm.0000000000002052]
4. Tsukita K, Sakamaki-Tsukita H, Takahashi R. Long-term effect of regular physical activity and exercise habits in patients with early Parkinson disease. Neurology 2022 Jan 12;98(8):e859-e871. [doi: 10.1212/wnl.0000000000013218]
5. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. Lancet Neurol 2019 Nov;18(11):998-1008. [doi: 10.1016/S1474-4422(19)30285-6] [Medline: 31521532]
6. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
7. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018 Feb;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
8. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57(4):387-395. [doi: 10.1016/j.pcad.2014.10.005] [Medline: 25446555]
9. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical asctivity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]

10. Hinde K, White G, Armstrong N. Wearable devices suitable for monitoring twenty four hour heart rate variability in military populations. Sensors (Basel) 2021 Feb 04;21(4):1061 [FREE Full text] [doi: 10.3390/s21041061] [Medline: 33557190]
11. Vermunicht P, Makayed K, Meysman P, Laukens K, Knaepen L, Vervoort Y, et al. Validation of Polar H10 chest strap and Fitbit Inspire 2 tracker for measuring continuous heart rate in cardiac patients: impact of artefact removal algorithm. EP Europace 2023;25(Supplement_1):euad122.550 [FREE Full text] [doi: 10.1093/europace/euad122.550]
12. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR wireless heart rate monitor: validation study conducted under free-living conditions. JMIR Mhealth Uhealth 2017 Oct 20;5(10):e157 [FREE Full text] [doi: 10.2196/mhealth.8233] [Medline: 29055881]
13. Martinato M, Lorenzoni G, Zanchi T, Bergamin A, Buratin A, Azzolina D, et al. Usability and accuracy of a smartwatch for the assessment of physical activity in the elderly population: observational study. JMIR Mhealth Uhealth 2021 May 05;9(5):e20966 [FREE Full text] [doi: 10.2196/20966] [Medline: 33949953]
14. Düking P, Tafler M, Wallmann-Sperlich B, Sperlich B, Kleih S. Behavior change techniques in wrist-Worn wearables to promote physical activity: content analysis. JMIR Mhealth Uhealth 2020 Nov 19;8(11):e20820 [FREE Full text] [doi: 10.2196/20820] [Medline: 33211023]
15. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
16. Chevance G, Golaszewski NM, Tipton E, Hekler EB, Buman M, Welk GJ, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 13;10(4):e35626 [FREE Full text] [doi: 10.2196/35626] [Medline: 35416777]
17. Nissen M, Slim S, Jäger K, Flaucher M, Huebner H, Danzberger N, et al. Heart rate measurement accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: device evaluation study. JMIR Form Res 2022 Mar 01;6(3):e33635 [FREE Full text] [doi: 10.2196/33635] [Medline: 35230250]
18. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, et al. Measurement of heart rate using the Polar OH1 and Fitbit Charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. JMIR Mhealth Uhealth 2021 Mar 25;9(3):e25313. [doi: 10.2196/25313] [Medline: 33764310]
19. Berg D, Adler CH, Bloem BR, Chan P, Gasser T, Goetz CG, et al. Movement disorder society criteria for clinically established early Parkinson's disease. Mov Disord 2018 Oct 25;33(10):1643-1646. [doi: 10.1002/mds.27431] [Medline: 30145841]
20. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology 1967 May 01;17(5):427-442 [FREE Full text] [doi: 10.1212/wnl.17.5.427] [Medline: 6067254]
21. van Nimwegen M, Speelman AD, Smulders K, Overeem S, Borm GF, Backx FJ, ParkFit Study Group. Design and baseline characteristics of the ParkFit study, a randomized controlled trial evaluating the effectiveness of a multifaceted behavioral program to increase physical activity in Parkinson patients. BMC Neurol 2010 Aug 19;10(1):70 [FREE Full text] [doi: 10.1186/1471-2377-10-70] [Medline: 20723221]
22. Fernandes B, Barbieri FA, Arthuso FZ, Silva FA, Moretto GF, Imaizumi LFI, et al. High-intensity interval versus moderate-intensity continuous training in individuals with Parkinson's Disease: hemodynamic and functional adaptation. J Phys Act Health 2020 Jan 01;17(1):85-91. [doi: 10.1123/jpah.2018-0588] [Medline: 31810064]
23. Combs S, Diehl M, Chrzastowski C, Didrick N, McCoin B, Mox N, et al. Community-based group exercise for persons with Parkinson disease: a randomized controlled trial. NRE 2013 Feb 28;32(1):117-124. [doi: 10.3233/nre-130828]
24. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, ESC National Cardiac Societies, ESC Scientific Document Group. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 07;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
25. Martinez J, Gordon T, Strath S. The use of consumer wearable physical activity monitors in clinical populations with functional limitations. J Phys Med Rehabil 2021;3(2):73-90 [FREE Full text] [doi: 10.33696/rehabilitation.3.022] [Medline: 35224553]
26. Dondzila CJ, Lewis CA, LOpez JR, Parker TM. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci 2018;11(7):575-584 [FREE Full text]
27. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
28. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]

Abbreviations

AIC: Akaike information criterion
AUC: area under the curve
Bpm: beats per minute

ECG: electrocardiogram
HR: heart rate
LoA: limits of agreement
NPV: negative predictive value
PD: Parkinson disease
PPG: photoplethysmography
PPV: positive predictive value
ROC: receiver operating characteristic

Edited by T Leung; submitted 02.08.23; peer-reviewed by M Nissen, D Lachant; comments to author 23.08.23; revised version received 27.10.23; accepted 09.11.23; published 08.12.23.

Please cite as:
Colonna G, Hoye J, de Laat B, Stanley G, Ibrahimy A, Tinaz S, Morris ED
Measuring Heart Rate Accurately in Patients With Parkinson Disease During Intense Exercise: Usability Study of Fitbit Charge 4
JMIR Biomed Eng 2023;8:e51515
URL: https://biomedeng.jmir.org/2023/1/e51515
doi:10.2196/51515
PMID:38875680

Original Paper

Assessment of Skin Maturity by LED Light at Birth and Its Association With Lung Maturity: Clinical Trial Secondary Outcomes

Gabriela Silveira Neves[1], PT, MSc; Zilma Silveira Nogueira Reis[1], MD, PhD; Roberta Romanelli[1], MD, PhD; James Batchelor[2], MD

[1]Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2]Faculty of Medicine, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Gabriela Silveira Neves, PT, MSc
Faculty of Medicine
Universidade Federal de Minas Gerais
190, Professor Alfredo Balena Avenue
Belo Horizonte, 30130-100
Brazil
Phone: 55 31995150401
Email: neves.gabriela87@gmail.com

Abstract

Background: Clinicians face barriers when assessing lung maturity at birth due to global inequalities. Still, strategies for testing based solely on gestational age to predict the likelihood of respiratory distress syndrome (RDS) do not offer a comprehensive approach to addressing the challenge of uncertain outcomes. We hypothesize that a noninvasive assessment of skin maturity may indicate lung maturity.

Objective: This study aimed to assess the association between a newborn's skin maturity and RDS occurrence.

Methods: We conducted a case-control nested in a prospective cohort study, a secondary endpoint of a multicenter clinical trial. The study was carried out in 5 Brazilian urban reference centers for highly complex perinatal care. Of 781 newborns from the cohort study, 640 were selected for the case-control analysis. Newborns with RDS formed the case group and newborns without RDS were the controls. All newborns with other diseases exhibiting respiratory manifestations were excluded. Skin maturity was assessed from the newborn's skin over the sole by an optical device that acquired a reflection signal through an LED sensor. The device, previously validated, measured and recorded skin reflectance. Clinical data related to respiratory outcomes were gathered from medical records during the 72-hour follow-up of the newborn, or until discharge or death, whichever occurred first. The main outcome measure was the association between skin reflectance and RDS using univariate and multivariate binary logistic regression. Additionally, we assessed the connection between skin reflectance and factors such as neonatal intensive care unit (NICU) admission and the need for ventilatory support.

Results: Out of 604 newborns, 470 (73.4%) were from the RDS group and 170 (26.6%) were from the control group. According to comparisons between the groups, newborns with RDS had a younger gestational age (31.6 vs 39.1 weeks, P<.001) and birth weight (1491 vs 3121 grams, P<.001) than controls. Skin reflectance was associated with RDS (odds ratio [OR] 0.982, 95% CI 0.979-0.985, R^2=0.632, P<.001). This relationship remained significant when adjusted by the cofactors antenatal corticosteroid and birth weight (OR 0.994, 95% CI 0.990-0.998, R^2=0.843, P<.001). Secondary outcomes also showed differences in skin reflectance. The mean difference was 0.219 (95% CI 0.200-0.238) between newborns that required ventilatory support versus those that did not and 0.223 (95% CI 0.205-0.241) between newborns that required NICU admission versus those that did not. Skin reflectance was associated with ventilatory support (OR 0.996, 95% CI 0.992-0.999, R^2=0.814, P=.01) and with NICU admission (OR 0.994, 95% CI 0.990-0.998, R^2=0.867, P=.004).

Conclusions: Our findings present a potential marker of lung immaturity at birth using the indirect method of skin assessment. Using the RDS clinical condition and a medical device, this study demonstrated the synchrony between lung and skin maturity.

Trial Registration: Registro Brasileiro de Ensaios Clínicos (ReBEC) RBR-3f5bm5; https://tinyurl.com/9fb7zrdb

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-027442

(JMIR Biomed Eng 2023;8:e52468) doi:10.2196/52468

KEYWORDS

newborn infant; prematurity; neonatal respiratory distress syndrome; skin physiological phenomena; photometer; gestational age

Introduction

Respiratory system maturation occurs in late gestation in preparation for the time of birth [1], extending into early childhood [2]. Essential for normal lung development, epigenetic mechanisms are influenced by the environment throughout gestation and postnatally [3]. Therefore, whether term or preterm, newborns might have immature ventilatory function that may foster respiratory instability [1].

Difficulties in assessing lung maturity arise when a single parameter is considered since the systems and organs may be at different stages of matureness. This is noticeable when judging maturity by gestational age when divergent pulmonary functional maturity is found between peers of the same age [4]. At different stages of development, organs can be affected by growth and differentiation factors released from other organs, as suggested by studies with multidimensional scaling and hierarchical cluster analysis of growth patterns during the fetal period [5].

Regarding the ability to interact with the external environment through the skin, a full-term newborn presents a complete or near entirely competent epidermal barrier at birth. The skin barrier at maturity, with an efficient stratum corneum achieved at around 34 weeks [6], prevents transepidermal water loss from the skin surface, maintaining the newborn's temperature [7]. An immature skin barrier at birth leads to hypothermia, which in turn increases the likelihood of developing respiratory distress syndrome (RDS), intraventricular hemorrhage, late-onset sepsis, and mortality [8].

Similar to the skin, the lung shows signs of readiness for extrauterine life in the last trimester of gestation [9], with peak of alveoli maturation and surfactant production at 35 weeks of gestation [10]. Reduced surfactant, the major cause of RDS, causes low functional residual lung capacity, increasing the work of breathing and leading to terminal airway collapse. As a result, an increased ventilation-perfusion mismatch can lead to the need for ventilatory support [11], which in turn increases the metabolic and caloric demand to maintain temperature [12]. Although the point of view of care for lung immaturity is often related to the inability of the epidermal barrier to retain heat, there are few studies in this regard. Taesch et al (1972) [13] demonstrated that skin age is an indicator of lung age by studying rabbits [13].

In terms of access to lung maturity assessment, as well as advanced neonatal care, there is inequality around the world [14]. In these circumstances, knowing the risks of respiratory morbidity with accuracy might help in making more balanced decisions and determining the most appropriate care. Still, testing strategies based on gestational age for predicting the likelihood of RDS do not provide a complete approach to addressing the dilemma of indeterminate outcomes [10]. In response, noninvasive assessment methods have been proposed. A new photobiological device proved to correctly classify preterm newborns with 91.4% accuracy using a mathematical algorithm based on skin maturity and clinical adjusters [15]. In this context, the ability to accurately assess skin maturation and the potential synchrony of skin-lung development enables the study of a possible marker of lung maturation. The aim of this study was to assess the relationship between newborn skin maturity and RDS.

Methods

Setting

This study was carried out in 5 Brazilian urban reference centers for highly complex perinatal care in different regions: in the southeast, Hospital de Clínicas of Universidade Federal de Minas Gerais (as coordinator) and Hospital Sofia Feldman; in the south, Hospital of Universidade Luterana do Brasil; in the center-west, the Hospital Materno Infantil de Brasília; and in the northeast, the University Hospital of Universidade Federal do Maranhão.

Ethics Approval

The trial protocol received approval from the independent ethics review board at each reference center under the number 81347817.6.1001.5149 at the Brazilian National Research Council. The procedures followed the Helsinki Declaration of 1975, as revised in 2013 [16], and all parents provided informed consent on behalf of their newborns before participating in the clinical trial.

Study Design

This was a case-control nested in a prospective cohort study to investigate a secondary outcome within a single-blinded multicenter clinical trial investigation with a single group and single arm. The clinical trial protocol was disclosed in the World Health Organization's International Clinical Trial Platform—Brazilian Clinical Trials (registered under trial number RBR-3f5bm5).

Participants

In the primary cohort, a concurrent and sequential process enrolled newborns who were up to 24 hours old, had a gestational age of at least 24 weeks as determined by standard ultrasound, and were recruited between January 2, 2019, and May 30, 2021. Skin maturity assessment was conducted within the first 24 hours of life, regardless of the newborn's location, whether it was in an incubator, heated crib, bassinet in the hospital room, or on the mother's lap. All participants were followed for a period of 72 hours or until discharge or death, whichever occurred first, for the assessment of lung maturity. The examiner, who was blind to the results of the skin assessment, collected respiratory outcome data from medical charts. This study focused on the 72-hour follow-up data. More details about the study protocol can be found in a previous

publication [17]. The timeframe of enrolment, intervention with the optical device, and respiratory outcome measurements is depicted in Table 1, adapted from Reis et al [18].

Table 1. Study timeline.

	Study period			
	Enrolment	Assessment	Close-out	Allocation
Time point	0 hours	0 hours	72 hours	Analysis
Enrolment				
Eligibility	X			
Informed consent	X			
Optical device intervention		X		
Assessment and analysis				
Optical device data acquisition		X		X
Standard ultrasound	X			X
Case-control nested study		X	X	

We included newborns diagnosed with RDS based on clinical and radiological criteria after reviewing their clinical records. Newborns with immature lungs diagnosed with RDS formed the case group, and newborns with mature lungs without a respiratory diagnosis were randomly paired by gestational age ranges to form the control group. Newborns with extrapulmonary conditions, tachypnea due to causes other than prematurity, and diagnosis of infection were excluded.

Skin Assessment

The skin assessment occurred with an optical device previously detailed [15] (Figure 1). Briefly, to obtain the skin reflectance, an LED sensor of wavelengths 400 to 1200 nm was touched to the newborn's sole for a few seconds to trigger 10 automated measurements. We performed 3 measurements, resulting in 30 automatic values to obtain the average reflection. The processor then captured the variations resulting from the interaction of the skin and LED light and kept it in storage for analysis. The data processor estimated lung maturity using machine learning algorithms. The evaluation advocated minimal manipulation, being performed in the position where the newborn was, after hand hygiene and sensor disinfection. The best body position to assess skin reflectance and possible influences, such as humidity, temperature, ambient light, and the skin tone of the newborn, were evaluated beforehand [19,20]. The reliability of skin assessment with the device was previously reported. The intraobserver and interobserver variability were 1.97% (95% CI 1.84%-2.11%) and 2.6% (95% CI 2.1%-3.1%), respectively [15].

Figure 1. Steps of skin assessment in 8 steps: (1) the device touches the skin; (2) the standard body position for assessing skin reflectance in newborns is the sole; (3) 3 measurements are taken simultaneously; (4) the LED light interacts with the skin, scattering the light, and the returned light (reflectance) toward the sensor is processed by a control unit and stored for analysis; (5) the user inputs clinical data, such as birth weight and prenatal corticosteroid use; (6) the user collects vital data during the procedure; (7) the data are recorded and stored for analysis; and (8) the data processor estimates lung maturity using machine learning algorithms, associating light reflection and respiratory outcomes.

The results obtained by the equipment were concealed from the researchers. The readings were stored in the processor and later transmitted to an electronic database, where they were stored on data servers. The results, apart from being inaccessible to the examiner, were also not shared with the professionals responsible for the child's care in the actual scenario. This approach ensured that the study did not interfere with the clinical decisions made by the health care professionals.

Clinical Data

To ensure proper data acquisition, all examiners were trained according to good clinical practice as recommended by the Brazilian Regulatory Health Agency [17]. Data related to respiratory outcomes were collected from medical records. During data curation, the senior clinician analyzed and confirmed the RDS diagnosis according to the guidelines previously described in the study protocol [21]. The framework of the clinical variables and skin acquisitions is available in Multimedia Appendix 1, as documented in the previous report by Reis et al [15].

We developed dedicated software to collect structured clinical data and associate them with the skin reflection of each newborn from 5 perinatal centers simultaneously. Examiners used individual sets of instruments, including a tablet, optical device, and paper versions of the clinical data forms. A double approach, on paper and electronically, allowed verification of clinical data for reliability and validity and was later validated by specialists in data curation.

Primary Outcome

The primary outcome was the association between RDS occurrence and skin light reflection.

The diagnosis of RDS was based on a previously published clinical trial protocol [21]. In brief, it considered clinical, laboratory, and radiological findings. The observations made during the first 72 hours of life included tachydyspnea, the need for oxygen or ventilatory support after 24 hours of age, the requirement for surfactant replacement, and abnormal X-ray findings. The main radiological signs included underinflated lungs and a pattern of diffuse "ground glass" reticulogranular opacities, along with reduced lung volume and air bronchograms.

Secondary Outcomes

The secondary outcomes were the association between the skin reflectance and NICU admission and the need for ventilatory support. Both invasive and noninvasive ventilatory supports were taken into account, including supplemental oxygen by nasal cannula or hood, nasal continuous positive airway pressure, noninvasive ventilation with biphasic positive airway pressure, and invasive mechanical ventilation through the endotracheal tube.

Statistical Analysis

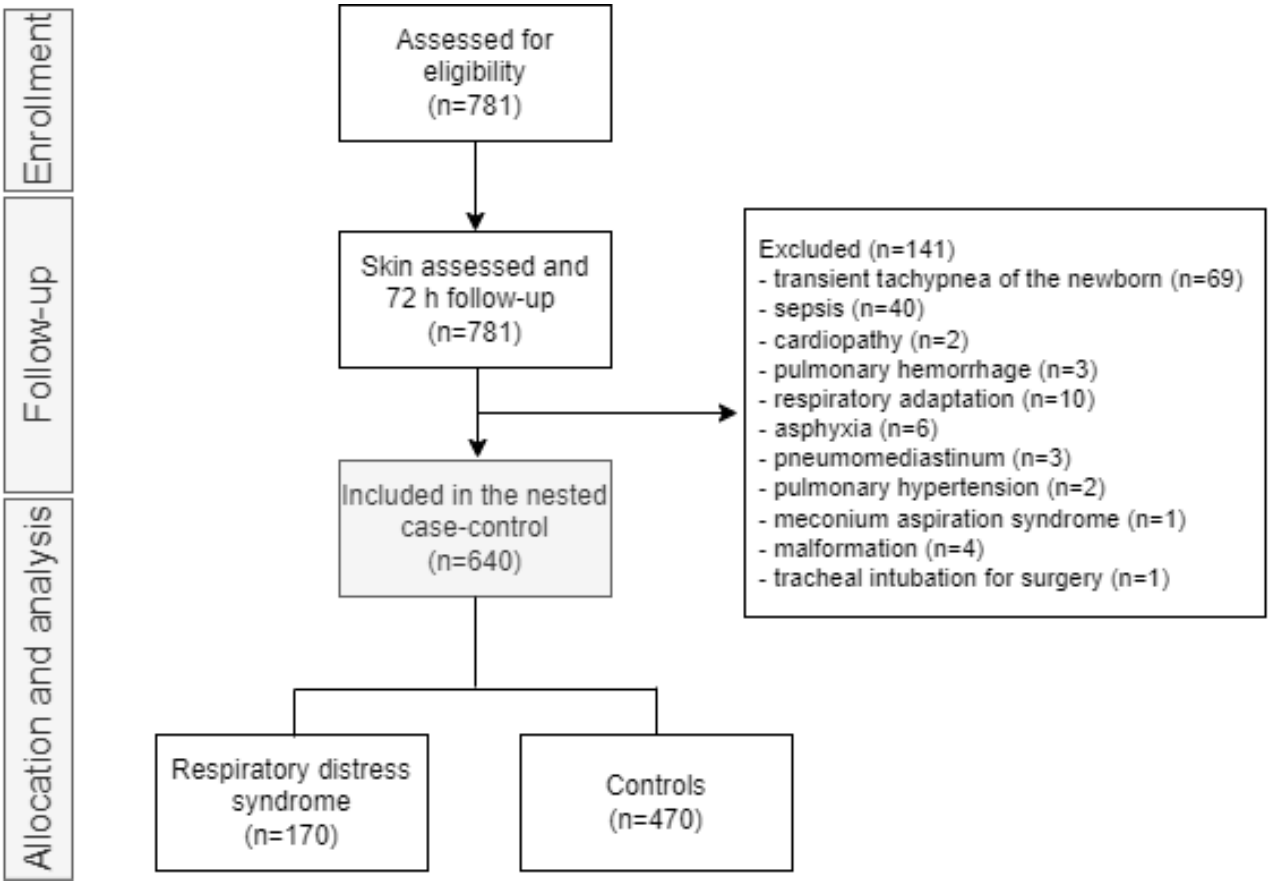
Descriptive statistics were used to explore the demographic and clinical characteristics of newborns according to groups of interest among the RDS and control groups. The analysis was performed by calculating the frequencies and percentages of categorical variables. The central tendency, mean (SD), median (IQR), and dispersion were calculated for quantitative variables. The independent sample *t* test or Mann-Whitney test was used to compare continuous variables, and the χ^2 test or Fisher exact test was used to compare categorical variables according to the nature of their distribution.

The inferential statistical analysis evaluated the relationship between skin reflectance and outcome occurrence. The sensor acquisition produced by the skin reflection was the independent variable. Logistic regression was used to identify potential influencers of RDS occurrence as birth weight and antenatal corticoid exposition. Similar analyses were conducted with the secondary outcomes, NICU admission and ventilatory support needs. Inference was estimated by calculating odds ratios (ORs) with 95% CIs. The Nagelkerke R^2 was used to measure how well the independent variables explained the variance in the model. A Wald test was used to confirm if a set of independent variables were collectively significant for the model. The variables corresponding to P values <.05 in the univariate analysis were selected for the multivariate analysis, and analyses were performed using the available data with imputation of the missing data. The statistical software SPSS 25.0 (IBM) was used for the analysis.

Results

At the end of the cohort study of 781 newborns, 640 were selected for the case-control analysis according to the eligibility criteria (Figure 2).

Figure 2. Flow diagram of participants included in the 72-hour follow-up study.



The main characteristics of antenatal care and newborn infants are shown in Table 2. There were 5 missing data points from 4 newborns, including information on the use of antenatal corticosteroid therapy for fetal maturation exposition (ACTMF), the presence of diabetes, the first minute Apgar score, and the fifth minute Apgar score. According to comparisons between groups, newborns with RDS had a younger gestational age (31.6 vs 39.1 weeks, P<.001) and lower birth weight (1491 vs 3121 grams, P<.001) than controls. During the 72-hour follow-up, significantly different rates were found between groups regarding NICU admission (RDS: n=170, 100%; control: n=16, 3.4%; P<.001), mortality (RDS: n=5, 9.2%; control: n=0, 0%; P<.001), and the need for ventilatory support (RDS: n=160, 100%; control: n=6, 1.3%; P<.001).

Table 2. Clinical characteristics of the studied newborns.

Variable	Total (n=640)	RDS[a] (n=170)	Controls (n=470)	*P* value (RDS vs control)
Maternal characteristics, n (%)				
ACTMF[b]	192 (30)	147 (87)	45 (9.6)	<.001
Diabetes	90 (14.1)	41 (24.1)	49 (10.4)	<.001
HDP[c]	129 (20.2)	66 (38.8)	63 (13.4)	<.001
Multiple gestation	101 (15.8)	67 (39.4)	34 (7.2)	<.001
Demographic data at birth				
GA[d] (weeks), median (IQR)	38.1 (33.8-39.9)	31.6 (29.9-33.4)	39.1 (37.4-40.1)	<.001
Preterm, n (%)	250 (39.1)	169 (99.4)	81 (17.2)	<.001
Sex, male, n (%)	320 (50)	89 (52.4)	231 (49.1)	.53
Birth weight (g), median (IQR)	2688 (905)	1491 (513)	3121 (561)	<.001
Birth weight classification, n (%)				<.001
>2500 g	420 (65.6)	4 (1)	416 (99)	
LBW[e]	130 (20.3)	79 (46.5)	51 (10.9)	
VLBW[f]	58 (9.1)	55 (32.4)	3 (0.6)	
ELBW[g]	32 (5)	32 (18.8)	0 (0)	
1-minute Apgar score, median (IQR)	9 (8-9)	8 (6-9)	9 (8-9)	<.001
5-minute Apgar score, median (IQR)	9 (9-10)	9 (8-10)	9 (9-10)	<.001
Neonatal resuscitation first steps, n (%)	268 (41.9)	160 (94.1)	108 (23)	<.001
Neonatal resuscitation steps, PPV[i], n (%)	95 (14.8)	75 (44.1)	20 (4.3)	<.001
Neonatal resuscitation steps, intubation, n (%)	26 (4.1)	24 (14.1)	2 (0.4)	<.001
Advanced resuscitation, n (%)	2 (0.3)	2 (1.2)	0 (0)	<.001
Follow-up within 72 hours, n (%)				
NICU[j] admission	186 (29.1)	170 (100)	16 (3.4)	<.001
Discharge	395 (61.7)	0 (0)	395 (84)	<.001
Mortality	5 (0.8)	5 (2.9)	0 (0)	.07
Incubator	166 (25.9)	152 (89.4)	14 (3)	<.001
Ventilatory support	176 (27.5)	170 (100)	6 (1.3)	<.001
MV[k]	54 (8.4)	51 (30)	3 (0.6)	<.001
NIV[l]	43 (6.7)	43 (25.3)	0 (0)	<.001
CPAP[m]	156 (24.4)	153 (90)	3 (0.6)	<.001
NC[n]	2 (0.3)	2 (1.2)	0 (0)	<.001
Surfactant therapy	73 (11.4)	73 (42.9)	0 (0)	<.001

[a]RDS: respiratory distress syndrome.

[b]ACTMF: antenatal corticosteroid therapy for fetal maturation exposition.

[c]HDP: hypertensive disorders of pregnancy.

[d]GA: gestational age.

[e]LBW: low birth weight (<2500 g).

[f]VLBW: very low birth weight (<1500 g).

[g]ELBW: extremely low birth weight (<1000 g).

[h]Not available.

[i]PPV: positive pressure ventilation.

[j]NICU: neonatal intensive care unit.

[k]MV: invasive mechanical ventilation.

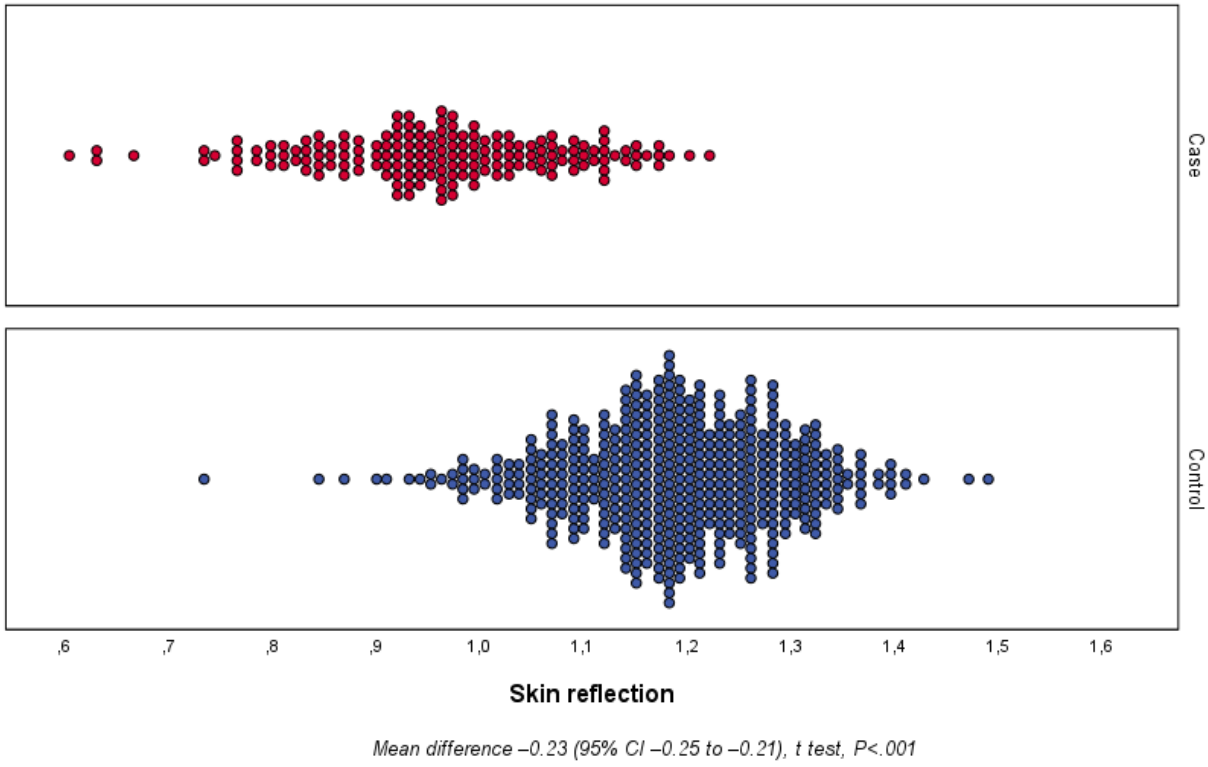
[l]NIV: noninvasive mechanical ventilation with bilevel-positive airway pressure.

[m]CPAP: continuous positive airway pressure.

[n]NC: oxygen by nasal cannula.

Concerning the primary outcome, different reflectance of the skin over the sole was observed between the groups studied (Figure 3). The reflectance range for the RDS group was 0.588-1.208 with a mean of 0.945 (SD 0.118), and that for the control group was 0.717-1.274 with a mean of 1.172 (SD 0.103). The mean difference of reflectance between groups was –0.227 (95% CI –0.246 to –0.208; P<.001).

Figure 3. Primary outcome: newborn skin reflection acquired on the sole of the foot on the first day of life for the case and control groups.



The univariate analysis showed a correlation between skin reflection and RDS. The skin reflectance was associated with RDS in the univariate analysis (OR 0.982, 95% CI 0.979-0.985, R^2=0.632, P<.001) as well as in the cofactor-adjusted analysis (OR 0.994, 95% CI 0.990-0.998, R^2=0.843, P<.001) (Table 3). Skin reflection was associated with RDS regardless of the inclusion of ACTMF and birth weight in the multivariate model.

Table 3. Univariate and multivariate analyses of the association between skin maturity and the occurrence of respiratory distress syndrome acquired by the optical device.

Variable	Univariate analysis			Multivariate analysis		
	OR[a] (95% CI)	*P* value (Wald test)	R^2	OR (95% CI)	*P* value (Wald test)	R^2
Skin reflection[b]	0.982 (0.979-0.985)	<.001	0.632	0.994 (0.990-0.998)	.001	0.843
Birth weight	0.995 (0.994-0.996)	<.001	0.825	0.996 (0.996-0.997)	<.001	
ACTMF[c]	63.106 (36.655-108.646)	<.001	0.621	2.854 (1.207-6.749)	.02	

[a]OR: odds ratio.

[b]OR and 95% CI values for skin reflection are × 10^3.

[c]ACTMF: antenatal corticosteroid therapy for fetal maturation exposition.

Secondary outcome data showed differences in skin reflectance between the studied groups for both ventilatory support use (yes vs no; Figure 4) and NICU admission (yes vs no; Figure 5). Regarding ventilatory support, the skin reflectance ranged from

0.588 to 1.305 with a mean of 0.952 (SD 0.009) and 0.717 to 1.474 with a mean of 1.172 (SD 0.005), for the yes and no groups, respectively. The mean difference was 0.219 (95% CI 0.200-0.238; *P*<.001). For NICU admission, the skin reflectance ranged from 0.588 to 1.304 with a mean of 0.953 (SD 0.009) and 0.717 to 1.473 with a mean of 1.176 (SD 0.005) for the yes and no groups, respectively. The mean difference was 0.223 (95% CI 0.205-0.241; *P*<.001).

Figure 4. Secondary outcome: newborn skin reflection acquired on the sole of the foot in the first 24 hours of life according to ventilatory support use (yes or no).

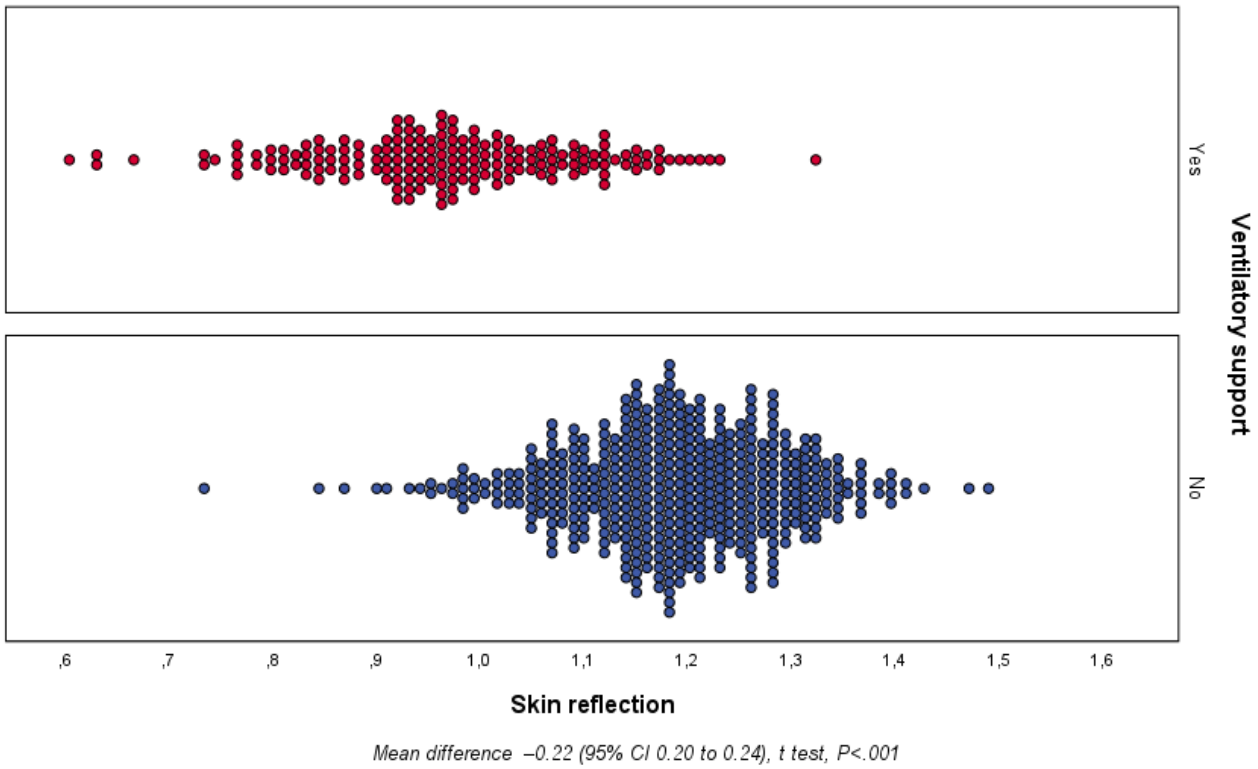
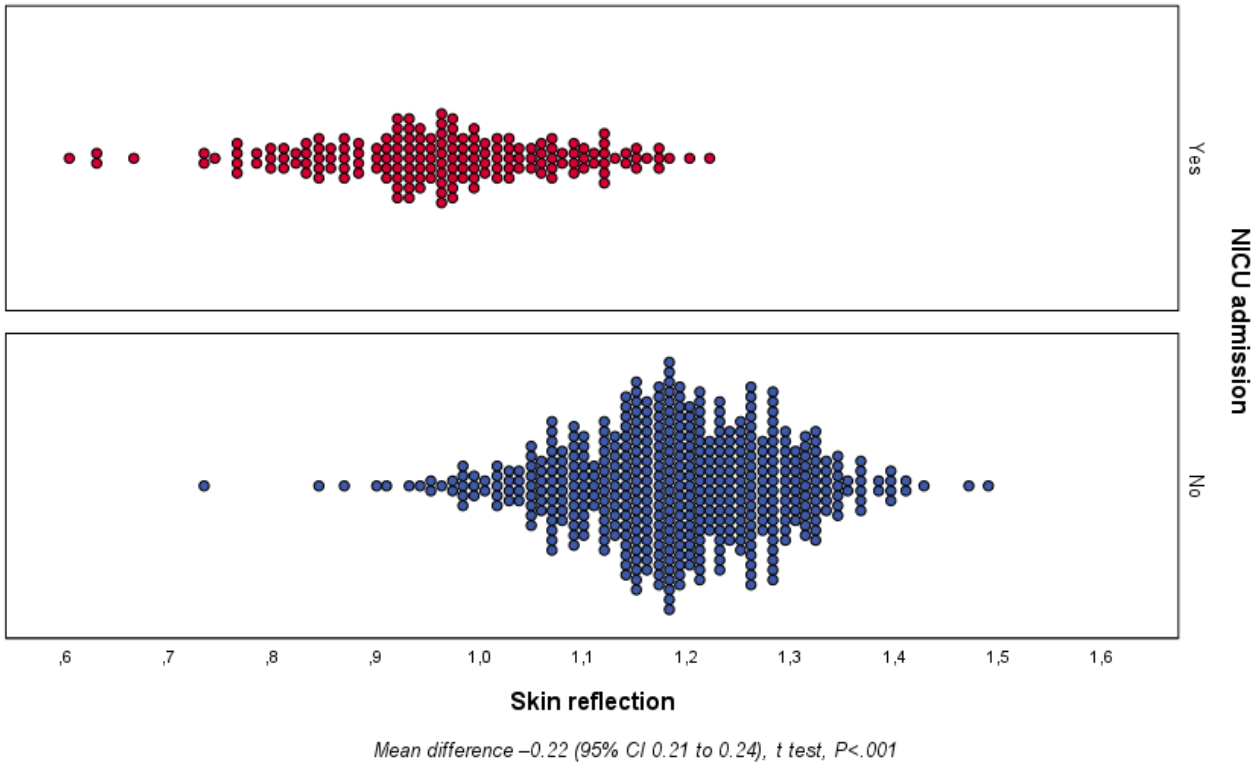


Figure 5. Secondary outcome: newborn skin reflection acquired on the sole of the foot in the first 24 hours of life according to NICU admission (yes or no). NICU: neonatal intensive care unit.



The univariate and multivariate analyses of secondary outcomes are summarized in Tables 4 and 5 for ventilatory support use and NICU admission, respectively. Skin reflection was associated with the need for ventilatory support in the univariate analysis (OR 0.983, 95% CI 0.981-0.986, R^2=0.598, P<.001) as well as in the cofactor-adjusted analysis (OR 0.996, 95% CI 0.992-0.999, R^2=0.814, P=.01). Similarly, there was an association between skin maturity and NICU admission in the univariate analysis (OR 0.928, 95% CI 0.979-0.985, R^2=0.635, P<.001) and multivariate analysis (OR 0.994, 95% CI 0.990-0.998, R^2=0.867, P=.004).

Table 4. Univariate and multivariate analyses of the association between skin maturity and the need for ventilatory support during the first 72 hours of life.

Variable	Univariate analysis			Multivariate analysis		
	OR[a] (95% CI)	*P* value (Wald test)	R^2	OR (95% CI)	*P* value (Wald test)	R^2
Skin reflection[b]	0.983 (0.981-0.986)	<.001	0.598	0.996 (0.992-0.999)	.01	0.814
Birth weight	0.996 (0.995-0.993)	<.001	0.801	0.997 (0.996-0.998)	<.001	
ACTMF[c]	56.108 (33.307-94.520)	<.001	0.607	2.677 (1.209-5.924)	.01	

[a]OR: odds ratio.

[b]OR and 95% CI values for skin reflection are × 10^3.

[c]ACTMF: antenatal corticosteroid therapy for fetal maturation exposition.

Table 5. Univariate and multivariate analyses of the association between skin maturity and neonatal intensive care unit admission during the first 72 hours of life.

Variable	Univariate analysis			Multivariate analysis		
	OR[a] (95% CI)	*P* value (Wald test)	R^2	OR (95% CI)	*P* value (Wald test)	R^2
Skin reflection[b]	0.928 (0.979-0.985)	<.001	0.635	0.994 (0.990-0.998)	.004	0.867
Birth weight	0.995 (0.994-0.996)	<.001	0.852	0.996 (0.995-0.997)	<.001	
ACTMF[c]	72.288 (42.238-123.715)	<.001	0.648	2.908 (1.223-6.9155)	.02	

[a]OR: odds ratio.

[b]OR and 95% CI values for skin reflection are × 10^3.

[c]ACTMF: antenatal corticosteroid therapy for fetal maturation exposition.

Discussion

Principal Findings

The main aim of this study was to demonstrate the use of skin maturity assessment as a potential marker of lung maturation. We found an association between skin immaturity and the occurrence of RDS, as well as a similar association with other respiratory outcomes, such as NICU admission and the need for ventilatory support. Skin reflectance at newborns' soles, assessed within the first 24 hours of life using an optical device, indicated its maturity. Respiratory outcomes related to lung maturity at 72 hours of life were obtained from medical records. These results could enhance neonatal care, as knowledge of lung maturity, regardless of the newborn's gestational age, can facilitate individualized care in the first hours of life.

Our findings reinforce the theory of the parallel development of the organs, highlighting the similarity between stages, which may allow the indirect evaluation of an organic system based on the measurements of another, regardless of age [13]. Studies in animal models have shown the similarity in the process of lipid production between the stratum corneum and pulmonary surfactant [22]. Predicting surfactant deficiency before respiratory deterioration depends on a combination of clinical signs and lung imaging [23]. Therefore, a timely indication of surfactant therapy may be postponed due to various factors, such as the lack of specificity in the initial phase of the imaging methods, the absence of the exam in low- and middle-income countries, and the masked signs of RDS severity by early continuous positive airway pressure protocols [24,25]. In this context, skin assessment appears to be a potential alternative, as this study demonstrated an association between skin reflectance and the need for ventilatory support in the first 72 hours of life. To the best of our knowledge, this is a pioneering study of pulmonary assessment using an indirect and noninvasive method.

This study presented no intention of predicting RDS or other complications related to pulmonary immaturity since it had a nested case-control design. Removing confounding diseases, such as sepsis, malformations, and other respiratory diagnoses, was essential to analyze the relationship between skin maturity and RDS. This analysis plan was stated in the clinical trial protocol [18]. Additionally, ACTMF and birth weight were included in the multivariate model to investigate the independent and adjusted association of skin-lung maturity. A comprehensive sample of newborns with train-test procedures on machine

learning approaches is still necessary to provide valuable models of RDS prediction using skin reflection.

Furthermore, this study found an association between skin reflection and NICU admission when adjusted by the cofactors of birth weight and ACTMF. Previous reports demonstrated the likelihood of preterm newborns, especially those with low birth weight, requiring NICU monitoring due to hypothermia [22]. This population is prone to uncontrolled heat loss, leading to hypoglycemia and hypoxemia, resulting in metabolic acidosis and associated respiratory distress [26]. Predicting which infants will become symptomatic of RDS is not always possible before birth. If not recognized and managed quickly, respiratory distress can escalate to respiratory failure and cardiopulmonary arrest [27]. Clinical evaluation of fetal lung maturity based on the analysis of the lecithin to sphingomyelin ratio and lamellar body count demands amniocentesis, which is an invasive procedure that poses potential risks, such as preterm labor, fetomaternal hemorrhage, or even death [28]. Furthermore, there is a large difference in sensitivity and specificity among laboratory analyses depending on the test [25]. Therefore, antenatal assessment of fetal lung maturity is limited, and we believe that a noninvasive method to assess lung maturity at birth can meet the actual clinical needs.

Medical technologies for monitoring fetal and maternal health are not equally accessible [29]. In low- and middle-income countries, among the challenges is the transfer of the newborn to specialized services due to suboptimal modes of transport and difficult and time-consuming routes [30]. Therefore, early assessment of lung maturity could improve resource allocation, supporting the indication for transport, and potentially lowering the mortality risk. The studied device can be easily applied by several health professionals, favoring assistance especially in low-resource settings. This study is the first step in this context, showing a possible agreement between lung and skin maturation, relying on RDS strictly due to immaturity.

Limitations

To assess pulmonary readiness for the extrauterine life, our data analysis was based on the RDS scenario strictly caused by immaturity and is therefore unable to predict the occurrence of RDS. Since this case-control study excluded other neonatal conditions, a comprehensive analysis including all causes of respiratory distress at birth is necessary.

Strengths

As far as we know, this was the first study to demonstrate the phenomenon of association between skin and lung maturity assessed postnatally. Through an accessible technological tool that can be integrated into current clinical practice, a therapeutic possibility arises for the indirect assessment of lung maturity.

Despite the case-control methodology, data collection was prospective, the research protocol was previously published, and the team was trained and certified [17].

Future Directions

To develop and validate a predictive model for RDS, a study that includes all causes of respiratory distress at birth is necessary. We already have a study underway for this purpose.

Conclusion

This study showed the potential for identifying RDS and immediate respiratory complications in the first 72 hours of life through skin assessment based on the synchronous development of the lungs and skin. The results, however, may not be applicable for predicting RDS, ventilatory support use, or NICU admission.

Acknowledgments

This study was sponsored by Fundação Oswaldo Cruz (grant VPPIS-002-FEX-20) and the Grand Challenges Canada's program, Government of Canada (grant R-ST-POC-1807-13515). ZSNR is a researcher with a grant from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant 305837/2021-4). The funders played no role in the design of the study, the collection and analysis of the data, the manuscript preparation, or the decision to publish the results. We acknowledge open-access funding by the Bill & Melinda Gates Foundation.

This work was supported during the development of the photometer by the Bill & Melinda Gates Foundation (grant OPP1128907) and the Fundação de Amparo a Pesquisa de Minas Gerais (grants AUC-00032-15 and PPM 0247/16; nonprofit sector).

GSN thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico of the Ministry of Science, Technology and Innovation from Brazil for the Split Fellowship Program (process no. 200516/2022-1).

Data Availability

The data sets generated analyzed in this study are available from the corresponding author upon reasonable request. Access to the data is provided after anonymization to ensure ethical and legal sharing while preserving the confidentiality of the study participants.

Authors' Contributions

All authors contributed to the study protocol's design. GSN participated in the conceptualization of the study, data collection, formal analysis of the data, and writing of the original draft of the manuscript. ZSNR and RR participated in the conceptualization,

formal analysis, and writing of the original draft. JB contributed to the review and editing. All authors have reviewed and approved the manuscript for publication.

Conflicts of Interest

The authors declare a patent deposit on behalf of the Universidade Federal de Minas Gerais and Birthtech dispositivos para a saúde LTDA (BR 102020020982A2).

Multimedia Appendix 1
Database of clinical variables collected from each newborn.
[DOC File , 113 KB - biomedeng_v8i1e52468_app1.doc]

References

1. Mammel D, Kemp J. Prematurity, the diagnosis of bronchopulmonary dysplasia, and maturation of ventilatory control. Pediatr Pulmonol 2021 Nov;56(11):3533-3545. [doi: 10.1002/ppul.25519] [Medline: 34042316]
2. Azad M, Moyce B, Guillemette L, Pascoe C, Wicklow B, McGavock J, et al. Diabetes in pregnancy and lung health in offspring: developmental origins of respiratory disease. Paediatr Respir Rev 2017 Jan;21:19-26. [doi: 10.1016/j.prrv.2016.08.007] [Medline: 27665512]
3. Ornoy A, Reece EA, Pavlinkova G, Kappen C, Miller RK. Effect of maternal diabetes on the embryo, fetus, and children: congenital anomalies, genetic and epigenetic changes and developmental outcomes. Birth Defects Res C Embryo Today 2015 Mar;105(1):53-72. [doi: 10.1002/bdrc.21090] [Medline: 25783684]
4. Moore ES, Elnaggar AC, Wareham JA, Ramsey CJ, Sumners JE. Neonatal functional lung maturity relative to gestational age at delivery, fetal growth, and pregnancy characteristics in triplet births. J Matern Fetal Neonatal Med 2012 Jan;25(1):78-83. [doi: 10.3109/14767058.2011.613973] [Medline: 21854124]
5. Udagawa J, Yasuda A, Naito K, Otani H. Analysis of the harmonized growth pattern of fetal organs by multidimensional scaling and hierarchical clustering. Congenit Anom (Kyoto) 2010 Sep;50(3):175-185. [doi: 10.1111/j.1741-4520.2010.00284.x] [Medline: 20584035]
6. Visscher MO, Adam R, Brink S, Odio M. Newborn infant skin: physiology, development, and care. Clin Dermatol 2015;33(3):271-280. [doi: 10.1016/j.clindermatol.2014.12.003] [Medline: 25889127]
7. Trevisanuto D, Testoni D, Almeida MFB. Maintaining normothermia: Why and how? Semin Fetal Neonatal Med 2018 Oct;23(5):333-339. [doi: 10.1016/j.siny.2018.03.009] [Medline: 29599071]
8. Walani SR. Global burden of preterm birth. Int J Gynaecol Obstet 2020 Jul;150(1):31-33. [doi: 10.1002/ijgo.13195] [Medline: 32524596]
9. Rubarth LB, Quinn J. Respiratory Development and Respiratory Distress Syndrome. Neonatal Netw 2015;34(4):231-238. [doi: 10.1891/0730-0832.34.4.231] [Medline: 26802638]
10. Grenache DG, Gronowski AM. Fetal lung maturity. Clin Biochem 2006 Jan;39(1):1-10. [doi: 10.1016/j.clinbiochem.2005.10.008] [Medline: 16303123]
11. Ho J, Subramaniam P, Davis P. Continuous positive airway pressure (CPAP) for respiratory distress in preterm infants. Cochrane Database Syst Rev 2020 Oct 15;10(10):CD002271 [FREE Full text] [doi: 10.1002/14651858.CD002271.pub3] [Medline: 33058208]
12. Shankaran S, Bell EF, Laptook AR, Saha S, Newman NS, Kazzi SNJ, et al. Weaning of Moderately Preterm Infants from the Incubator to the Crib: A Randomized Clinical Trial. J Pediatr 2019 Jan;204(7):96-102.e4 [FREE Full text] [doi: 10.1016/j.jpeds.2018.08.079] [Medline: 30337189]
13. Taeusch HW, Wang NS, Avery ME. Studies on organ maturation: "skin age" as an indicator of "lung age" in fetal rabbits. Pediatrics 1972 Mar;49(3):400-405. [Medline: 5062268]
14. Patel AB, Bann CM, Kolhe CS, Lokangaka A, Tshefu A, Bauserman M, et al. The Global Network Socioeconomic Status Index as a predictor of stillbirths, perinatal mortality, and neonatal mortality in rural communities in low and lower middle income country sites of the Global Network for Women's and Children's Health Research. PLoS One 2022;17(8):e0272712 [FREE Full text] [doi: 10.1371/journal.pone.0272712] [Medline: 35972913]
15. Reis ZSN, Romanelli RMDC, Guimarães RN, Gaspar JDS, Neves GS, Vale MS, et al. Newborn Skin Maturity Medical Device Validation for Gestational Age Prediction: Clinical Trial. J Med Internet Res 2022 Sep 07;24(9):e38727 [FREE Full text] [doi: 10.2196/38727] [Medline: 36069805]
16. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
17. Reis ZSN, Guimarães RN, Rego MAS, Romanelli RMC, Gaspar JDS, Vitral GLN, et al. Prematurity detection evaluating interaction between the skin of the newborn and light: protocol for the preemie-test multicentre clinical trial in Brazilian hospitals to validate a new medical device. BMJ Open 2019 Mar 05;9(3):e027442 [FREE Full text] [doi: 10.1136/bmjopen-2018-027442] [Medline: 30842119]

18. Reis ZSN, Vitral GLN, Guimarães RN, Gaspar JDS, Colosimo E, Taunde S, et al. Premature or Small for Gestational Age Discrimination: International Multicenter Trial Protocol for Classification of the Low-Birth-Weight Newborn Through the Optical Properties of the Skin. JMIR Res Protoc 2020 Jul 14;9(7):e16477 [FREE Full text] [doi: 10.2196/16477] [Medline: 32673275]
19. Reis ZSN, Vitral GLN, Souza IMF, Rego MAS, Guimaraes RN. Newborn skin reflection: Proof of concept for a new approach for predicting gestational age at birth. A cross-sectional study. PLoS One 2017;12(9):e0184734 [FREE Full text] [doi: 10.1371/journal.pone.0184734] [Medline: 28931040]
20. Silva PC, Guimarães RN, Souza RG, Reis ZSN. A quantitative cross-sectional analysis of the melanin index in the skin of preterm newborns and its association with gestational age at birth. Skin Res Technol 2020 May;26(3):356-361 [FREE Full text] [doi: 10.1111/srt.12810] [Medline: 31763716]
21. Reis MAA, Romanelli RMC, Reis ZSN. Respiratory distress syndrome of the newborntransient tachypnea of the newborn diagnosis V. Protocols.io. URL: https://www.protocols.io/view/respiratory-distress-syndrome-of-the-newborn-and-t-q26g7bp33lwz/v2 [accessed 2023-12-13]
22. Chiou Y, Blume-Peytavi U. Stratum corneum maturation. A review of neonatal skin function. Skin Pharmacol Physiol 2004;17(2):57-66. [doi: 10.1159/000076015] [Medline: 14976382]
23. Sweet DG, Carnielli VP, Greisen G, Hallman M, Klebermass-Schrehof K, Ozek E, et al. European Consensus Guidelines on the Management of Respiratory Distress Syndrome: 2022 Update. Neonatology 2023;120(1):3-23 [FREE Full text] [doi: 10.1159/000528914] [Medline: 36863329]
24. Shukla VV, Eggleston B, Ambalavanan N, McClure EM, Mwenechanya M, Chomba E, et al. Predictive Modeling for Perinatal Mortality in Resource-Limited Settings. JAMA Netw Open 2020 Nov 02;3(11):e2026750 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.26750] [Medline: 33206194]
25. Li Q, Xu L, Zhang L. The Role and Clinical Value of Optimized Fetal Main Pulmonary Artery Doppler Parameters in the Diagnosis and Prognosis Monitoring of Neonatal Respiratory Distress Syndrome. Comput Math Methods Med 2022;2022:1682225 [FREE Full text] [doi: 10.1155/2022/1682225] [Medline: 35979041]
26. Fawcett K. Preventing admission hypothermia in very low birth weight neonates. Neonatal Netw 2014;33(3):143-149. [doi: 10.1891/0730-0832.33.3.143] [Medline: 24816875]
27. Reuter S, Moser C, Baack M. Respiratory distress in the newborn. Pediatr Rev 2014 Oct;35(10):417-28; quiz 429 [FREE Full text] [doi: 10.1542/pir.35-10-417] [Medline: 25274969]
28. Han T, Jin X, Yang J, Tang Y. Clinical Analysis of Fetal Lung Development Index and Pregnancy Outcome in Pregnant Women with Gestational Diabetes Mellitus with Satisfactory Blood Glucose Control. Contrast Media Mol Imaging 2022;2022:5777804 [FREE Full text] [doi: 10.1155/2022/5777804] [Medline: 36262988]
29. Dusingizimana T, Ramilan T, Weber JL, Iversen PO, Mugabowindekwe M, Ahishakiye J, et al. Predictors for achieving adequate antenatal care visits during pregnancy: a cross-sectional study in rural Northwest Rwanda. BMC Pregnancy Childbirth 2023 Jan 26;23(1):69 [FREE Full text] [doi: 10.1186/s12884-023-05384-0] [Medline: 36703102]
30. Cavallin F, Contin A, Alfeu N, Macmillian B, Seni AHA, Cebola BR, et al. Prognostic role of TOPS in ambulance-transferred neonates in a low-resource setting: a retrospective observational study. BMC Pregnancy Childbirth 2022 Sep 23;22(1):726 [FREE Full text] [doi: 10.1186/s12884-022-05060-9] [Medline: 36151540]

Abbreviations

ACTFM: antenatal corticosteroid therapy for fetal maturation
NICU: neonatal intensive care unit
RDS: respiratory distress syndrome

Edited by T Leung; submitted 05.09.23; peer-reviewed by N Akhundi, M Castro Fernández; comments to author 06.10.23; revised version received 26.10.23; accepted 27.10.23; published 25.12.23.

Please cite as:
Neves GS, Reis ZSN, Romanelli R, Batchelor J
Assessment of Skin Maturity by LED Light at Birth and Its Association With Lung Maturity: Clinical Trial Secondary Outcomes
JMIR Biomed Eng 2023;8:e52468
URL: https://biomedeng.jmir.org/2023/1/e52468
doi:10.2196/52468
PMID:38875690

Original Paper

Enhancing Ultrasound Image Quality Across Disease Domains: Application of Cycle-Consistent Generative Adversarial Network and Perceptual Loss

Shreeram Athreya[1], MS; Ashwath Radhachandran[2], BS; Vedrana Ivezić[3], BS; Vivek R Sant[4], MD; Corey W Arnold[1,2,3,5,6], PhD; William Speier[2,3,5], PhD

[1]Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, United States

[2]Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, United States

[3]Medical Informatics, University of California Los Angeles, Los Angeles, CA, United States

[4]Department of Surgery, The University of Texas Southwestern Medical Center, Dallas, TX, United States

[5]Department of Radiological Sciences, University of California Los Angeles, Los Angeles, CA, United States

[6]Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Shreeram Athreya, MS
Department of Electrical and Computer Engineering
University of California Los Angeles
924 Westwood Boulevard, Suite 600
Los Angeles, CA, 90024
United States
Phone: 1 4244206158
Email: shreeram@ucla.edu

Abstract

Background: Numerous studies have explored image processing techniques aimed at enhancing ultrasound images to narrow the performance gap between low-quality portable devices and high-end ultrasound equipment. These investigations often use registered image pairs created by modifying the same image through methods like down sampling or adding noise, rather than using separate images from different machines. Additionally, they rely on organ-specific features, limiting the models' generalizability across various imaging conditions and devices. The challenge remains to develop a universal framework capable of improving image quality across different devices and conditions, independent of registration or specific organ characteristics.

Objective: This study aims to develop a robust framework that enhances the quality of ultrasound images, particularly those captured with compact, portable devices, which are often constrained by low quality due to hardware limitations. The framework is designed to effectively process nonregistered ultrasound image pairs, a common challenge in medical imaging, across various clinical settings and device types. By addressing these challenges, the research seeks to provide a more generalized and adaptable solution that can be widely applied across diverse medical scenarios, improving the accessibility and quality of diagnostic imaging.

Methods: A retrospective analysis was conducted by using a cycle-consistent generative adversarial network (CycleGAN) framework enhanced with perceptual loss to improve the quality of ultrasound images, focusing on nonregistered image pairs from various organ systems. The perceptual loss was integrated to preserve anatomical integrity by comparing deep features extracted from pretrained neural networks. The model's performance was evaluated against corresponding high-resolution images, ensuring that the enhanced outputs closely mimic those from high-end ultrasound devices. The model was trained and validated using a publicly available, diverse dataset to ensure robustness and generalizability across different imaging scenarios.

Results: The advanced CycleGAN framework, enhanced with perceptual loss, significantly outperformed the previous state-of-the-art, stable CycleGAN, in multiple evaluation metrics. Specifically, our method achieved a structural similarity index of 0.2889 versus 0.2502 (P<.001), a peak signal-to-noise ratio of 15.8935 versus 14.9430 (P<.001), and a learned perceptual image patch similarity score of 0.4490 versus 0.5005 (P<.001). These results demonstrate the model's superior ability to enhance image quality while preserving critical anatomical details, thereby improving diagnostic usefulness.

Conclusions: This study presents a significant advancement in ultrasound imaging by leveraging a CycleGAN model enhanced with perceptual loss to bridge the quality gap between images from different devices. By processing nonregistered image pairs,

the model not only enhances visual quality but also ensures the preservation of essential anatomical structures, crucial for accurate diagnosis. This approach holds the potential to democratize high-quality ultrasound imaging, making it accessible through low-cost portable devices, thereby improving health care outcomes, particularly in resource-limited settings. Future research will focus on further validation and optimization for clinical use.

(JMIR Biomed Eng 2024;9:e58911) doi:10.2196/58911

KEYWORDS

generative networks; cycle generative adversarial network; image enhancement; perceptual loss; ultrasound scans; ultrasound images; imaging; machine learning; portable handheld devices

Introduction

Ultrasound imaging is crucial in medical diagnostics due to its noninvasive nature and high accuracy. It provides point-of-care assessments that have been increasingly adopted by health care professionals [1,2]. Historically, technology has been limited to large, expensive devices typically found in specialized medical settings. However, there has been a transformative shift toward the development and adoption of compact, handheld ultrasound devices [3,4]. These smaller devices promise to democratize access to medical imaging by making it more affordable and widely available. Yet, the miniaturization and cost-effectiveness often come at the expense of image quality, a trade-off primarily attributable to hardware constraints [5-7].

Machine learning algorithms have been explored to enhance low-quality images without the need for hardware improvements [8]. For instance, generative adversarial networks (GANs) [9] have been used to create high-quality reconstructions of ultrasound images and videos, providing a cost-efficient avenue for the enhancement of portable ultrasound devices [10-12]. The cycle-consistent generative adversarial network (CycleGAN) framework, which is particularly useful for image-to-image translations without requiring paired data, has become increasingly popular [13]. The technology has been applied across a spectrum of tasks including, style transfer [14], where the appearance of one image is transformed to match another style, and object transfiguration [13,15,16], which involves changing 1 object in an image into another while retaining the overall structure. In medical imaging, CycleGANs have been used in tasks such as pixel-wise translation in echocardiography [17]. CycleGANs have also been applied in cross-modality medical image translation such as converting computed tomography to magnetic resonance imaging [18]. The architecture has even found use in histopathology to standardize microscopy staining for more accurate diagnoses [19].

We hypothesize that the integration of computational algorithms, particularly CycleGAN, can mitigate the disparities in images acquired from different medical imaging devices. Traditional training approaches for these models artificially introduce corruption into medical images to create pixel-wise pairs [20-22]. However, these methods typically fail to encapsulate the different characteristics of images acquired using different devices. Acquiring paired images using different devices leads to technical issues as images are captured at different time instances with varying orientations, leading to structural changes that cannot be completely resolved using image registration.

In this work, we benchmark several key models that are highly relevant to our task of ultrasound image enhancement. Pix2Pix [4] uses conditional adversarial networks for paired image-to-image translation, making it effective for directly comparing low- and high-quality images. CycleGAN [5] enables unpaired image-to-image translation, which is crucial when paired datasets are not available. Registration GAN (RegGAN) [6] focuses on medical image translation by aligning structural content using a registration network, and multilevel structure-preserved GAN (MSPGAN) [7] introduces a multilevel structure-preserved GAN for domain adaptation in intravascular ultrasound analysis. However, the current state-of-the-art is the stability-enhanced CycleGAN [1], which specifically addresses domain transformation challenges in unpaired ultrasound images, making it particularly relevant and effective for our specific application.

Evaluation metrics play a critical role in assessing the effectiveness of image enhancement models. Commonly used metrics include structural similarity index (SSI), peak signal-to-noise ratio (PSNR), and locally normalized cross-correlation (LNCC) [5,10-12,23]. While these metrics are widely accepted, they primarily capture low-frequency information and may not adequately reflect true image quality, particularly in preserving high-frequency details, which are crucial for medical diagnostics. Models that perform well on these traditional metrics may produce visually appealing images but fail to retain essential high-frequency content, leading to a loss of critical diagnostic information and perceptual quality [24]. To address this limitation, we incorporate the learned perceptual image patch similarity (LPIPS) [24] metric in our evaluations. LPIPS is designed to capture perceptual differences that align more closely with human visual perception, ensuring that our method not only performs well quantitatively but also produces qualitatively superior images, preserving both low- and high-frequency details essential for accurate medical analysis.

To overcome these challenges, our approach leverages perceptual loss, which can eliminate the need for registration and more accurately relate images from disparate domains. Traditional loss functions used in CycleGAN can result in hallucinated features in the enhanced images [25]. By incorporating perceptual loss [24], more interpretable images are generated that are more robust to registration artifacts [26]. This method can enhance the reliability and consistency of images from handheld ultrasound devices to bridge the gap with expensive high-end systems for greater equity in access to health care.

Methods

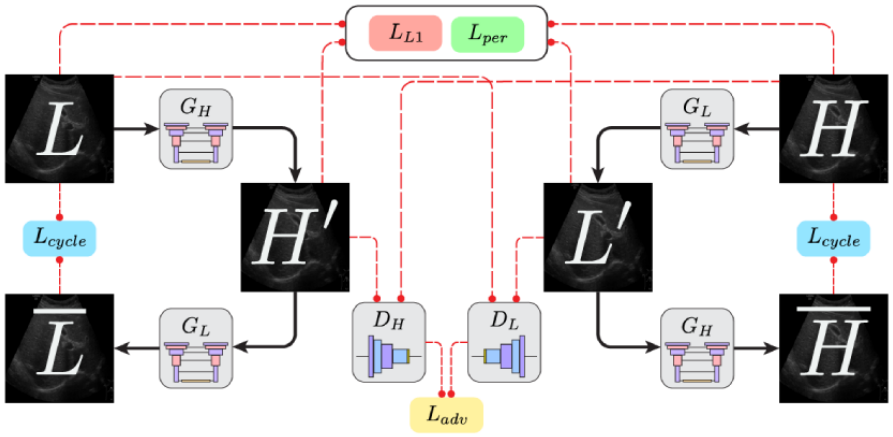
Study Design

In this study, we aim to address the challenge of enhancing ultrasound image quality, particularly for images captured by compact, portable devices that often suffer from lower quality due to hardware limitations. To achieve this, we used a CycleGAN framework enhanced with perceptual loss. This approach focuses on processing nonregistered image pairs from various organ systems, ensuring that the enhanced images retain anatomical integrity and closely mimic high-resolution outputs. Our method is designed to be robust, versatile, and applicable across diverse clinical settings.

Model Overview

Our framework for generating high-quality images is a modification of the CycleGAN architecture, designed to map between 2 distinct imaging domains. In ultrasound image enhancement, these domains correspond to low-quality (domain *L*) and high-quality (domain *H*) images. The model uses 2 generators, G_L and G_H, and 2 discriminators, D_L and D_H (Figure 1). Note that the generators G_L and G_H share the same model architecture. Similarly, the discriminators D_L and D_H share the same model architecture. The generator G_L is responsible for converting an image from domain H, which represents high-quality images, to domain L, characterized by low-quality images. Conversely, the generator G_H performs the opposite transformation, taking an image from domain L and converting it to align with domain H. This bidirectional transformation process is essential for the task of image enhancement, as it allows for the improvement of low-quality images by translating them into their high-quality counterparts. The discriminators aim to distinguish real images in their respective domains from those transformed by the generators. A unique feature of this approach is the cycle consistency loss [13], which plays a crucial role in image quality enhancement. This loss ensures that when an image is translated to the other domain and then reverted to its original domain, it closely resembles the original image. Specifically, for enhancing low-quality images to high-quality images, the cycle consistency loss maintains the integrity of the image content throughout the transformation process. This prevents the introduction of artifacts and ensures that the enhanced image retains the essential features of the original low-quality image, resulting in a high-quality output that remains true to the source. After training, the G_H generator is used to enhance images, maintaining essential structural attributes while improving clarity and resolution.

Figure 1. An overview of the cycle generative adversarial network model training and loss computation framework. The solid black arrows indicate the flow of data. The dashed red arrows indicate the flow of information for loss computation.



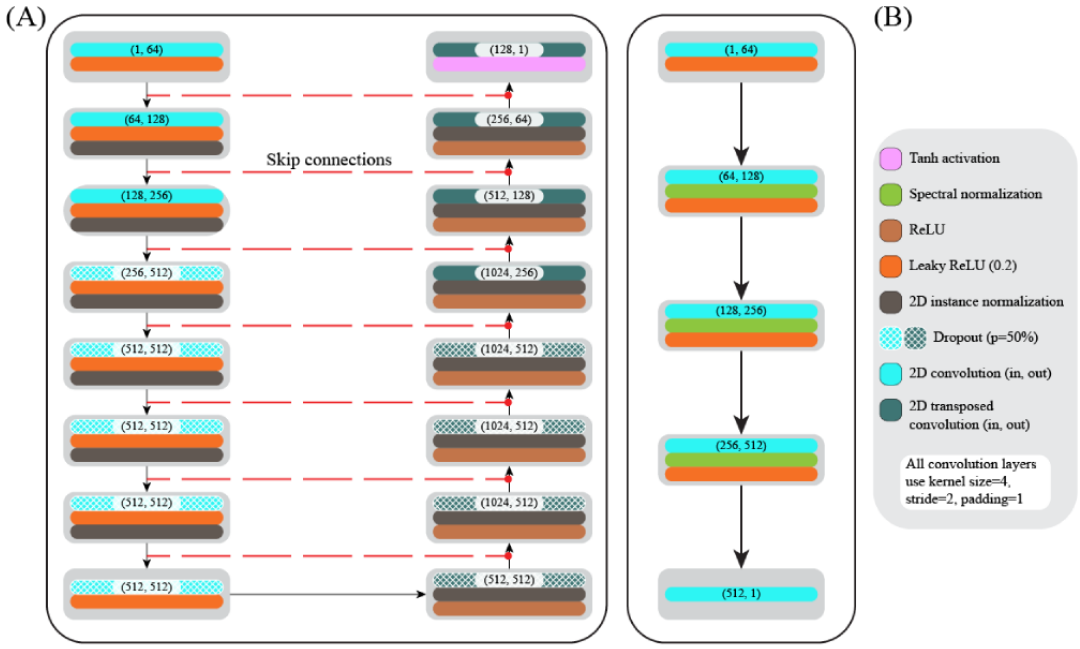
Model Description

GANs have seen transformative advancements, with CycleGAN [13] representing a significant milestone in facilitating unsupervised image-to-image translations. The GAN architecture comprises 2 primary modules: the generator and the discriminator.

The generator (Figure 2A) architecture is inspired by the generator used by Isola et al [27]. The generator network is structured as a UNet [28], divided into encoding and decoding phases, incorporating detailed mechanisms for both down-sampling and up-sampling the input data. The encoder initiates with a 64-channel 2D convolutional layer designed to capture broad contextual details. This phase uses multiple down-sampling layers, each comprising a convolutional layer with instance normalization and leaky rectified linear unit (ReLU) activation functions. The instance normalization layers stabilize the training process by normalizing the feature maps, while leaky ReLU activations introduce nonlinearity and mitigate the vanishing gradient problem. To enhance model robustness, dropout layers are included in deeper layers of the encoder. The down-sampling process reduces the spatial dimensions while increasing the depth, thereby emphasizing hierarchical feature extraction. In the decoding phase, the model uses transposed convolutional layers for up-sampling, which restores the spatial dimensions. Each up-sampling step involves skipping connections from the corresponding down-sampling layers, preserving detailed features from earlier stages. These layers also incorporate instance normalization and ReLU activations, where ReLU functions introduce nonlinearity, promoting sparse activations and efficient learning. The final layer uses a tanh activation function, scaling the output values to [–1, 1], suitable for image generation tasks. This design ensures effective image enhancement by maintaining high-quality feature extraction and reconstruction.

The discriminator (Figure 2B) distinguishes between real and generated images. We use spectral normalization [29] to ensure stability during training. Its architecture begins with a convolutional layer that compresses spatial information and expands depth followed by a leaky ReLU activation layer. Subsequent layers maintain the use of spectral normalization to ensure 1 – Lipschitz continuity. This constraint on the spectral norm of each layer's weights helps to balance the generator and discriminator during training. The PatchGAN [27,30] style discriminator output is a 30×30 grid with a depth of 1, which provides a spatial map indicating the likelihood of each region in the input image being real or generated. This final classification output allows for a more detailed evaluation of the image, helping to distinguish between authentic and synthetic content across different spatial locations.

Figure 2. The model architectures. The (A) generator and (B) discriminator model architectures. The figure legend lists the different layers in the models. ReLU: rectified linear unit.



Loss Function

Perceptual Loss

Conventional methodologies like mean squared error (MSE) and SSI rely on pixel-wise alignment which makes them unsuitable for nonregistered image pairs acquired using different devices. The LPIPS metric addresses these constraints by evaluating the perceptual similarity between images [24]. LPIPS leverages deep features extracted from pretrained convolutional networks, such as the visual geometry group network [31]. The LPIPS metric comparing images X and Y is given by the following equation.

where F_i and w_i denotes the feature maps and optimized weights from the i^{th} layer of the pretrained network. Deep feature maps are systematically extracted from every layer within the network, ensuring a comprehensive reflection of the multi-scale characteristics of human perceptual judgment. These features are then unified through linear combination, optimizing the weights to align with perceptual judgments assessed by human evaluators. The LPIPS metric consistently outranks traditional metrics, showcasing superior performance across an array of perceptual judgment tasks [26]. This loss is calculated between real images L and H, and those generated through the CycleGAN framework's generators as follows.

where H' represents $G_H(L)$ and L' represents $G_L(H)$.

Generator Loss

The generator's loss function is a linear combination of several distinct loss terms, each playing a pivotal role in optimizing image translation between the 2 domains. First, adversarial loss L_{adv} (D_H,D_L) induces the discriminators to perceive generated images as genuine, whether they are translated from low to high quality or vice versa.

$$L_{adv}(D_H,D_L) = MSE(1,D_H) + MSE(1,D_L)$$

Specifically, MSE calculates the discrepancy between the discriminator's predictions and an array of ones. The array of ones represents the target output for real images, indicating that the discriminator should classify these images as genuine. By comparing the discriminator's predictions to this ideal output, the MSE helps measure how far the generated images are from being perceived as real. These terms push the generator to produce images that can convince the discriminator they belong to the high-quality domain, thereby improving the realism and quality of the generated images. Using an array of ones ensures that the generator is continuously driven to reduce the difference between its output and real high-quality images, enhancing its performance over time. The cycle loss L_{cycle} (G_H,G_L) prevents

the loss of critical image features by ensuring that an image translated to the other domain and back yields the original image.

where  . and  . The $L1$ loss $L_{L1}(G_H,G_L)$ ensures that the generated images are closer to true high-quality images in an $L1$ distance sense.

Finally, the aggregate generator loss, $L(G_H,G_L,D_H,D_L)$, is computed by combining all individual loss terms weighed by their respective lambda constants.

By using this multifaceted loss function, the model ensures that the generators achieve high-quality image translations while preserving the intrinsic characteristics of the source domain.

Discriminator Loss

The discriminator loss function is designed to evaluate the authenticity of images, incorporating the principle of label smoothing to further enhance the model's generalizability. The discriminator is tasked with distinguishing between real images from the dataset and synthetic images generated by the corresponding generator.

For each domain, the discriminator computes scores for both real and synthetic images. Conventionally, discriminators are trained using hard labels, where real images are labeled as "1" and synthetic images as "0." However, hard labels can cause vulnerability to adversarial perturbations—small, intentional changes to the input that can deceive the model into making incorrect predictions—and lead to overconfidence, where the discriminator becomes excessively certain in its predictions. Label smoothing improves the generalization and robustness of neural networks by preventing overconfidence in predictions. Szegedy et al [32] demonstrated its effectiveness in reducing overfitting and enhancing performance in image classification. Similarly, Salimans et al [33] applied 1-sided label smoothing in the training of GANs, which helped stabilize training and improve the quality of generated images. These studies support the use of label smoothing as a strategy to mitigate the negative effects of hard labels. In our framework, if the mean scores of both real and synthetic images for the high-quality domain are less than 0.9, the label 1.0 is used. Otherwise, a smoothing factor of 0.9 is applied, meaning the real images are given a target value slightly less than 1, to prevent overconfidence and promote model robustness. The total discriminator loss, $L(D_H,D_L)$, is then computed by aggregating the individual MSE losses  . and  . for high and low-quality domains, respectively.

Implementation Details

All models were trained for 300 epochs with a batch size of 4 images. We used the Adam optimizer for model optimization, with a learning rate (3×10^{-4}) set for both the generators and the discriminators. A beta value of 0.9 for the first and 0.999 for the second moments were used in each optimizer. A learning rate scheduler reduced learning rates by half (γ=.5) every 100 epochs, to allow adaptability during training. Weights were assigned to each loss term: λ_{adv}=1 for adversarial loss, λ_{cycle}=10 for cycle-consistency loss, λ_{L1}=2 for $L1$ loss, and λ_{per}=10 for perceptual loss. The overall dataset was split with 70% for model training, 10% for model validation, and 20% for the hold-out test set. Gradient scaling was used to optimize the model's precision and speed. Code implementation will be made publicly available.

Evaluation

The synthetic high-quality images generated by the model are evaluated using 4 key metrics: SSI, LNCC, PSNR, and LPIPS. Each of these metrics provides a unique perspective on the fidelity and quality of the generated images compared to the ground truth.

- SSI evaluates the structural fidelity between the generated image H' and the actual high-quality image H. It considers 3 aspects: luminance, contrast, and structure. The SSI is computed as follows.

where μ_H and $\mu_{H'}$ are the mean intensities, and  are the variances, and  is the covariance between H' and H. The constants C_1 and C_2 are used to stabilize the division.

- LNCC measures the local similarity in intensity patterns between the generated and actual high-quality images. This metric is particularly sensitive to local differences in intensity, making it useful for detecting fine-grained discrepancies. LNCC is calculated by dividing the cross-correlation of local image patches by the product of their local SDs.

where  . and  . are the local means of the patches.

- PSNR quantifies the error signal strength between the generated and actual images, derived from the MSE. It is defined as follows.

where MAX_I is the maximum possible pixel value of the image, and $MSE\ (H, H')$ is the mean squared error between H' and H. A higher PSNR indicates that the generated image is closer to the high-quality reference.

- LPIPS [24] is a perceptual metric that compares deep features extracted from a neural network rather than directly

comparing pixel values. This approach better aligns with human perception of image similarity. LPIPS is computed by passing both the generated and actual images through a pretrained deep network and measuring the distance between their respective feature representations. Lower LPIPS values indicate higher perceptual similarity.

Together, these metrics provide a comprehensive evaluation of the generated images, which is particularly relevant for our image enhancement task, where unregistered low-high quality image pairs are compared. SSI and LNCC assess how well the structural and intensity patterns are preserved during enhancement, even when images are not perfectly aligned. PSNR quantifies the reduction in error relative to the original image, indicating overall fidelity. LPIPS, on the other hand, evaluates perceptual quality, ensuring that the enhancement appears natural and realistic to human observers, even in challenging scenarios with misaligned inputs.

Ethical Considerations

This study did not involve the collection of new data from human participants. The dataset used is publicly available and provided as part of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2023) USEnhance Challenge [34,35]. Therefore, no ethics review or approval was required for this study. As the study used publicly available data provided by the organizers of the challenge, informed consent specific to this research was not required. It is assumed that the original informed consent for data collection includes provisions for secondary analysis without requiring additional consent. All images used in this study are fully deidentified with no personal health information included. The dataset provided by the challenge organizers ensured anonymity, thus protecting the privacy and confidentiality of any potential human participants. There was no direct interaction with human participants in this study; hence, no compensation was provided. No images in this study or its supplementary materials allow for the identification of individual participants. All data are deidentified and anonymous, ensuring that no individual can be recognized from the images used.

Results

Overview

This study uses a dataset consisting of 2100 ultrasound images, including 1050 pairs of low- and high-quality images (Table 1 [34]). These images were collected from 131 patients with suspected thyroid tumors, carotid plaque, or breast cancer, along with healthy participants. During scans, volunteers were instructed to hold their breath for approximately 10 seconds to minimize deformation, and landmark points were noted for nonrigid registration to ensure the creation of accurate data pairs. This well-curated dataset provides a robust foundation for this study. This dataset was provided by the organizers of the MICCAI 2023 USEnhance Challenge [34]. Our baseline compares low-quality images directly to high-quality images without any enhancement or learning-based processing, serving as the starting point for evaluating the effectiveness of various models, including our approach. The models benchmarked in this study include Pix2Pix [27], MSPGAN [11], CycleGAN [13], RegGAN [23], and stable CycleGAN [12]. Among these, MSPGAN, RegGAN, and stable CycleGAN are the most recent advancements and are considered state-of-the-art for this task. To rigorously assess the improvements offered by our method, we computed the statistical significance of our results using the 1-sided Wilcoxon signed rank test.

Table 1. Dataset summary across different ultrasound devices and organs.

Organ	Low-end device	High-end device	Patients (n=131), n (%)	Ultrasound image pairs, n (%)		
				Training (n=840)	Testing (n=210)	Total (n=1050)
Thyroid	mSonics MU1	Toshiba Aplio 500	33 (25.2)	233 (27.7)	58 (27.6)	291 (27.7)
Carotid	SSUN	Toshiba Aplio 500	54 (41.2)	229 (27.3)	57 (27.1)	286 (27.2)
Abdomen	SSUN	General Electric LOGIQ E9	21 (16)	217 (25.8)	54 (25.7)	271 (25.8)
Breast	mSonics MU1	Aixplorer ultrasound system (SuperSonic Imaging SA)	23 (17.6)	161 (19.2)	41 (19.5)	202 (19.2)

Quantitative Results

In the evaluation of the SSI, our proposed method achieved a score of 0.2889 (Table 2), surpassing the reference low baseline (0.2363; P<.001), as well as CycleGAN (0.2622; P<.001) and stable CycleGAN (0.2502; P<.001). This places our method on par with the top-performing models like Pix2Pix (0.2862; PP>.99), MSPGAN (0.2796; P<.001), and RegGAN (0.2809; P<.001). Among the methods evaluated, stable CycleGAN exhibited the lowest SSI score, indicating the least effective structural preservation. Pix2Pix, on the other hand, performed slightly better than MSPGAN and RegGAN, highlighting its strength in maintaining structural details. For LNCC, our method recorded a score of 0.8454, which is significantly higher than the reference low baseline (0.7836; P<.001) and comparable to the scores achieved by MSPGAN (0.8535; P>.99) and Pix2Pix (0.8491; P>.99). While MSPGAN led in LNCC, the differences between the top performing methods are minimal, underscoring the similar performance levels across these models. Notably, CycleGAN and stable CycleGAN scored 0.8271 (P<.001) and 0.8145 (P<.001), respectively, showing lower but still competitive performance.

In terms of PSNR, the proposed method achieved a score of 15.8935, which is a marked improvement over the reference low baseline (14.2978; P<.001). Although Pix2Pix (16.3914; P>.99) and MSPGAN (16.2602; P>.99) reported higher PSNR

values, indicating lower overall error between the generated and high-quality images, the differences between these models and our approach are modest. RegGAN also performed well with a score of (16.2721; *P*>.99), while CycleGAN (14.9126; *P*<.001) and stable CycleGAN (14.9430; *P*<.001) had lower PSNR values, indicating higher error rates. Finally, for the LPIPS metric, our method demonstrated the best performance with a score of 0.4490, significantly lower than Pix2Pix (0.4664; *P*<.001), MSPGAN (0.4709; *P*<.001), and RegGAN (0.4855; *P*<.001). This indicates that our method produced images that were perceptually closer to high-quality outputs. CycleGAN and stable CycleGAN reported LPIPS scores of 0.4828 (*P*<.001) and 0.5005 (*P*<.001), respectively, with stable CycleGAN showing the least favorable performance among all models in terms of perceptual quality. Across these metrics, while certain models like Pix2Pix and MSPGAN excel in specific metrics such as LNCC and PSNR, our approach consistently delivers competitive performance, particularly in SSI and LPIPS, making it a robust framework for ultrasound image enhancement.

Table 2. Performance evaluation of models on the test set.

Model configurations	LNCC[a]↑[b]	SSI[c]↑	PSNR[d]↑	LPIPS[e]↓[f]
Reference low	0.7836[g]	0.2363[g]	14.2978[g]	0.5080[g]
Pix2Pix [27]	0.8491	0.2862	16.3914	0.4664[g]
MSPGAN[h] [11]	0.8535	0.2796[g]	16.2602	0.4709[g]
CycleGAN[i] [13]	0.8271[g]	0.2622[g]	14.9126[g]	0.4828[g]
RegGAN[j] [23]	0.8475	0.2809[g]	16.2721	0.4855[g]
Stable CycleGAN [12]	0.8145[g]	0.2502[g]	14.9430[g]	0.5005[g]
Proposed method	0.8454	0.2889	15.8935	0.4490

[a]LNCC: locally normalized cross-correlation.

[b]↑: higher scores are better.

[c]SSI: structural similarity index.

[d]PSNR: peak signal-to-noise ratio.

[e]LPIPS: learned perceptual image patch similarity

[f]↓: lower scores are better.

[g]Statistically significant improvement.

[h]MSPGAN: multilevel structure-preserved generative adversarial network.

[i]CycleGAN: cycle-consistent generative adversarial network.

[j]RegGAN: registration generative adversarial network.

Qualitative Results

Qualitative analysis further illustrates the differences in the generated images across the models. As shown in Figures 3 and 4, methods such as Pix2Pix, MSPGAN, and RegGAN, despite their higher scores in SSI, LNCC, and PSNR, often produce images that lack anatomical detail and introduce distortions that may affect clinical interpretation. In contrast, methods like CycleGAN, Stable CycleGAN, and our proposed approach maintain the integrity of anatomical structures, ensuring that the generated images closely resemble the original high-quality images. Our approach is particularly effective in preventing the loss of critical diagnostic information, which is essential for accurate medical assessments. While quantitative metrics provide a useful evaluation framework, the qualitative results underscore the importance of preserving anatomical integrity, an area where our method excels.

Figure 3. A comparative visualization of ultrasound scans from the test set, showcasing the performance of different enhancement frameworks on the same high-low quality image pair. Each subfigure corresponds to a different model's output, allowing for a direct comparison of the anatomical preservation and image quality achieved by each approach. (A) reference low, (B) Pix2Pix [27], (C) MSPGAN [11], (D) RegGAN [23], (E) reference high, (F) CycleGAN [13], (G) stable CycleGAN [12], and (H) proposed method. CycleGAN: cycle-consistent generative adversarial network; MSPGAN: multilevel structure-preserved generative adversarial network; RegGAN: registration generative adversarial network.

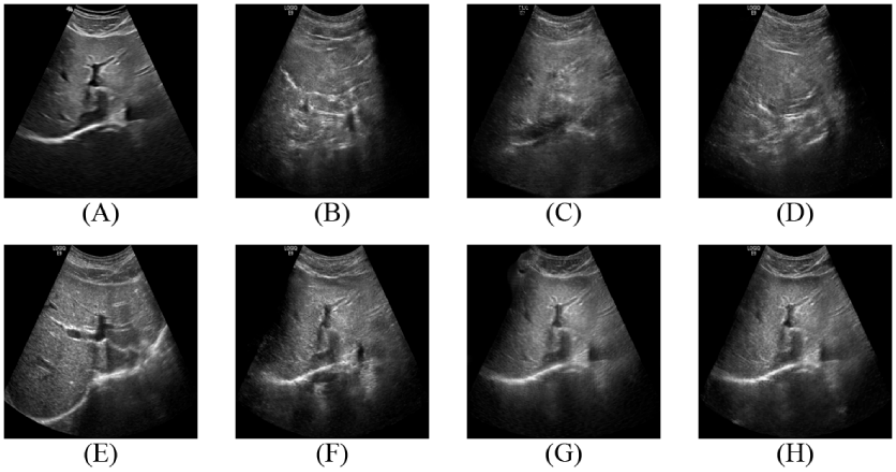
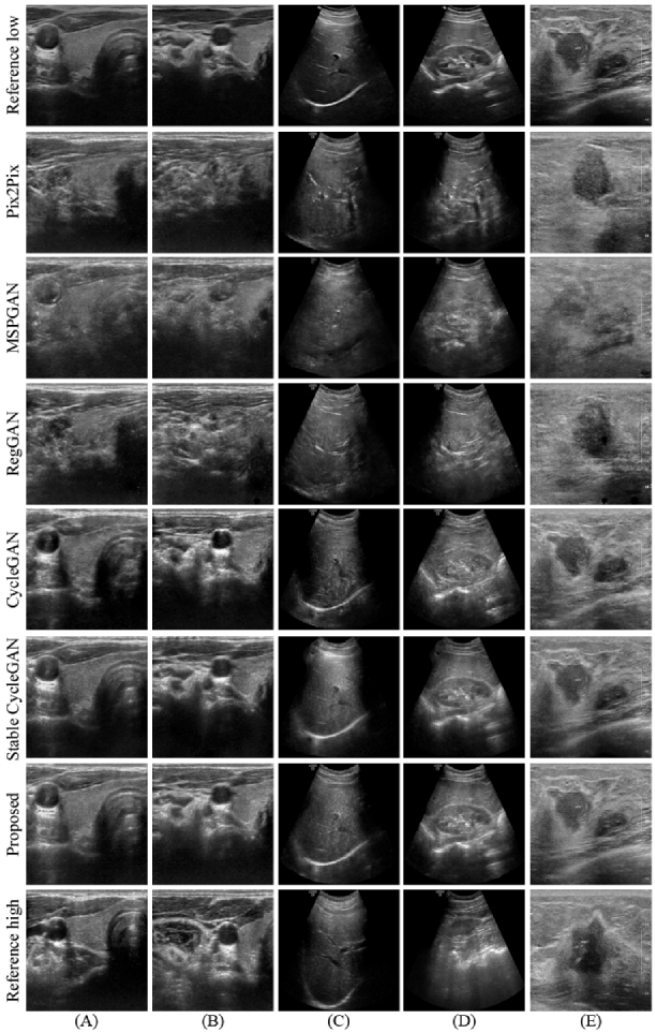


Figure 4. A comparative visualization of ultrasound scans from the test set, showcasing the performance of different enhancement frameworks on the same high-low quality image pairs. (A) Thyroid, (B) carotid, (C) liver, (D) kidney, and (E) breast. CycleGAN: cycle-consistent generative adversarial network; MSPGAN: multilevel structure-preserved generative adversarial network; RegGAN: registration generative adversarial network.



Discussion

In this study, we developed and evaluated an advanced CycleGAN framework enhanced with perceptual loss to address the challenge of varying image quality in ultrasound imaging across different devices. Our primary objective was to improve the quality of low-resolution ultrasound images captured by portable devices while preserving anatomical integrity, which is critical for accurate clinical diagnostics. The results from our evaluation demonstrated that the integration of perceptual loss enhanced the quality of the generated images, achieving strong performance in key metrics such as SSI and LPIPS, though slightly lower in LNCC and PSNR compared to other models. These outcomes suggest that our approach represents a significant step toward bridging the gap between low- and high-quality ultrasound images, making it particularly beneficial for portable, handheld devices that often struggle with image quality due to hardware limitations.

The use of perceptual loss in our model allowed for a more direct and meaningful comparison between low- and high-quality images, which contrasts with previous studies that treated these domains as independent [12]. By leveraging paired images from different devices, our model was able to learn the nuances of quality differences in a manner that closely mirrors real-world clinical scenarios. This pairing led to significant improvements in metrics such as SSI and LPIPS, indicating that our model preserves structural fidelity and local intensity patterns more effectively than current state-of-the-art approaches. However, it is important to note that while some models, such as Pix2Pix, MSPGAN, and RegGAN, achieve high scores in SSI, LNCC, and PSNR, they often do so at the expense of anatomical integrity. These models tend to remove or alter critical anatomical structures, leading to a loss of valuable diagnostic information. In contrast, our approach retains the anatomical content while producing comparable performance in these metrics and outperforming all other models in LPIPS, which measures perceptual quality. This balance between maintaining anatomical fidelity and achieving high image quality is a significant strength of our method, making it more suitable for clinical applications where accuracy is paramount.

Despite these promising results, there are some limitations to our approach that need to be addressed in future work. The reliance on perceptual loss, while beneficial for maintaining image fidelity, introduces additional computational complexity, leading to longer training times. This requirement could be a limitation in scenarios where computational resources are limited or rapid model deployment is necessary. Additionally, while our model has demonstrated strong performance across a well-curated dataset, the findings need to be validated through extensive real-world applications across diverse datasets and imaging conditions to ensure robustness and generalizability. Furthermore, the current model is designed to work across various organ systems and diseases, but future research could explore the development of more specialized models tailored to specific clinical contexts, potentially optimizing performance for targeted diagnostic tasks.

To conclude, this work introduced an advanced CycleGAN-based framework that effectively enhances ultrasound image quality across devices by using perceptual loss to train on paired images. Our findings demonstrate the feasibility of bridging the image quality gap between low- and high-quality ultrasound images, thereby improving the accessibility and equity of high-quality diagnostic imaging. As we move forward, it will be crucial to conduct clinical validation of this approach across a wide range of medical scenarios and explore its application to other imaging modalities. This result could pave the way for integrating our model into routine clinical practice, ultimately enhancing diagnostic accuracy and improving patient outcomes. By making high-quality imaging more accessible, particularly through portable ultrasound devices, our approach holds the potential to significantly impact health care delivery and patient care on a global scale.

Acknowledgments

The authors thank the editor and the reviewers for their valuable input, suggestions, and comments. The dataset used for this study was provided by the organizers of the MICCAI (Medical Image Computing and Computer Assisted Intervention) 2023 USEnhance Challenge [34]. This work was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award R21EB030691 and a University of California Los Angeles Radiology Exploratory Research Grant.

Data Availability

The datasets generated during and/or analyzed during this study are available in the MICCAI (Medical Image Computing and Computer Assisted Intervention) 2023 USEnhance Challenge repository [34].

Authors' Contributions

SA, AR, VI, VS, CA, and WS contributed to the overall conception, design, oversight, and guidance of the work. SA contributed to the CycleGAN (cycle-consistent generative adversarial network) framework development, training, and further analyses. SA drafted the initial manuscript. All authors contributed to the critical revision of the manuscript, approved the final draft for submission and publication, and agreed to be accountable for the work.

Conflicts of Interest

None declared.

References

1. Moore CL, Copel JA. Point-of-care ultrasonography. N Engl J Med 2011;364(8):749-757. [doi: 10.1056/NEJMra0909487] [Medline: 21345104]
2. Martocchia A, Bentivegna E, Sergi D, Luciani M, Barlattani M, Notarangelo MF, et al. The point-of-care ultrasound (POCUS) by the handheld ultrasound devices (HUDs) in the COVID-19 scenario: a review of the literature. SN Compr Clin Med 2023;5(1):1 [FREE Full text] [doi: 10.1007/s42399-022-01316-9] [Medline: 36407770]
3. Le PT, Voigt L, Nathanson R, Maw AM, Johnson G, Dancel R, et al. Comparison of four handheld point-of-care ultrasound devices by expert users. Ultrasound J 2022;14(1):27. [doi: 10.1186/s13089-022-00274-6] [Medline: 35796842]
4. Laastad Sørensen M, Oterhals K, Pönitz V, Morken IM. Point-of-care examinations using handheld ultrasound devices performed by intensive care nurses in a cardiac intensive care unit. Eur J Cardiovasc Nurs 2023;22(5):482-488. [doi: 10.1093/eurjcn/zvac089] [Medline: 36149811]
5. Zhou Z, Guo Y, Wang Y. Handheld ultrasound video high-quality reconstruction using a low-rank representation multipathway generative adversarial network. IEEE Trans Neural Netw Learning Syst 2021;32(2):575-588. [doi: 10.1109/tnnls.2020.3025380]
6. van der Pol HGA, van Karnenbeek LM, Wijkhuizen M, Geldof F, Dashtbozorg B. Deep learning for point-of-care ultrasound image quality enhancement: a review. Appl Sci 2024;14(16):7132. [doi: 10.3390/app14167132]
7. Loizou C, Pattichis SC, Istepanian R, Pantziaris M, Kyriakou E, Tyllis T, et al. Ultrasound image quality evaluation. 2003 Presented at: 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003; April 24-26, 2003; Birmingham, United Kingdom. [doi: 10.1109/itab.2003.1222492]
8. Chen C, Raymond C, Speier W, Jin X, Cloughesy TF, Enzmann D, et al. Synthesizing MR image contrast enhancement using 3D high-resolution ConvNets. IEEE Trans. Biomed. Eng 2023;70(2):401-412. [doi: 10.1109/tbme.2022.3192309]
9. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020;63(11):139-144 [FREE Full text] [doi: 10.1145/3422622]
10. Zhou Z, Wang Y, Guo Y, Qi Y, Yu J. Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network. IEEE Trans Biomed Eng 2020;67(1):298-311. [doi: 10.1109/tbme.2019.2912986]
11. Xia M, Yang H, Qu Y, Guo Y, Zhou G, Zhang F, et al. Multilevel structure-preserved GAN for domain adaptation in intravascular ultrasound analysis. Med Image Anal 2022;82:102614. [doi: 10.1016/j.media.2022.102614] [Medline: 36115099]
12. Huang L, Zhou Z, Guo Y, Wang Y. A stability-enhanced CycleGAN for effective domain transformation of unpaired ultrasound images. Biomed Signal Process Control 2022;77:103831. [doi: 10.1016/j.bspc.2022.103831]
13. Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 Presented at: IEEE International Conference on Computer Vision (ICCV); October 22-29, 2017; Venice, Italy. [doi: 10.1109/iccv.2017.244]
14. Chang H, Lu J, Yu F, Finkelstein A. PairedCycleGAN: asymmetric style transfer for applying and removing makeup. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00012]
15. Ye Z, Lyu F, Li L, Sun Y, Fu Q, Hu F. Unsupervised object transfiguration with attention. Cogn Comput 2019;11(6):869-878. [doi: 10.1007/s12559-019-09633-3]
16. Chen X, Xu C, Yang X, Tao D. Attention-Gan for object transfiguration in wild images. 2018 Presented at: European Conference on Computer Vision 2018 (ECCV 2018); September 8-14, 2018; Munich, Germany p. 164. [doi: 10.1007/978-3-030-01216-8_11]
17. Teng L, Fu Z, Yao Y. Interactive translation in echocardiography training system with enhanced cycle-GAN. IEEE Access 2020;8:106147-106156. [doi: 10.1109/access.2020.3000666]
18. Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00963]
19. Gadermayr M, Gupta L, Appel V, Boor P, Klinkhammer BM, Merhof D. Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: a study on kidney histology. IEEE Trans Med Imaging 2019;38(10):2293-2302. [doi: 10.1109/tmi.2019.2899364]
20. Geng M, Deng Y, Zhao Q, Xie Q, Zeng D, Zeng D, et al. Unsupervised/semi-supervised deep learning for low-dose CT enhancement. ArXiv Preprint posted online on August 8, 2018. [doi: 10.48550/ARXIV.1808.02603]
21. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, et al. Low-dose CT via convolutional neural network. Biomed Opt Express 2017;8(2):679. [doi: 10.1364/boe.8.000679]

22. Eun D, Jang R, Ha WS, Lee H, Jung SC, Kim N. Deep-learning-based image quality enhancement of compressed sensing magnetic resonance imaging of vessel wall: comparison of self-supervised and unsupervised approaches. Sci Rep 2020 Aug 18;10(1):13950 [FREE Full text] [doi: 10.1038/s41598-020-69932-w] [Medline: 32811848]
23. Kong L, Lian C, Huang D, Li Z, Hu Y, Zhou Q. Breaking the dilemma of medical image-to-image translation. 2021 Presented at: NIPS'21: 35th International Conference on Neural Information Processing Systems; December 6-14, 2021; Virtual-only conference URL: https://openreview.net/forum?id=C0GmZH2RnVR
24. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The Unreasonable effectiveness of deep features as a perceptual metric. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00068]
25. Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. 2018 Presented at: 21st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018); September 16-20, 2018; Granada, Spain p. 529-536. [doi: 10.1007/978-3-030-00928-1_60]
26. Armanious K, Jiang C, Fischer M, Küstner T, Hepp T, Nikolaou K, et al. MedGAN: medical image translation using GANs. Comput Med Imaging Graph 2020;79:101684. [doi: 10.1016/j.compmedimag.2019.101684] [Medline: 31812132]
27. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI. [doi: 10.1109/cvpr.2017.632]
28. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. 2015 Presented at: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015); October 5-9, 2015; Munich, Germany p. 234-241. [doi: 10.1007/978-3-319-24574-4_28]
29. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. 2018 Presented at: 6th International Conference on Learning Representations (ICLR 2018); April 30-May 3, 2018; Vancouver, BC URL: https://openreview.net/forum?id=B1QRgziT-
30. Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. Cham: Springer; 2016 Presented at: European Conference on Computer Vision 2016 (ECCV 2016); October 11-14, 2016; Amsterdam, The Netherlands p. 702-716. [doi: 10.1007/978-3-319-46487-9_43]
31. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv Preprint posted online on September 4, 2014 [FREE Full text]
32. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.308]
33. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training Gans. Cambridge, MA: Adv Neural Inf Process Syst; 2016 Presented at: 30th International Conference on Neural Information Processing Systems (NIPS'16); December 5-10, 2016; Barcelona, Spain.
34. Guo Y, Shichong Z, Shi J, Wang Y. Ultrasound image enhancement challenge 2023. Zenodo. URL: https://zenodo.org/records/7841250 [accessed 2023-04-08]
35. USenhance 2023: ultrasound image enhancement challenge (multiple organs). Grand Challenge. URL: https://ultrasoundenhance2023.grand-challenge.org/datasets/ [accessed 2023-10-12]

Abbreviations

CycleGAN: cycle-consistent generative adversarial network
GAN: generative adversarial network
LNCC: locally normalized cross-correlation
LPIPS: learned perceptual image patch similarity
MICCAI: Medical Image Computing and Computer Assisted Intervention
MSE: mean squared error
MSPGAN: multilevel structure-preserved generative adversarial network
PSNR: peak signal-to-noise ratio
RegGAN: registration generative adversarial network
ReLU: rectified linear unit
SSI: structural similarity index

Edited by S Rizvi; submitted 27.03.24; peer-reviewed by N Puzhakkal, J Peng, BS Shin; comments to author 15.06.24; revised version received 24.08.24; accepted 13.09.24; published 17.12.24.

Please cite as:
Athreya S, Radhachandran A, Ivezić V, Sant VR, Arnold CW, Speier W
Enhancing Ultrasound Image Quality Across Disease Domains: Application of Cycle-Consistent Generative Adversarial Network and Perceptual Loss
JMIR Biomed Eng 2024;9:e58911
URL: https://biomedeng.jmir.org/2024/1/e58911
doi:10.2196/58911
PMID:

Original Paper

The Variability of Lumbar Sequential Motion Patterns: Observational Study

Inge Caelers[1,2,3*], MD; Toon Boselie[1,3*], MD, PhD; Wouter van Hemert[1*], MD, PhD; Kim Rijkers[1,3*], MD, PhD; Rob De Bie[4*], PhD, PT; Henk van Santbrink[1,2,3*], MD, PhD

[1]Department of Neurosurgery, Zuyderland Medical Center, Heerlen, Netherlands

[2]Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[3]Department of Neurosurgery, Maastricht University Medical Center +, Maastricht, Netherlands

[4]Department of Epidemiology, Maastricht University, Maastricht, Netherlands

[*]all authors contributed equally

Corresponding Author:
Inge Caelers, MD
Department of Neurosurgery, Zuyderland Medical Center
Henri Dunantstraat 5
Heerlen, 6419 PC
Netherlands
Phone: 31 652593545
Email: inge.caelers@mumc.nl

Abstract

Background: Physiological motion of the lumbar spine is a topic of interest for musculoskeletal health care professionals since abnormal motion is believed to be related to lumbar complaints. Many researchers have described ranges of motion for the lumbar spine, but only few have mentioned specific motion patterns of each individual segment during flexion and extension, mostly comprising the sequence of segmental initiation in sagittal rotation. However, an adequate definition of physiological motion is still lacking. For the lower cervical spine, a consistent pattern of segmental contributions in a flexion-extension movement in young healthy individuals was described, resulting in a definition of physiological motion of the cervical spine.

Objective: This study aimed to define the lumbar spines' physiological motion pattern by determining the sequence of segmental contribution in sagittal rotation of each vertebra during maximum flexion and extension in healthy male participants.

Methods: Cinematographic recordings were performed twice in 11 healthy male participants, aged 18-25 years, without a history of spine problems, with a 2-week interval (time point T1 and T2). Image recognition software was used to identify specific patterns in the sequence of segmental contributions per individual by plotting segmental rotation of each individual segment against the cumulative rotation of segments L1 to S1. Intraindividual variability was determined by testing T1 against T2. Intraclass correlation coefficients were tested by reevaluation of 30 intervertebral sequences by a second researcher.

Results: No consistent pattern was found when studying the graphs of the cinematographic recordings during flexion. A much more consistent pattern was found during extension, especially in the last phase. It consisted of a peak in rotation in L3L4, followed by a peak in L2L3, and finally, in L1L2. This pattern was present in 71% (15/21) of all recordings; 64% (7/11) of the participants had a consistent pattern at both time points. Sequence of segmental contribution was less consistent in the lumbar spine than the cervical spine, possibly caused by differences in facet orientation, intervertebral discs, overprojection of the pelvis, and muscle recruitment.

Conclusions: In 64% (7/11) of the recordings, a consistent motion pattern was found in the upper lumbar spine during the last phase of extension in asymptomatic young male participants. Physiological motion of the lumbar spine is a broad concept, influenced by multiple factors, which cannot be captured in a firm definition yet.

Trial Registration: ClinicalTrials.gov NCT03737227; https://clinicaltrials.gov/ct2/show/NCT03737227

International Registered Report Identifier (IRRID): RR2-10.2196/14741

(JMIR Biomed Eng 2023;8:e41906) doi:10.2196/41906

KEYWORDS

lumbar spine; cinematographic recordings; sequence; motion pattern; flexion; extension; rotation; physiological; musculoskeletal; motion; spine; upper lumbar; observational study; physiological motion

Introduction

Physiological motion of the lumbar spine is of interest for musculoskeletal health care professionals. Motion of the lumbar spine is dependent on multiple structures, for example facet joint orientation, spinal-pelvic relations, intervertebral disc loading, and muscle recruitment. Although the concept of physiological motion is used in many instances, a proper definition is still lacking. Over the last 90 years, several attempts to define physiological motion have been made. In 1931, Dittmar et al [1] were the first to use sagittal radiographs to analyze the normal range of flexion and extension for the lumbar spine. Subsequently, more motion research followed using other techniques including computed tomography and magnetic resonance–based 3D imaging [2-4]. Based on these data, segmental ranges of motion with a high intra, and interindividual variability were described [5,6]. For this reason, researchers started to investigate sequences, like sequence of segmental initiation of motion. Studies that report sequence of segmental initiation of motion in flexion and extension also showed variable results. The lack of consistent segmental ranges of motion or sequence hampers the definition of physiological motion of the lumbar spine [7-17].

Our research group described a consistent sequence of segmental contribution in the lower cervical spine during extension using sagittal cinematographic recordings [18]. This research was used to create a definition of physiological motion in young healthy individuals without spinal complaints. To our knowledge, similar analysis of the sequence of segmental contribution for the lumbar spine has not been carried out previously.

This study aimed to analyze the sequence of segmental contribution of L1 to S1 in sagittal rotation during flexion and extension in individual participants. A consistent pattern of segmental contribution in asymptomatic participants could be seen as a definition of psychological motion. In the future, this pattern could be used to investigate potential abnormal motion in lumbar conditions. It might be possible to better diagnose instability and the impact of it on lumbar spine motion. Furthermore, we can determine if differences in motion lead to back pain and can be resolved by physiotherapy.

Methods

Ethics Approval

The study was approved by the Medical Research Ethics Committee of Zuyderland Hospital and Zuyd University of Applied Sciences, the Netherlands (METCZ20180094).

Participant Inclusion

The study protocol was published [19]. After approval, this study included men, aged between 18 and 25 years, with a BMI <25 kg/m^2, with no medical history of spine problems, and able to perform maximum lumbar flexion and extension without complaints. No medical history of spine problems was defined as no visits to a doctor or physical therapist for spine complaints, no former spine surgery, total scores of Oswestry Disability Index and Visual Analogue Scale for back pain of zero, and a Kellgrens' classification of 0-1 in levels L4L5 and L5S1 on cinematographic recordings evaluated by 2 spine surgeons (TB, HvS, and WvH) [20-22]. Female participants were excluded to protect their ovaries from direct radiation exposure. Potential participants were excluded if x-rays of the abdomen, pelvis, hip, lumbar, or sacral spine were taken in the previous year or in cases of active spinal infection, immature bone, lumbar tumor, previous lumbar radiotherapy, congenital lumbar spine abnormality, or planned pregnancy of the participants' partner in the coming year. Sample size, based on previous studies, was set on 11 participants [13,14,23].

Informed consent was acquired from all participants. Radiological data were stored along with the number of participants and recordings. Handling of personal data will comply with the guidelines of the Dutch Personal Data Protection Act.

Study Procedures

Flexion and extension cinematographic recordings were acquired twice for each participant during afternoons and evenings. An interval of 2 weeks was maintained to determine reproducibility and consistency of the sequence between 2 time points (T1 and T2) [18,24]. Cinematographic recordings were made from a lateral perspective to obtain sagittal images, using the Philips Allura Xper FD20 x-ray system. The following settings were used: frames of 1024×1024 pixels, 7.5 frames per second, tube voltage of 75-90 kV, filter of 0.9 mm copper + 1 mm aluminum, and a detector distance of 48 cm. The total radiation dose for participants was categorized in category 2A, using the Neurocritical Care Society guidelines on risks of radiation dose (0.1-1.0 mSv) [25]. During cinematographic recordings, participants were seated in a customized wooden chair, designed to keep the pelvis in a fixed position (Figure 1). A 3-point fixation was located on the anterior superior iliac spine, posterior inferior iliac spine, and the upper legs, which could be adjusted to the participants' physique. Participants were asked to remove clothes that could disturb the cinematographic recordings. From a neutral seating position with the knees in 90 degrees flexion, participants were asked to perform maximum extension, followed by maximum flexion, and then a return to maximum extension in 14 seconds, using a metronome. Maximum flexion and extension was determined as the maximum achievable position of the participant and practiced before the final cinematographic recordings. During the active motion task, arms were crossed in front of the chest (Figure 1). This duration was chosen based on the pulse frequency of the image technique (7.5 pulses per second) and the number of necessary images (104 images) for image recognition.

Figure 1. Customized wooden chair with 3-point fixation of the pelvis. The 3-point fixation is located on the anterior superior iliac spine, posterior inferior iliac spine, and the upper legs.

Radiological Data Processing and Analysis

For this research, we have previously developed custom software that uses image recognition algorithms to track vertebrae during flexion and extension [26]. The software follows bony structures within user-defined template areas throughout all frames, using a best-fit principle to match normalized gradient field images. To define these template areas, the user draws polygons around all vertebrae on the median frame of the recording [26]. After the software has completed tracking these structures, they can be manually evaluated. Corrections can be made if necessary. Finally, graphs are made for both flexion and extension cinematographic recordings for each individual participant to identify specific patterns in the sequence of segmental contributions. Segmental rotation of each individual segment (L1 to S1) between each pair of successive frames was plotted against the cumulative rotation in segments L1 to S1 together. A more detailed description of the image recognition software can be found in a previously published study [26]. Analyses were first performed for T1 and tested against T2. Time spent on radiological data processing and analysis was 2 to 3 days per cinematographic recording. Analyses were performed by

researcher IC, with reevaluation of 30 intervertebral sequences by a second researcher (TB) to determine reproducibility, using a two-way mixed intraclass correlation coefficient (ICC). An ICC above 0.60 was considered adequate. A consistent motion pattern was defined as a similar pattern shown in at least 80% (8/10) of the cinematographic recordings in 2 time points. This was comparable with the results of the cervical spine [18].

Results

A total of 11 participants were recruited and included, all undergoing 2 cinematographic recordings. This resulted in a total of 22 recordings, of which 1 (P1-recording 1) was excluded from analyses, since L1 could not be followed in the field of view. No consistent pattern was found when studying the graphs of the cinematographic recordings during flexion (Multimedia Appendix 1). During extension, segments L4L5 and L5S1 showed an inconsistent pattern (Multimedia Appendix 2). Leaving L4L5 and L5S1 out of the analyses, a much more consistent pattern on the sequence of segmental contribution was found, especially in the last phase of the extension motion. It consisted of a peak in rotation in L3L4, followed by a peak in L2L3, and finally, in L1L2 (Figure 2; Multimedia Appendix 3). Only the sequence of the peaks was important, not the height of the peaks itself, since a peak represents the largest contribution of a specific segment at a specific point in the total motion despite the height. As discussed in the study of Boselie et al [18], peaks with a rotation lower than 0.3 were deemed to fall within the measurement error and were not taken into consideration. In total, 71% (15/21) of extension graphs showed the abovementioned sequence, which represents 80% (8/10) at T1 and 64% (7/11) at T2 (Multimedia Appendix 3). At both time points, P5 and P7 did not show a consistent motion sequence with different motion patterns at each time point. P9 only showed a consistent motion sequence in T1. ICC was determined for each segment in 5 cinematographic recordings (Table 1).

Figure 2. Sagittal rotation in segments in the upper lumbar spine (segments L1 to L4) during extension of the lumbar spine in healthy young male participants (P2-T1). On the y-axis, the rotation is shown in degrees between successive frames. On the x-axis, cumulative degrees of extension in block L1 to L4 are shown. Peaks of the graphs per segments (L1L2, L2L3, and L3L4) depict maximum contribution of the segment in a specific phase of the extension. At the last phase of the extension, the L3L4 peak was followed by an L2L3 peak and finally the L1L2 peak. Each series of values undergoes smoothing by means of a low-pass Gaussian digital filter.

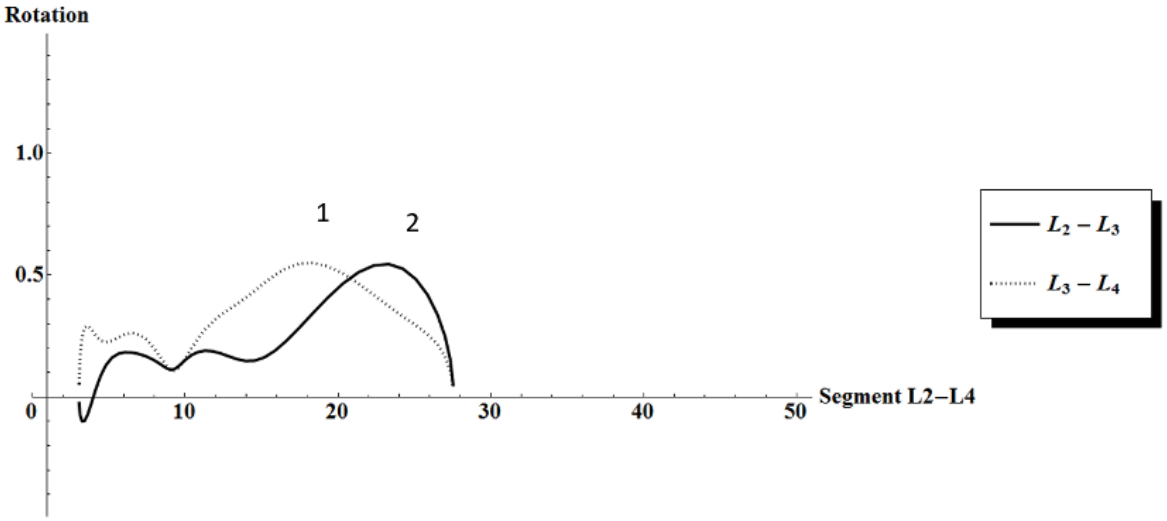


Table 1. Intraclass correlation coefficients (ICC) per segment of 5 randomly chosen cinematographic recordings. An ICC below 0.60 is determined as inadequate and indicated in italics.

Segments	Cinematographic recordings					
	2-1	3-2	4-1	8-1	11-2	Mean
L1L2	0.601	0.858	*0.389*	0.662	*0.550*	0.612
L2L3	0.799	0.615	0.808	0.662	0.722	0.721
L3L4	*0.577*	0.819	0.886	0.932	0.694	0.782
L4L5	0.691	*0.437*	0.876	0.917	*0.553*	0.695
L5S1	0.763	*0.258*	0.750	0.902	*0.268*	*0.588*

Discussion

Principal Findings

The aim of this study was to ascertain the sequence of segmental contribution and to possibly understand physiological motion in sagittal rotation during maximum flexion and extension of the lumbar spine in asymptomatic male participants. Results showed a consistent pattern in 71% (15/21) of the recordings during the last phase of the extension with a peak in rotation in L3L4, followed by a peak in L2L3, and finally, in L1L2. However, this pattern was consistent in only 64% (7/11) of the recordings over the 2 time points.

Previous studies have used different imaging techniques to describe the range of motion and the sequence of initiation of motion of individual segments during flexion and extension of the lumbar spine. Dvorak et al [16], Pearcy et al [15], and Staub et al [11] described range of motion in rotation during maximal passive flexion and extension of each level. Since range of motion differed between studies and resulted in a high inter, and intraindividual variability, a more consistent method to define physiological motion was pursued. Initiation of motion was described by several previous studies. Because of limitations (eg, reporting pooled data instead of individual sequences, limited range of motion, analyses of part of the lumbar spine, and describing intervertebral rotation at specific time points or specific ranges of motion instead of between successive frames), results differ between studies with high inter and intraindividual variability [7-10,13,14,17].

In a cervical spine study [18], a more consistent sequence of segmental contribution during the end of the extension, namely in 80% of the participants in T1 and 90% in T2, was found using the same measurement method and setup as this study. The reliability, sensitivity, and specificity of this measurement method showed high scores, with a reliability, determined in Fleiss Kappa, of 0.80-0.84, average sensitivity of 90%, and average specificity of 85% [18]. We believe that these findings show that this method is accurate and reproducible to determine the sequence of segmental contribution in cervical spine. Even though the setup of this study was similar, we found less consistent motion patterns in the lumbar spine. We believe several variables between the cervical spine and lumbar spine contribute to our differences in consistency in motion patterns. These variables are as follows: facet orientation, intervertebral discs loadings, the spino-pelvic relationship, and muscle recruitment.

Cervical facet joint surfaces between C3 and T1 have a 45 degrees angle to the transverse plane [27]. In the lumbar spine, the superior articular process is medially orientated, and the inferior articular process is laterally orientated in the sagittal plane, with right angles to the transverse plane [28]. These differences in orientation result in less constrained facet joints of the lumbar spine, resulting in a greater freedom of motion, which could explain a less consistent movement compared to the cervical spine [27]. The uncinate process and uncovertebral joints, found from C3 to C7, also provide stability and mobility of the cervical spine by functioning as a guide rail during flexion and extension and limit rotation and bending, resembling a saddle joint [29]. Since these structures are not present in the lumbar spine, it could lead to less consistent motion patterns due to less constraint of the motion segments. Intervertebral discs of the cervical spine and lumbar spine are both wedged shaped with a larger anterior side of the disc compared to the posterior side [30]. In addition, both discs are elliptical shaped, with a larger cross-sectional area of lumbar intervertebral discs than cervical spine [30]. In this study, it is possible that the axial loading of the intervertebral disc is altered by fixation of the pelvis and the seating position. Nachemson et al [27] described a relative increase in intervertebral disc pressure, ascending from supine to standing to sitting position and from neutral position to flexion. Furthermore, forced anteversion or retroversion of the pelvis caused by the fixed position could influence motion patterns of the lumbar spine during flexion and extension. There is no study that compares motion of the lumbar spine in a standing versus sitting position. The pelvis and abdominal structures also led to overprojection in segment L5S1, making it challenging to trace these segments with the computer software. For this reason, ICCs were determined in this study, resulting in an average ICC of all segments mostly above 0.60, except for L5S1. Furthermore, analyses showed that the lower lumbar spine segments, L4L5 and L5S1, showed inconsistent patterns throughout all recordings. When excluding them from analyses, a more consistent pattern from L1 to L4 appeared. In addition to the difficulty due to overprojection at L5S1, the motion of L4L5 and L5S1 is influenced by more variables compared to the upper lumbar spine, leading to less consistency. The lower lumbar segments function as a kinematic transition zone from a highly mobile region (ie, upper lumbar spine) to an immobile sacroiliac region [31]. For this reason, it is also plausible that pathology mostly occurs in lower lumbar segments.

Finally, muscle recruitment differs between the cervical and lumbar spine. In the cervical spine, the range of rotations is mostly influenced by muscle recruitment, except for the end stages of the motion, which are influenced by gravity [27]. In the lumbar spine, rotation is controlled by muscle recruitments throughout the whole motion. Muscle recruitment and strength is affected by age, sex, motivation, pain, as well as muscle and joint physiology and geometry [27]. This means that mostly interindividual motion differences can be explained by differences in muscle recruitment and strength, which plays a larger part in the lumbar spine compared to the cervical spine motion. In addition, the 4 abdominal muscles (ie, rectus abdominis as well as external and internal oblique and transverse abdominal muscle) have a great influence on flexion of the lumbar spine, with an increased muscle recruitment per degree of flexion [27]. This could also be an explanation for the fact that lumbar flexion patterns are less consistent than lumbar extension patterns.

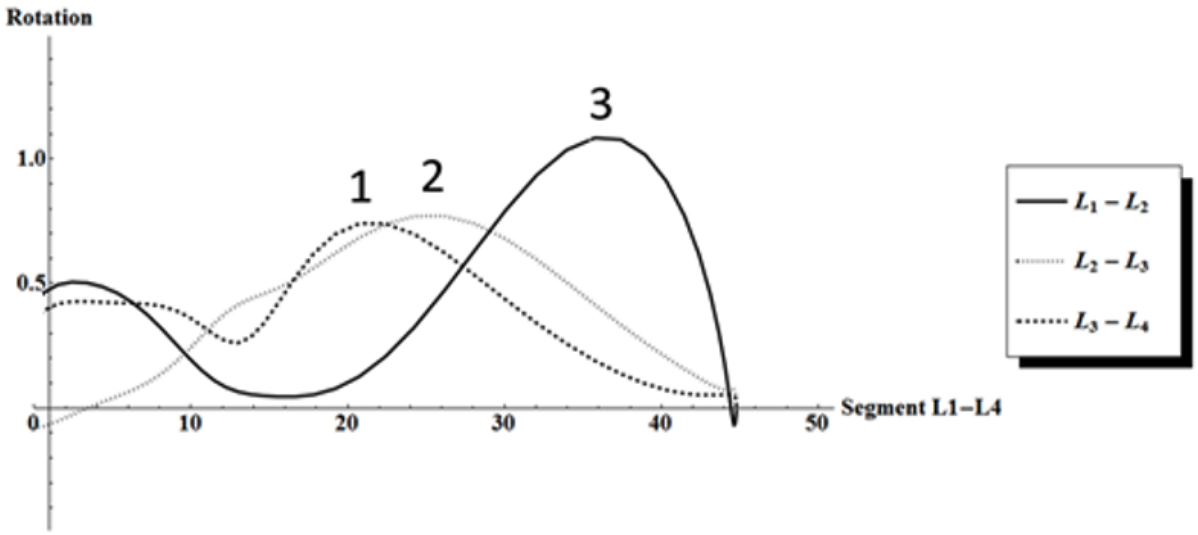
Strengths and Limitations

There are multiple strengths to this study. First, the intention of this study was to determine motion patterns of L1 to S1, instead of a selection of vertebrae by using a sufficient field of view. However, especially segment L5S1 was difficult to track due to the overprojection of the pelvis and abdominal structures. Additionally, it is possible that L5S1 also had less focus since

this segment was placed at the maximum bottom of the field of view. This resulted in a mean ICC below 0.60 for L5S1 and an inconsistent motion pattern throughout the recordings.

Second, this study described motion patterns during maximum flexion and extension of an individual instead of the usually reported fixed ranges to determine physiological motion. Maximum flexion and extension represents the lumbar motion in daily activity better, as it does not limit a person to move within a strict range. Furthermore, patients could move differently because they had to stay within a range of motion, which could influence the muscle recruitment. The downside of using maximum range motion patterns is the possibility of segments moving outside the field of view. This happened once in P1-T1 (Figure 3), resulting in the exclusion of this cinematographic recording from final analyses. However, Figure 2 shows a peak in L3L4, followed by a peak in L2L3 during the last phase of extension, comparable with the abovementioned most consistent sequence of motion.

Figure 3. Sagittal rotation in segments in the upper lumbar spine (segments L2 to L4, since L1 fell outside the field of view during extension of the lumbar spine in P1-T1). On the y-axis, the rotation is shown in degrees between successive frames. On the x-axis, cumulative degrees of extension in block L2 to L4 are shown. Peaks of the graphs per segments (L2L3 and L3L4) depict maximum contribution of the segment in a specific phase of the extension. At the end of the extension, the peak of L3L4 was followed by a peak of L2L3. Each series of values undergoes smoothing by means of a low-pass Gaussian digital filter.



Finally, since sequence of segmental contribution in the cervical spine showed consistent motion patterns in the study of Boselie et al [18], we used the same imaging technique for recordings, the same computer tracking software, and the same research team in this study [18]. Additionally, cinematographic recordings of all participants were supervised by the same team (IC and CH Christoph) and performed with the use of the same customized chair. The included participants were all male, around the same age, and with a BMI below 25 kg/m^2 to minimize the influence of age, sex, and body habitus on muscle recruitment and overprojection of abdominal structures. Female participants were excluded to protect their ovaries from direct radiation exposure. However, Staub et al [11], Troke et al [5], Dvorak et al [16], and Wong et al [8] showed no statistically significant difference between sexes in motion of the lumbar spine.

This study also had some limitations. First, sagittal balance parameters were not determined during this study, as femoral heads were not shown in the cinematographic recording. A fixed pelvis could influence the motion of the lumbar spine by forced anteversion or retroversion, which could have been determined using these parameters. Second, the measurement method used to develop the graphics was a time-consuming method. For this reason, the possibility of using artificial intelligence should be investigated to determine if it could lower the workload without losing reliability of the measurements. However, this would be more important for cervical spine analyses, as lumbar spine analyses using this method showed less consistent motion patterns, and therefore, it will have less clinical relevance. It could be possible that another analyzing method should be used to determine physiological motion of the lumbar spine. It has been suggested that center of rotation (COR), defined as the point around which motion segments of the lumbar spine move, could quantify the kinematic features of the lumbar spine [32]. COR of the lumbar spine was the main topic in many previous studies. However, conditions to determine COR varied between studies (eg, symptomatic and asymptomatic participants, different motion tasks, as well as before and after surgery). A current systematic review [32] is analyzing and summarizing data of these different studies to determine if COR could be used to define physiological motion of the lumbar spine. Unfortunately, results are not yet available.

Third, this study was conducted with 11 participants, resulting in 22 cinematographic recordings over 2 time points. Despite this small sample size, we believe that expansion of the study group would not have led to more conclusive results, since there were also intraindividual variabilities between the 2 time points

besides interindividual variabilities, and previous research showed consistent results with similar group sizes.

Conclusions

This study aimed to provide physiological motion patterns of the lumbar spine based on the sequence of segmental contribution. A total of 64% (7/11) of the cinematographic recordings of asymptomatic young male participants showed a consistent pattern at both time points during the last phase of extension, with a peak in rotation in L3L4, followed by a peak in L2L3, and finally, in L1L2. Since 36% (4/11) of the cinematographic recordings did not show a consistent pattern, we believe that physiological motion of the lumbar spine is a broad concept, which cannot be stated in a firm definition using this method. Even in healthy participants, multiple factors are responsible for inconsistencies in lumbar spine motion patterns, which can be aggravated in case of lumbar pathology. For this reason and because of the time-consuming method for analysis, we believe the clinical relevance in this form will be limited, and it should not be used as a diagnostic tool to distinguish between physiological and pathological motions.

Acknowledgments

We would like to thank CH Christoph for the assistance during the cinematographic recordings. We would also like to thank Dr SN Van Laarhoven for his contribution to the customized wooden chair.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Graphs flexion of lumbar motion.
[PDF File (Adobe PDF File), 663 KB - biomedeng_v8i1e41906_app1.pdf]

Multimedia Appendix 2
Graphs extension L1 to S1.
[PDF File (Adobe PDF File), 663 KB - biomedeng_v8i1e41906_app2.pdf]

Multimedia Appendix 3
Graphs of extension cinematographic recordings L1 to L4; T1 and t2.
[PDF File (Adobe PDF File), 699 KB - biomedeng_v8i1e41906_app3.pdf]

References

1. Dittmar O. Röntgenstudien zur Mechanologie der Wirbelsäule. Z Orthop Chir 1931;55:321-351.
2. Jinkins JR, Dworkin JS, Damadian RV. Upright, weight-bearing, dynamic-kinetic MRI of the spine: initial results. Eur Radiol 2005 Sep 20;15(9):1815-1825. [doi: 10.1007/s00330-005-2666-4] [Medline: 15906040]
3. Svedmark P, Tullberg T, Noz ME, Maguire GQ, Zeleznik MP, Weidenhielm L, et al. Three-dimensional movements of the lumbar spine facet joints and segmental movements: in vivo examinations of normal subjects with a new non-invasive method. Eur Spine J 2012 Apr 1;21(4):599-605 [FREE Full text] [doi: 10.1007/s00586-011-1988-y] [Medline: 21881866]
4. Powers CM, Kulig K, Harrison J, Bergman G. Segmental mobility of the lumbar spine during a posterior to anterior mobilization: assessment using dynamic MRI. Clin Biomech (Bristol, Avon) 2003 Jan;18(1):80-83. [doi: 10.1016/s0268-0033(02)00174-2] [Medline: 12527251]
5. Troke M, Moore AP, Maillardet FJ, Cheek E. A normative database of lumbar spine ranges of motion. Man Ther 2005 Aug;10(3):198-206. [doi: 10.1016/j.math.2004.10.004] [Medline: 16038855]
6. Russell P, Weld A, Pearcy MJ, Hogg R, Unsworth A. Variation in lumbar spine mobility measured over a 24-hour period. Br J Rheumatol 1992 May;31(5):329-332. [doi: 10.1093/rheumatology/31.5.329] [Medline: 1581775]
7. Wong KWN, Luk KDK, Leong JCY, Wong SF, Wong KKY. Continuous dynamic spinal motion analysis. Spine (Phila Pa 1976) 2006 Feb 15;31(4):414-419. [doi: 10.1097/01.brs.0000199955.87517.82] [Medline: 16481951]
8. Wong KW, Leong JC, Chan M, Luk KD, Lu WW. The flexion-extension profile of lumbar spine in 100 healthy volunteers. Spine (Phila Pa 1976) 2004 Aug 01;29(15):1636-1641. [doi: 10.1097/01.brs.0000132320.39297.6c] [Medline: 15284509]
9. Takayanagi K, Takahashi K, Yamagata M, Moriya H, Kitahara H, Tamaki T. Using cineradiography for continuous dynamic-motion analysis of the lumbar spine. Spine (Phila Pa 1976) 2001 Sep 01;26(17):1858-1865. [doi: 10.1097/00007632-200109010-00008] [Medline: 11568694]

10. Lee S, Wong KWN, Chan M, Yeung H, Chiu JLF, Leong JCY. Development and validation of a new technique for assessing lumbar spine motion. Spine (Phila Pa 1976) 2002 Apr 15;27(8):E215-E220. [doi: 10.1097/00007632-200204150-00022] [Medline: 11935121]
11. Staub BN, Holman PJ, Reitman CA, Hipp J. Sagittal plane lumbar intervertebral motion during seated flexion-extension radiographs of 658 asymptomatic nondegenerated levels. SPI 2015 Dec;23(6):731-738. [doi: 10.3171/2015.3.spine14898]
12. Wu S, Jou J, Lee H, Chen H, Su F, Kuo L. The reproducibility comparison of two intervertebral translation measurements in cervical flexion-extension. Spine J 2015 May 01;15(5):1083-1091. [doi: 10.1016/j.spinee.2013.06.097] [Medline: 24239486]
13. Harada M, Abumi K, Ito M, Kaneda K. Cineradiographic motion analysis of normal lumbar spine during forward and backward flexion. Spine (Phila Pa 1976) 2000 Aug 01;25(15):1932-1937. [doi: 10.1097/00007632-200008010-00011] [Medline: 10908936]
14. Kanayama M, Abumi K, Kaneda K, Tadano S, Ukai T. Phase lag of the intersegmental motion in flexion-extension of the lumbar and lumbosacral spine. An in vivo study. Spine (Phila Pa 1976) 1996 Jun 15;21(12):1416-1422. [doi: 10.1097/00007632-199606150-00004] [Medline: 8792517]
15. Pearcy M, Portek I, Shepherd J. Three-dimensional x-ray analysis of normal movement in the lumbar spine. Spine (Phila Pa 1976) 1984 Apr;9(3):294-297. [doi: 10.1097/00007632-198404000-00013] [Medline: 6374922]
16. Dvorák J, Panjabi MM, Chang DG, Theiler R, Grob D. Functional radiographic diagnosis of the lumbar spine. Flexion-extension and lateral bending. Spine (Phila Pa 1976) 1991 May;16(5):562-571 [FREE Full text] [doi: 10.1097/00007632-199105000-00014] [Medline: 2052999]
17. Okawa A, Shinomiya K, Komori H, Muneta T, Arai Y, Nakai O. Dynamic motion study of the whole lumbar spine by videofluoroscopy. Spine (Phila Pa 1976) 1998 Aug 15;23(16):1743-1749. [doi: 10.1097/00007632-199808150-00007] [Medline: 9728375]
18. Boselie T, van Santbrink H, de Bie RA, van Mameren H. Pilot study of sequence of segmental contributions in the lower cervical spine during active extension and flexion: healthy controls versus cervical degenerative disc disease patients. Spine (Phila Pa 1976) 2017 Jun 01;42(11):E642-E647. [doi: 10.1097/BRS.0000000000001914] [Medline: 27879563]
19. Caelers IJMH, Boselie TFM, Rijkers K, Van Hemert WLW, De Bie RA, Van Santbrink H. Lumbar intervertebral motion analysis during flexion and extension cinematographic recordings in healthy Male Participants: Protocol. JMIR Res Protoc 2020 Feb 24;9(2):e14741 [FREE Full text] [doi: 10.2196/14741] [Medline: 32130199]
20. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000 Nov 15;25(22):2940-52; discussion 2952. [doi: 10.1097/00007632-200011150-00017] [Medline: 11074683]
21. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957 Dec;16(4):494-502 [FREE Full text] [doi: 10.1136/ard.16.4.494] [Medline: 13498604]
22. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov 07;63 Suppl 11(S11):S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
23. Boselie TF, van Mameren H, de Bie RA, van Santbrink H. Cervical spine kinematics after anterior cervical discectomy with or without implantation of a mobile cervical disc prosthesis; an RCT. BMC Musculoskelet Disord 2015 Feb 21;16(1):34 [FREE Full text] [doi: 10.1186/s12891-015-0479-4] [Medline: 25887569]
24. Van Mameren H, Drukker J, Sanches H, Beursgens J. Cervical spine motion in the sagittal plane (I) range of motion of actually performed movements, an X-ray cinematographic study. Eur J Morphol 1990;28(1):47-68. [Medline: 2390411]
25. NEDERLANDSE COMMISSIE VOOR STRALINGSDOSIMETRIE. Publication of the Netherlands Commission on Radiation Dosimetry 2017:1-20 [FREE Full text]
26. Reinartz R, Platel B, Boselie T, van Mameren H, van Santbrink H, Romeny B. Cervical vertebrae tracking in video-fluoroscopy using the normalized gradient field. Med Image Comput Comput Assist Interv 2009;12(Pt 1):524-531. [doi: 10.1007/978-3-642-04268-3_65] [Medline: 20426028]
27. Buckwalter J, Einhorn TA, Simon SR. Chapter 28, Kinesiology. In: Orthopaedic basic science, biology and biomechanics of the muskuloskeletal system. Rosemont, IL: American Academy of Orthopaedic surgeons; 2000:766-782.
28. Schünke M. Prometheus, General Anatomy and Muskuloskeletal System. Houten, the Netherlands: Bohn Staflau van Loghum; 2010.
29. Tubbs RS, Rompala OJ, Verma K, Mortazavi MM, Benninger B, Loukas M, et al. Analysis of the uncinate processes of the cervical spine: an anatomical study. J Neurosurg Spine 2012 Apr;16(4):402-407. [doi: 10.3171/2011.12.SPINE11541] [Medline: 22264177]
30. Pooni J, Hukins D, Harris P, Hilton R, Davies K. Comparison of the structure of human intervertebral discs in the cervical, thoracic and lumbar regions of the spine. Surg Radiol Anat 1986 Sep;8(3):175-182. [doi: 10.1007/bf02427846]
31. Cunningham BW, Mueller KB, Hawken JB, Rolle NP. Biomechanical considerations and mechanisms of injury in spondylolisthesis. Semin Spin Surg 2020 Sep;32(3):100803. [doi: 10.1016/j.semss.2020.100803]

32. Funabashi M, Breen A, De Carvalho D, Henry A, Murnaghan K, Pagé I, et al. Center of rotation locations during lumbar spine movements: a scoping review protocol. JBI Evid Synth 2020 Jun;18(6):1305-1312. [doi: 10.11124/JBISRIR-D-19-00080] [Medline: 32813379]

Abbreviations

COR: center of rotation
ICC: intraclass correlation coefficient
T: time point

Edited by T Leung; submitted 14.08.22; peer-reviewed by S Muraki, A Breen, Y Wu; comments to author 12.01.23; revised version received 15.01.23; accepted 31.01.23; published 20.06.23.

Please cite as:
Caelers I, Boselie T, van Hemert W, Rijkers K, De Bie R, van Santbrink H
The Variability of Lumbar Sequential Motion Patterns: Observational Study
JMIR Biomed Eng 2023;8:e41906
URL: https://biomedeng.jmir.org/2023/1/e41906
doi:10.2196/41906
PMID:38875682

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org