
JMIR Biomedical Engineering

Engineering for health technologies, medical devices, and innovative medical treatments and procedures
Volume 9 (2024) ISSN 2561-3278 Editor in Chief: Syed A. A. Rizvi, MD, PhD, MBA, MPH, BSN

Contents

Reviews

- Finite Element Analysis for Degenerative Cervical Myelopathy: Scoping Review of the Current Findings and Design Approaches, Including Recommendations on the Choice of Material Properties ([e48146](#))
Benjamin Davies, Samuel Schaefer, Amir Rafati Fard, Virginia Newcombe, Michael Sutcliffe. 3
- Trends in South Korean Medical Device Development for Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder: Narrative Review ([e60399](#))
Yunah Cho, Sharon Talboys. 25

Original Papers

- Sacroiliac Joint Dysfunction in Endurance Runners Using Wearable Technology as a Clinical Monitoring Tool: Systematic Review ([e46067](#))
Stuart Evans. 42
- Investigation of Deepfake Voice Detection Using Speech Pause Patterns: Algorithm Development and Validation ([e56245](#))
Nikhil Kulangareth, Jaycee Kaufman, Jessica Oreskovic, Yan Fossat. 61
- Validation of a Novel Noninvasive Technology to Estimate Blood Oxygen Saturation Using Green Light: Observational Study ([e46974](#))
Sanjay Gokhale, Vinoop Daggubati, Georgios Alexandrakis. 72
- Preliminary Assessment of an Ambulatory Device Dedicated to Upper Airway Muscle Training in Patients With Sleep Apnea: Proof-of-Concept Study ([e51901](#))
Patrice Roberge, Jean Ruel, André Bégin-Drolet, Jean Lemay, Simon Gakwaya, Jean-François Masse, Frédéric Sériès. 81
- Impact of Audio Data Compression on Feature Extraction for Vocal Biomarker Detection: Validation Study ([e56246](#))
Jessica Oreskovic, Jaycee Kaufman, Yan Fossat. 92
- Agreement Between Apple Watch and Actical Step Counts in a Community Setting: Cross-Sectional Investigation From the Framingham Heart Study ([e54631](#))
Nicole Spartano, Yuankai Zhang, Chunyu Liu, Ariel Chernofsky, Honghuang Lin, Ludovic Trinquart, Belinda Borrelli, Chaturangi Pathirivasan, Vik Kheterpal, Christopher Nowak, Ramachandran Vasam, Emelia Benjamin, David McManus, Joanne Murabito. 99

Assessing the Accuracy of Smartwatch-Based Estimation of Maximum Oxygen Uptake Using the Apple Watch Series 7: Validation Study (e59459)	
Polona Caserman, Sungsoo Yum, Stefan Göbel, Andreas Reif, Silke Matura.	115
A Deep Learning Framework for Predicting Patient Decannulation on Extracorporeal Membrane Oxygenation Devices: Development and Model Analysis Study (e48497)	
Joshua Fuller, Alexey Abramov, Dana Mullin, James Beck, Philippe Lemaitre, Elham Azizi.	127
Enhancing Energy Efficiency in Telehealth Internet of Things Systems Through Fog and Cloud Computing Integration: Simulation Study (e50175)	
Yunyong Guo, Sudhakar Ganti, Yi Wu.	142
An Engineering Alternative to Lockdown During COVID-19 and Other Airborne Infectious Disease Pandemics: Feasibility Study (e54666)	
Yusaku Fujii.	157
Stroke Survivors' Interaction With Hand Rehabilitation Devices: Observational Study (e54159)	
Chioma Wodu, Gillian Sweeney, Milena Slachetka, Andrew Kerr.	170
Classifying Residual Stroke Severity Using Robotics-Assisted Stroke Rehabilitation: Machine Learning Approach (e56980)	
Russell Jeter, Raymond Greenfield, Stephen Housley, Igor Belykh.	181

Review

Finite Element Analysis for Degenerative Cervical Myelopathy: Scoping Review of the Current Findings and Design Approaches, Including Recommendations on the Choice of Material Properties

Benjamin Davies¹, BSc, MBChB (Hons), MPhil; Samuel Schaefer², MEng; Amir Rafati Fard¹, BA; Virginia Newcombe¹, MD, PhD; Michael Sutcliffe², MA, PhD

¹Department of Medicine, University of Cambridge, Cambridge, United Kingdom

²Department of Engineering, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:

Benjamin Davies, BSc, MBChB (Hons), MPhil

Department of Medicine

University of Cambridge

Addenbrooke's Hospital, Hills Road

Cambridge, CB2 0QQ

United Kingdom

Phone: 44 07766692608

Email: bd375@cam.ac.uk

Abstract

Background: Degenerative cervical myelopathy (DCM) is a slow-motion spinal cord injury caused via chronic mechanical loading by spinal degenerative changes. A range of different degenerative changes can occur. Finite element analysis (FEA) can predict the distribution of mechanical stress and strain on the spinal cord to help understand the implications of any mechanical loading. One of the critical assumptions for FEA is the behavior of each anatomical element under loading (ie, its material properties).

Objective: This scoping review aims to undertake a structured process to select the most appropriate material properties for use in DCM FEA. In doing so, it also provides an overview of existing modeling approaches in spinal cord disease and clinical insights into DCM.

Methods: We conducted a scoping review using qualitative synthesis. Observational studies that discussed the use of FEA models involving the spinal cord in either health or disease (including DCM) were eligible for inclusion in the review. We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. The MEDLINE and Embase databases were searched to September 1, 2021. This was supplemented with citation searching to retrieve the literature used to define material properties. Duplicate title and abstract screening and data extraction were performed. The quality of evidence was appraised using the quality assessment tool we developed, adapted from the Newcastle-Ottawa Scale, and shortlisted with respect to DCM material properties, with a final recommendation provided. A qualitative synthesis of the literature is presented according to the Synthesis Without Meta-Analysis reporting guidelines.

Results: A total of 60 papers were included: 41 (68%) "FEA articles" and 19 (32%) "source articles." Most FEA articles (33/41, 80%) modeled the gray matter and white matter separately, with models typically based on tabulated data or, less frequently, a hyperelastic Ogden variant or linear elastic function. Of the 19 source articles, 14 (74%) were identified as describing the material properties of the spinal cord, of which 3 (21%) were considered most relevant to DCM. Of the 41 FEA articles, 15 (37%) focused on DCM, of which 9 (60%) focused on ossification of the posterior longitudinal ligament. Our aggregated results of DCM FEA indicate that spinal cord loading is influenced by the pattern of degenerative changes, with decompression alone (eg, laminectomy) sufficient to address this as opposed to decompression combined with other procedures (eg, laminectomy and fusion).

Conclusions: FEA is a promising technique for exploring the pathobiology of DCM and informing clinical care. This review describes a structured approach to help future investigators deploy FEA for DCM. However, there are limitations to these recommendations and wider uncertainties. It is likely that these will need to be overcome to support the clinical translation of FEA to DCM.

(*JMIR Biomed Eng* 2024;9:e48146) doi:[10.2196/48146](https://doi.org/10.2196/48146)

KEYWORDS

scoping review; fine element analysis; cervical spine; spinal cord; degenerative cervical myelopathy

Introduction

Degenerative cervical myelopathy (DCM) occurs when arthritic changes to the structure of the cervical spine injure the spinal cord, causing a slowly progressive spinal cord injury (SCI) [1]. This leads to a range of different symptoms that can affect the whole body, including loss of dexterity, imbalance, altered sensation, bladder and bowel dysfunction, and pain [2]. Although DCM is estimated to affect 1 in 50 adults, <20% are estimated to receive a diagnosis. This is likely, in part, as most are only mildly affected [3,4]. Treatment is currently limited to surgery but, due to inherent risks, is reserved for those with progressive or moderate-to-severe disease [5]. Notably, <5% of patients with DCM will make a complete recovery after surgery, and instead are left with lifelong disabilities and dependence having among the lowest quality of life scores of any disease [6,7]. Consequently, this was recently estimated to cost GBP £0.7 billion (approximately US \$0.9 billion) per year [8].

The etiology and pathophysiology of DCM are poorly understood [1,9]. At a macroscopic level, this is a cohort that displays progressive cervical myelopathy with degenerative changes to the structure of their cervical spine, typically causing some deformation of the spinal cord on magnetic resonance imaging (MRI), which responds to decompressive surgery. This led to the hypothesis that DCM is triggered by a chronic mechanical injury, specifically compression loading.

However, this is likely to be an oversimplification. Spinal cord compression is most commonly an incidental finding [3]; the amount of compression visualized on the MRI poorly correlates with the disease severity and does not predict the treatment response [10-12]. Moreover, many other forms of mechanical loading also occur, including stretching or shear loading. These are recognized to be capable of causing tissue injury independently [1]. For example, stretching is considered the etiology of myelopathy in tethered cord syndrome and some forms of deformity [13]. Consequently, it is more likely that the mechanical trigger in DCM is the interaction of these mechanical forces rather than one alone. As the structural changes within the spine highly vary between patients, this is likely to be a very individualized phenomenon [14]. This presents a problem for clinical practice, as conventional diagnostic tests such as MRI cannot measure mechanical stress; however, the goal of surgery is to alleviate it [12,15].

Finite element analysis (FEA) is an engineering technique that uses a computational model to derive the extent and severity of mechanical stress from an assumed loading [16]. This has frequently been applied to health care, including, to some extent, SCI and, more recently, DCM [16-18]. FEA could have important applications in DCM, both to improve our understanding of the pathobiology and to represent an individual's injury and objectively inform surgical strategy.

To perform an FEA, a computer model incorporating the geometry, motion, and material properties of each structure

must be created [17]. Geometry and motion, to a large extent, can be defined based on an individual's clinical imaging. However, the material properties must be chosen from other sources. These choices will influence the results of the FEA. For spinal cord FEA to date, these choices have been made on a project-by-project basis, typically informed by the experience of the investigators, their interpretation and knowledge of the literature, and their specific project aims. To inform the development of FEA for DCM, we adopted an iterative approach using a scoping review methodology with the following aims:

- To describe how FEA models have been constructed with respect to spinal cord disease
- To identify and appraise the experimental literature that has informed their material property choices to make recommendations on the material properties for DCM FEA
- To aggregate the findings from studies using FEA to explore DCM.

To the best of our knowledge, this represents a unique approach to selecting the material properties for a clinical FEA model and may represent an exemplar for similar initiatives.

Methods

A scoping review methodology was considered most appropriate to meet these objectives [19]. This scoping review was reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines ([Multimedia Appendix 1](#)).

Search Strategy

The search was conducted using a modified population, interventions, comparisons, and outcomes strategy, which states that the research question for a review must include the population, intervention, comparison, and outcome. Our research question was, "what are the current findings and design approaches for FEA in DCM?", with the population being patients with DCM, intervention being FEA, and outcomes being current findings and design approaches. To more comprehensively guide future decisions regarding the application of FEA methods to DCM, we broadened our inclusion criteria to incorporate any study that applied FEA to the spinal cord (in either health or disease). Consequently, the search terms were designed to capture observational studies that had developed FEA models that included the spinal cord in either health or disease, including DCM ([Multimedia Appendix 2](#)). Searches were conducted from inception (February 12, 2021) to September 1, 2021, in the MEDLINE and Embase databases. Search sensitivity was evaluated using 5 papers known to meet the inclusion criteria; all papers were successfully captured [18,20-23].

Inclusion and Exclusion Criteria

Papers were considered eligible for inclusion if they were observational studies that discussed the use of FEA models that

included the spinal cord of humans or animals in either health or disease, including DCM.

Papers were excluded if they were written in a language other than English, did not use FEA models, or did not include the spinal cord in the FEA model. Furthermore, systematic reviews, scoping reviews, editorials, and abstracts were excluded.

Study Screening and Data Extraction

Two reviewers (BMD and SS) independently performed title and abstract screening with blinding using Rayyan (Rayyan Systems Inc). A pilot screen of 100 publications was conducted to ensure concordance between reviewers. Any disagreements following unblinding were resolved by discussion between the reviewers until mutual agreement was reached. In this review, papers identified through our search strategy are termed “FEA articles”.

From the included FEA articles, the references used to justify a structure’s material properties were also screened to identify experimental studies reporting original data acquired from physical tissue tests. Studies exploring behavior computationally but including their original physical experiments, even if published elsewhere, were included. Studies that explored properties solely on a computational basis were excluded. This forward search continued within the references of a referenced study if the reference did not meet this criterion and had cited an alternative source.

Papers were retrieved for full-text screening and data extraction using a piloted pro forma. Data extracted from the papers included: author, year of publication, country, study objectives, study design (eg, human or animal study), disease of interest (if any), spinal segment (eg, cervical, thoracic, and lumbar), reference for anatomy (eg, cadaveric specimen and imaging), and details of how the FEA model was developed and validated (including the material properties of the anatomical elements).

Data extraction focused on the properties specifically referenced by the original FEA models and may not have included all the material properties discussed in the paper. To understand an investigator’s approach to model development, these were distinguished as those used to define the model a priori (ie, referenced data and the choice of material law and selected coefficients) or those used to validate the final model (if performed). However, for the purpose of selecting data to inform an FEA model, these references were aggregated and termed as “source articles” in this review.

In the absence of a standard quality assessment tool for experimental studies of biomechanics, we developed a classification to help appraise source articles that are most appropriate for a DCM FEA model [24]. This included a risk of bias assessment adapted from the Newcastle-Ottawa Scale, focusing on selection and reporting bias (Multimedia Appendix 3) [25].

Data Analysis and Reporting

Due to significant heterogeneity between methodologies, meta-analysis was not possible, and a qualitative Synthesis

Without Meta-Analysis (SWiM) was instead performed. Data were aggregated, where applicable, qualitatively, quantitatively, or using frequency statistics, as per the SWiM guidelines [26].

Given the small field size, with many papers published by single groups, citation networks were created to graphically consider which choices were made across the field and how they were informed. Using this framework and our judgment, we ranked source articles into approximate tertiles. For FEA articles that had cited top-source articles and represented the material properties using an equation, the performance of this equation was further evaluated graphically by generating stress-strain curves. These were exclusively either linear or hyperelastic. For models using a linear elastic equation, the Young modulus was used as the gradient of the stress-strain curve. For models using a hyperelastic equation, a 3×3 element cube was created using ABAQUS (Dassault Systèmes). The cube was stretched uniaxially, with no constraint applied in the orthogonal directions, linearly increasing the nominal strain in increments of 0.04 to a maximum of 0.4. The outputs of this model were then applied true stress as a function of the applied true strain. Finally, any primary clinical papers that conducted FEA for the investigation of DCM were aggregated separately and analyzed.

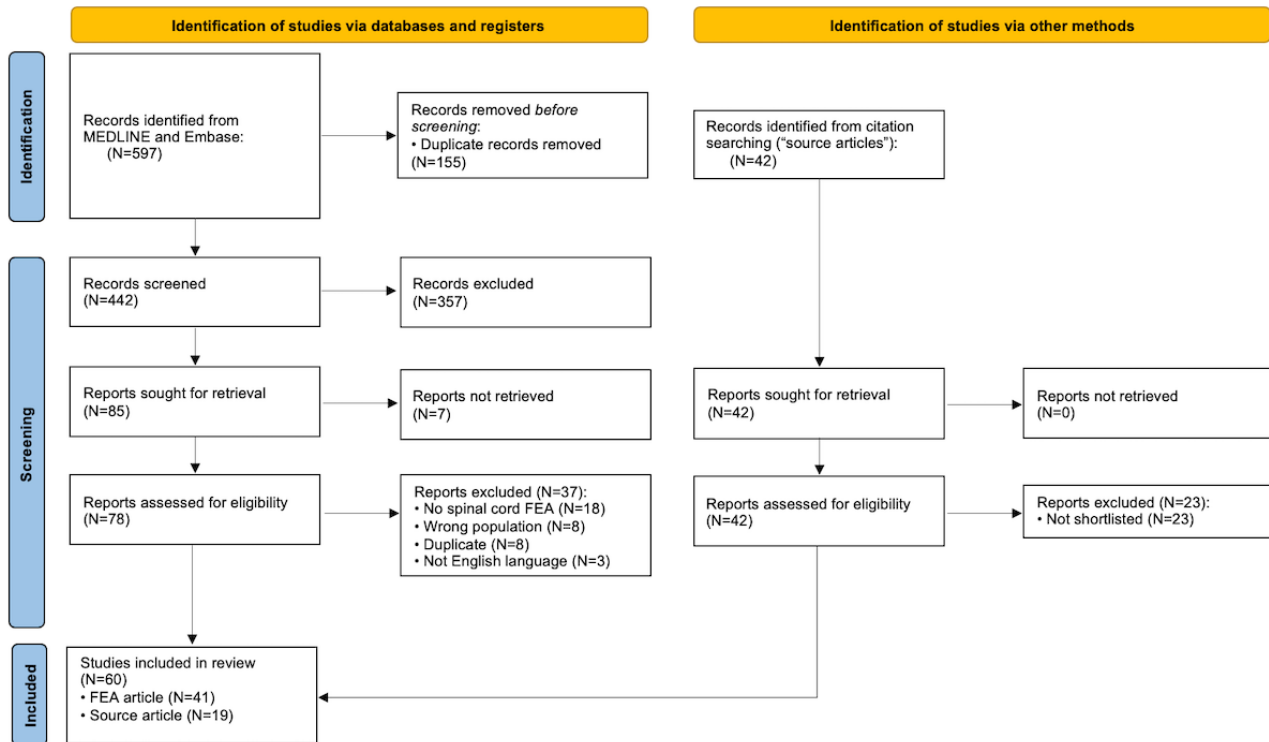
Data were displayed using a range of plots constructed using R Studio (version 4.0.3; Posit).

Results

Overall Approach of FEA Models of Spinal Cord Disease: Anatomy, Geometry, Motion, and Validation

The search returned 597 articles, of which 155 (25.9%) were duplicates (Figure 1). Following screening, 41 FEA articles were eligible for inclusion, of which 32 (78%) modeled the human spinal cord; a further 45 (7.54%) source articles were identified through citation search, of which 19 (42%) were shortlisted as suitable. Of the FEA articles, approximately half (21/41, 51%) focused on SCI [27-47]; 34% (14/41) on DCM [18,20-22,48-57]; and 5% (2/41) each on scoliosis [58,59], syringomyelia [60,61], and flexion myelopathy [62,63]. Most models (25/41, 61%) included only the spinal cord, whereas 24% (10/41) included the surrounding anatomy at multiple vertebral levels, and 17% (7/41) included the surrounding anatomy at only 1 motion segment (ie, 2 adjacent vertebrae). Physiological movement of the spine (flexion and extension) was incorporated into 17% (7/41) of the models, but none evaluated spinal cord oscillation. This was equally likely among the DCM and SCI models (Multimedia Appendix 4).

The anatomy of each model was built using a combination of imaging and cadaveric data in 27% (11/41) of the FEA articles. Typically, imaging was used for bones and cadavers for soft tissues, including the spinal cord. This included an open-source reference library called BodyWorks [64] and a review of spinal cord geometry [65]. MRI was used to define the spinal cord specifically in 20% (8/41) of the FEA articles.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. FEA: finite element analysis.

For most FEA articles (33/41, 80%), the spinal cord was modeled as gray matter and white matter separately and had a defined pial layer (26/41, 63%) or was encased within the dural layer (26/41, 63%). Defined pial and dural layers were used in combination in only half of these articles (13/41, 32%). Cerebrospinal fluid (CSF) was specifically modeled in 41% (17/41) of the FEA articles, while other elements were variably included. This choice was independent of the disease and publication date (Multimedia Appendix 4). Elements were modeled using solid shell elements, unless specified differently in the Material Properties of Anatomical Elements With Recommendations for DCM FEA section.

Validation methods were specified in 63% (26/41) of the FEA articles, with 15% (6/41) using their own experiments and 9% (20/41) using literature (Multimedia Appendix 5). These references pointed to 17 articles, of which 7 (42%) provided material property data for the spinal cord in healthy circumstances and 3 (18%) in traumatic SCI circumstances. Of the remaining 17 articles, 4 (24%) described motion of the spine [66-69] and 1 (6%) described the spinal cord in flexion and extension [70]. Of the 9 articles providing information on healthy spinal cord properties, 7 (78%) were also used in other studies to inform the selection of material property. No DCM-specific validation data sets were identified.

Material Properties of Anatomical Elements With Recommendations for DCM FEA

Spinal Cord

The material properties of the whole spinal cord were defined in 22% (9/41) of the FEA articles. This was rarely justified, but if so, qualified by its uncertain significance [71,72]. Typically, a hyperelastic Ogden variant (4/9, 44%) or a linear elastic (3/9, 33%) function was used.

For the remaining models, gray and white matter were modeled separately, except for the article that explored the impact of a range of white matter material properties, where the material law applied to gray matter was the same as that of white matter. The remaining 32 models were mostly based on tabulated data from the studies by Ichihara et al [72,73], and less frequently, Bilston and Thibault [74], Tunturi [75], and Ozawa et al [76]. Alternatively, a hyperelastic Ogden variant (10/41, 24%) or a linear elastic (4/41, 10%) function was used.

A total of 2 studies specifically compared different material properties with respect to a transverse contusion model of SCI. Jannesar et al [38] explored white matter properties on the basis that single constitutive models may not account for the dynamic (viscoelastic) and anisotropic properties. They identified that this could be improved by adding reinforcing functions. A second order reduced polynomial hyperelastic function combined with a quadratic reinforcing function in a 4-term Prony series performed best ($0.89 < R^2 < 0.99$), although this was principally in relation to the high strain rates of an SCI. Fournely et al [45] used a first-order Ogden function but varied the stiffness of the gray matter with respect to the white matter. Although this fell within the range of the validation data set, they observed differing responses to the load. When the gray matter was stiffer than the white matter, strain distribution was more diffuse and maximal within the white matter. When the stiffness was equivalent, strain was localized to the impact site. When the white matter was stiffer than the gray matter, strain was less localized, maximal within the gray matter and involved the contralateral gray matter. This was the principal factor determining behavior, ahead of other factors explored, including spinal cord diameter, curvature, and impactor angle.

A total of 2 studies similarly explored the implications of different gray and white matter material properties with respect

to DCM, with similar findings discussed in the Findings From the FEA Studies of DCM section [34,50].

A total of 14 source articles were identified describing the material properties of the spinal cord or its subcomponents (Multimedia Appendix 6 [46,72-75,77-90]), of which 3 (21%) were shortlisted with relevance to an FEA for DCM [72-74]. Their interpretations varied across studies (Figure 2). The choice of material laws and values of those who directly cited the

prioritized source articles and separately distinguished gray and white matter are listed in Tables 1-2. Broadly, these align with the source articles; however, there are differences across the strain range (Multimedia Appendix 6). Of these FEA articles representing material properties with an equation, studies by Jannesar et al [29] and Khuyagbaatar et al [53] were selected as these were most aligned for gray matter and white matter, respectively.

Figure 2. Network analysis of finite element analysis models, which is linked to a shortlisted source article, for the white matter (A) and gray matter (B) or the spinal cord as a whole (C). The original finite element analysis models are represented by their choice of material law as a star (linear elastic), square (hyperelastic), diamond (tabulated), or triangle (other) and their disease of interest as degenerative cervical myelopathy (DCM; red), spinal cord injury (SCI; blue), or other (green). These link to the primary source articles (dots). An intermediate article, that is, the one that did not include primary experimental data, is pale gray. A shortlisted source article is black. Each figure is additionally available as an interactive file; refer to Multimedia Appendix 7. The higher resolution version of this figure is available in Multimedia Appendix 8.

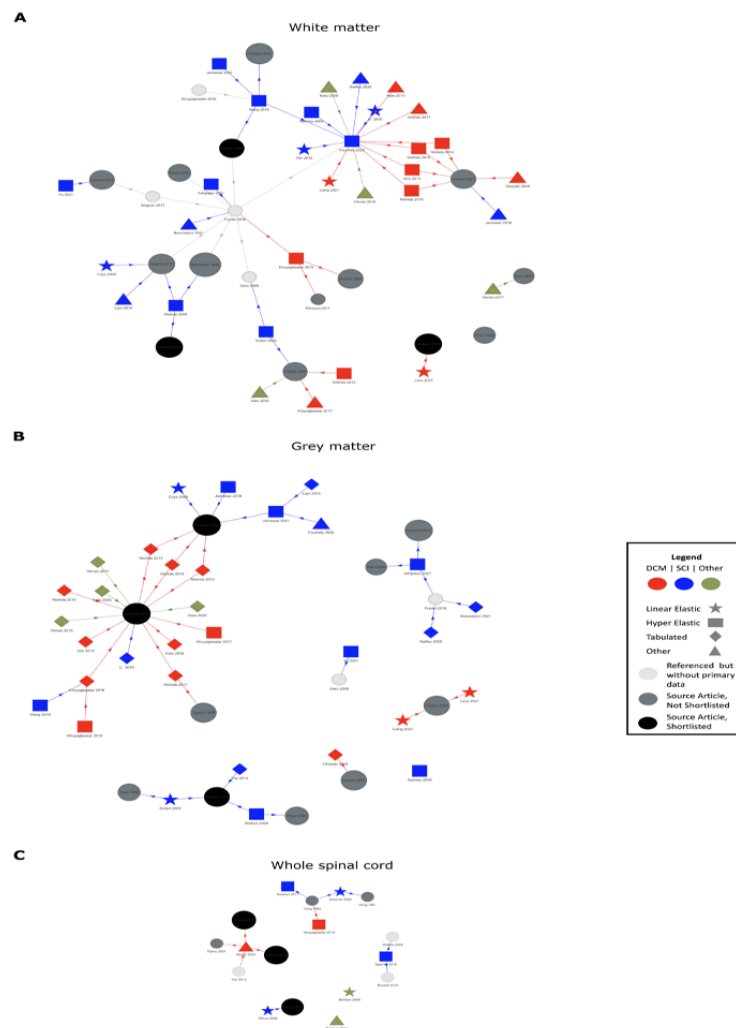


Table 1. Extracted material equations for the gray matter.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	ν^b	α^c	μ^d (MPa)	D ^e (MPa ⁻¹)
Jannesar et al [29], 2021 ^f	SCI ^g	Ichihara et al [72], 2003	Hyperelastic	Ogden, first Order	— ^h	0.49	10.57	0.0445	<i>0.905^{i,j}</i>
Khuyagbaatar et al [53], 2017	DCM ^k	Ichihara et al [73], 2001	Hyperelastic	Ogden, first order	—	<i>0.45</i>	14.7	0.0041	50.5
Jannesar et al [38], 2016	SCI	Ichihara et al [72], 2003	Hyperelastic	Ogden, first order	—	0.45	7.52	0.0306	6.77
Khuyagbaatar et al [39], 2016	SCI	Ichihara et al [73], 2001	Hyperelastic	Ogden, first order	—	<i>0.45</i>	14.7	0.0041	50.5
Czyz et al [42], 2008	SCI	Ichihara et al [72], 2003	Linear elastic	—	0.656	0.499	—	<i>0.2188</i>	—
Maikos et al [43], 2008	SCI	Bilston and Thibault [74], 1996	Hyperelastic	Ogden, first order	—	0.45	4.7	0.0320	6.47
Scifert et al [44], 2002	SCI	Bilston and Thibault [74], 1996	Linear elastic	—	0.0667	0.499	—	<i>0.0222</i>	—

^aE: Young modulus.

^b ν : Poisson ratio. Where missing, the value of ν was assumed to be 0.45.

^c α : material exponent parameter.

^d μ : ground shear hyperelastic modulus.

^eD: compressibility constant.

^fThe single preferred source of the authors based on modeling (Multimedia Appendix 6), where a range of equations were put forward.

^gSCI: spinal cord injury.

^hNot available.

ⁱDenotes a suspected error in original text and input value given.

^jValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu\{1+\nu\})$, and for linear elastic, $\mu=E/(2\{1+\nu\})$.

^kDCM: degenerative cervical myelopathy.

Table 2. Extracted material equations for the white matter.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	ν^b	α^c	μ^d (MPa)	D ^e (MPa ⁻¹)
Liang et al [48], 2021	DCM ^f	Ichihara et al [73], 2001	Linear elastic	— ^g	4.2	0.45	—	<i>1.4483^h</i>	—
Khuyagbaatar et al [52], 2017	DCM	Ichihara et al [73], 2001	Hyperelastic	Ogden, first order	—	<i>0.45</i>	12.5	0.0040	51.7
Khuyagbaatar et al [39], 2016	SCI ⁱ	Ichihara et al [73], 2001	Hyperelastic	Ogden, first order	—	<i>0.45</i>	12.5	0.0040	51.7
Czyz et al [42], 2008	SCI	Ichihara et al [72], 2003	Linear elastic	—	0.277	0.499	—	<i>0.0924</i>	—
Maikos et al [43], 2008	SCI	Bilston and Thibault [74], 1996	Hyperelastic	Ogden, first order	—	0.45	4.7	0.0320	6.47
Scifert et al [44], 2002	SCI	Bilston and Thibault [74], 1996	Linear elastic	—	0.0667	0.499	—	<i>0.0222</i>	—

^aE: Young modulus.

^b ν : Poisson ratio. Where missing, the value of ν was assumed to be 0.45.

^c α : material exponent parameter.

^d μ : ground shear hyperelastic modulus.

^eD: compressibility constant.

^fDCM: degenerative cervical myelopathy.

^gNot available.

^hValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu\{1+\nu\})$, and for linear elastic, $\mu=E/(2\{1+\nu\})$.

ⁱSCI: spinal cord injury.

Pia

Of the 26 FEA articles with defined pia, 14 (54%) used a linear elastic function, 9 (21%) did not report their method, and 2 (5%) used a hyperelastic Ogden variant function. The remaining study (1/26, 4%) used tabulated data from the study by Ichihara et al [73].

A total of 4 source articles were identified for the pia ([Multimedia Appendix 6](#)), of which 2 (50%) were shortlisted as suitable [75,77]. The choice of material laws and the values of those who directly cited these shortlisted source articles are listed in [Table 3](#). These equations have differences in how they represent the source article ([Multimedia Appendix 6](#)). Of the FEA articles representing material properties with an equation, the study by Jannesar et al [38] was selected as the most preferred.

Table 3. Extracted material equations for the pia.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	v ^b	α ^c	μ ^d (MPa)	D ^e (MPa ⁻¹)
Henao et al [58,59], 2017	Other	Tunturi [75], 1978	Linear elastic	— ^f	100	0.4	—	35.71 ^g	—
Nishida et al [91], 2016	DCM ^h	Tunturi [75], 1978	—	—	—	—	—	—	—
Nishida et al [54], 2015	DCM	Tunturi [75], 1978	—	—	—	—	—	—	—
Nishida et al [55], 2014	DCM	Tunturi [75], 1978	—	—	—	—	—	—	—
Nishida et al [22], 2012	DCM	Tunturi [75], 1978	—	—	—	—	—	—	—
Henao et al [58], 2018	Other	Tunturi [75], 1978	Linear elastic	—	100	0.4	—	35.71	—
Kato et al [56], 2010	DCM	Tunturi [75], 1978	—	—	—	—	—	—	—
Kato et al [62], 2008	Other	Tunturi [75], 1978	—	—	—	—	—	—	—
Kato et al [63], 2009	Other	Tunturi [75], 1978	—	—	—	—	—	—	—
Jannesar et al [38], 2016 ⁱ	SCI	Kimpara et al [77], 2006	Linear elastic	—	39.3	0.3	—	15.12	—

^aE: Young modulus.

^bv: Poisson ratio. Where missing, the value of v was assumed to be 0.45.

^cα: material exponent parameter.

^dμ: ground shear hyperelastic modulus.

^eD: compressibility constant.

^fNot available.

^gValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu\{1+\nu\})$, and for linear elastic, $\mu=E/(2\{1+\nu\})$.

^hDCM: degenerative cervical myelopathy.

ⁱThe single preferred source of the authors based on modelling ([Multimedia Appendix 6](#)).

Dura

Of the 26 models with defined dura, 18 (69%) used a linear elastic function, 5 (19%) used a hyperelastic Ogden variant, and 3 (12%) did not report their method.

Persson et al [46] compared the performance of a linear and hyperelastic function, which is summarized in the following CSF section.

A total of 9 source articles were referenced ([Multimedia Appendix 6](#)), of which 4 (44%) were shortlisted [78-81]. The choice of material laws and values of those who directly cited these prioritized source articles are listed in [Table 4](#). These equations have differences in how they represent the source article ([Multimedia Appendix 6](#)). Of the FEA articles representing material properties with an equation, the study by Sparrey et al [33] was selected as preferred.

Table 4. Extracted material equations for the dura.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	ν^b	α^c	μ^d (MPa)	D ^e (MPa ⁻¹)
Stoner et al [20], 2020	DCM ^f	Persson et al [92], 2020	Linear Elastic	— ^g	5	0.45	—	1.72 ^h	—
Khuyagbaatar et al [49], 2018	DCM	Persson et al [92], 2020	Linear Elastic	—	80	0.49	—	26.85	—
Henao et al [58,59], 2017	Other	Wilcox et al [47], 2004	Linear Elastic	—	231	0.45	—	79.66	—
Khuyagbaatar et al [52], 2017	DCM	Persson et al [92], 2020	Linear Elastic	—	80	0.49	—	26.85	—
Sparrey et al [33], 2016 ⁱ	SCI	Hong et al [78], 2011 and Zarzur et al [79], 1996	Hyper-elastic	Ogden, 1st Order	—	0.45	16.2	1.205	0.172
Yan et al [36], 2012	SCI	Wilcox et al [47], 2004	Linear Elastic	—	142	0.45	—	48.97	—
Henao et al [58], 2018	Other	Wilcox et al [47], 2004	Linear Elastic	—	231	0.45	—	79.66	—
Khuyagbaatar et al [39], 2016	SCI	Persson et al [92], 2020	Linear Elastic	—	80	0.49	—	26.85	—
Khuyagbaatar et al [57], 2015	DCM	Persson et al [92], 2020	Linear Elastic	—	80	0.49	—	26.85	—
Khuyagbaatar et al [57], 2015	SCI	Persson et al [92], 2020	Linear Elastic	—	80	0.49	—	26.85	—
Czyz et al [42], 2008	SCI	Wilcox et al [47], 2004	Linear Elastic	—	142	0.45	—	48.97	—
Persson et al [46], 2011	SCI	Persson et al [92], 2020	Linear Elastic	—	80	0.49	—	26.85	—
Wilcox et al [47], 2004	SCI	Wilcox et al [47], 2004	Anisotropic Elastic	—	Young modulus in the radial direction=142, Young modulus in the circumferential direction=142, Young modulus in the longitudinal direction=0.7	—	—	—	—

^aE: Young modulus.

^b ν : Poisson ratio. Where missing, the value of ν was assumed to be 0.45.

^c α : material exponent parameter.

^d μ : ground shear hyperelastic modulus.

^eD: compressibility constant.

^fDCM: degenerative cervical myelopathy.

^gNot available.

^hValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu\{1+\nu\})$, and for linear elastic, $\mu=E/(2\{1+\nu\})$.

ⁱThe single preferred source of the authors based on modelling ([Multimedia Appendix 6](#)).

Dentate Ligament

Of the 13 FEA articles that included the dentate ligament, 12 (92%) used a linear elastic function and 1 (8%) used tabulated data. Typically, these were modeled using shell elements (6/13,

46%) with geometric properties, but 8% (1/13) used link elements and 15% (2/13) used spring elements.

A total of 2 source articles were referenced ([Multimedia Appendix 6](#)), of which both were shortlisted [75,82]. The choice of material laws and values of those who directly cited these prioritized source articles are listed in [Table 5](#).

Table 5. Extracted material equations for the dentate.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	ν^b	α^c	μ^d (MPa)	D ^e (MPa ⁻¹)
Henao et al [58,59], 2017	Other	Tunturi [75], 1978	Linear elastic	— ^f	100	0.4	—	35.7 ^g	—
Henao et al [58], 2018	Other	Tunturi [75], 1978	Linear elastic	—	100	0.4	—	35.7	—
Greaves et al [41], 2008	SCI ^h	Tunturi [75], 1978	Linear elastic	—	5.8	—	—	2.0	—
Czyz et al [42], 2008	SCI	Tunturi [75], 1978	Linear elastic	—	100	0.3	—	38.5	—

^aE: Young modulus.

^b ν : Poisson ratio. Where missing, the value of ν was assumed to be 0.45.

^c α : material exponent parameter.

^d μ : ground shear hyperelastic modulus.

^eD: compressibility constant.

^fNot available.

^gValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu\{1+\nu\})$, and for linear elastic, $\mu=E/(2\{1+\nu\})$.

^hSCI: spinal cord injury.

Cerebrospinal Fluid

Of the 17 models that included CSF, 8 (47%) modeled it as a Newtonian fluid. Alternatives included modeling CSF as a pressurized fluid cavity (1/17, 6%), modeling it as a polynomial equation of state (1/17, 6%), modeling it as smoothed particular hydrodynamics (1/17, 6%), using a hyperelastic Mooney-Rivlin model (3/17, 18%), or using a linear elastic equation (1/17, 6%).

Persson et al [46] and Jones et al [93] specifically explored the implications of including a CSF cavity, with or without the dura. To measure cord deformation, Persson et al [46] used an FEA model with reference to a transverse bovine impaction model of SCI, whereas Jones et al [93] performed their own bovine and surrogate cord experiments. They observed that the presence of CSF reduced stress and strain (Persson et al [46]) on the spinal cord and deformation (Jones et al [93]) in the spinal cord. Persson et al [46] demonstrated this was through a greater longitudinal distribution, particularly when the dura was

included and modeled using a hyperelastic Ogden (as opposed to linear elastic) function. Furthermore, Persson et al [46] observed that cord deformation occurred upon contact with the dura (before the CSF between the spinal cord and the dura was redistributed). Jones et al [93] observed that the inclusion of the dura only changed behavior if CSF was also included.

Furthermore, Arhptsov and Marom [31] explored CSF pressure, alongside the presence or absence of epidural fat, using a computational contusion model of SCI based on a thoracic burst fracture. Both CSF and epidural fat were modeled using smoothed particular hydrodynamics. In a model without epidural fat, spinal cord stress and strain increased with increasing CSF pressure. However, in the model with epidural fat, spinal cord stress and strain decreased with increasing CSF pressure.

A total of 5 source articles were referenced (Multimedia Appendix 6), of which 3 (60%) were shortlisted [46,83,84]. The choice of material laws and values of those who directly cited these prioritized source articles are listed in Table 6.

Table 6. Extracted material equations for the cerebrospinal fluid.

Study, year	Pathology	Reference	Law	Viscosity (Pa/s)	Density (kg/m ³)
Khuyagbaatar et al [52], 2017	DCM ^a	Bloomfield et al [83], 1998	Newtonian Fluid	0.001	— ^b
Arhptsov [31], 2021	SCI ^c	Persson et al [46], 2011	Polynomial Equation of State	—	—
Khuyagbaatar et al [39], 2016	DCM	Bloomfield et al [83], 1998, Brydon et al [84], 1995	Newtonian Fluid	0.001	—
Khuyagbaatar et al [39], 2016	SCI	Bloomfield et al [83], 1998, Brydon et al [84], 1995	Newtonian Fluid	0.001	1000
Khuyagbaatar et al [57], 2015	DCM	Bloomfield et al [83], 1998, Brydon et al [84], 1995	Newtonian Fluid	0.001	—
Khuyagbaatar et al [57], 2015	SCI	Bloomfield et al [83], 1998, Brydon et al [84], 1995	Newtonian Fluid	0.001	—
Persson et al [46], 2011	SCI	Bloomfield et al [83], 1998	Newtonian Fluid	0.001	—

^aDCM: degenerative cervical myelopathy.

^bNot available.

^cSCI: spinal cord injury.

Posterior Longitudinal Ligament and Ligamentum Flavum

The analysis focused on the posterior longitudinal ligament and ligamentum flavum, given their specific involvement in the pathobiology of DCM. In all 6 instances included, they were included together and modeled in the same manner: using piecewise linear plasticity (2/6, 33%), linear elastic function

(2/6, 33%), hyperelastic Ogden variant (1/6, 17%), or tabulated data (1/6, 17%).

A total of 6 source articles were referenced (Multimedia Appendix 6), of which 3 (50%) were shortlisted [85-87]. The choice of material laws and values of those who directly cited these prioritized source articles are listed in Tables 7 and 8.

Table 7. Extracted material equations for the ligamentum flavum.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	v ^b	α ^c	μ ^d (MPa)	D ^e (MPa ⁻¹)
Greaves et al [41], 2008	SCI ^f	Yoganandan et al 1989 and 2000 [86,87]	Linear elastic	— ^g	3.8	—	—	1.3 ^h	—

^aE: Young modulus.

^bv: Poisson ratio. Where missing, the value of v was assumed to be 0.45.

^cα: material exponent parameter.

^dμ: ground shear hyperelastic modulus.

^eD: compressibility constant.

^fSCI: spinal cord injury.

^gNot available.

^hValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu(1+\nu))$, and for linear elastic, $\mu=E/(2(1+\nu))$.

Table 8. Extracted material equations for the posterior longitudinal ligament.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	v ^b	α ^c	μ ^d (MPa)	D ^e (MPa ⁻¹)
Greaves et al [41], 2008	SCI ^f	Przybylski et al [85], 1996 and Yoganandan 1989 and 2000 [86,87]	Linear elastic	— ^g	35.7	—	—	12.3 ^h	—

^aE: Young modulus.

^bv: Poisson ratio. Where missing, the value of v was assumed to be 0.45.

^cα: material exponent parameter.

^dμ: ground shear hyperelastic modulus.

^eD: compressibility constant.

^fSCI: spinal cord injury.

^gNot available.

^hValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu(1+\nu))$, and for linear elastic, $\mu=E/(2(1+\nu))$.

Spinal Roots

A total of 7 models included spinal nerve roots, of which 2 (29%) distinguished between the intradural and extradural components. These 2 models specifically explored the nature of C5 palsy in relation to surgery for DCM [49,57]. Nerve roots

were all modeled with spring elements, either as a spring (5/7, 71%) or with a linear elastic equation (2/7, 29%).

A total of 2 source articles of equivalent quality were referenced (Multimedia Appendix 6) [88,89]. The choice of material laws and values of those who directly cited these prioritized source articles are listed in Table 9.

Table 9. Extracted material equations for the nerve roots.

Study, year	Pathology	Reference	Law	E ^a (MPa)	ν ^b	Spring constant	Mass (g)
Lévy et al [18], 2021	DCM ^c	Kulkarni [88], 2007	Spring	— ^d	—	0.133	0.1
Khuyagbaatar et al [49], 2018	DCM	Singh [89], 2005	Linear Elastic	1.3	0.3	—	—
Henao et al [58,59], 2017	Other	Kulkarni [88], 2007	Spring	—	—	0.133	—
Khuyagbaatar et al [52], 2017	DCM	Singh [89], 2005	Linear Elastic	1.3	0.3	—	—
Henao et al [58], 2018	Other	Kulkarni [88], 2007	Spring	—	—	0.133	—

^aE: Young modulus.

^bν: Poisson ratio; where missing, ν was assumed to be 0.45. For Kulkarni et al [88], the unit is uncertain, with a range of different units referenced across its citations.

^cDCM: degenerative cervical myelopathy.

^dNot available.

Other Elements

Other elements included in some models were bone (14/41, 34%); intervertebral disks (IVDs; 13/41, 31%); and the remaining spinal ligaments, such as the anterior longitudinal or interspinous ligament.

The bone was generally modeled as a rigid body (8/14, 57%). Of the 8 models, 3 (21%) subdivided the vertebrae into anatomical subcomponents (eg, body, laminae, and spinous process), and 5 (36%) distinguished between cortical and cancellous bone, of which 3 (60%) applied an equation just to the cortical bone (linear elastic in all cases) and 2 (40%) applied a Johnson-Cook or plastic kinematic equation. We found no eligible source articles using our search process.

The IVD were modeled as a single entity in 54% (7/13) of the papers, typically as a rigid body (5/7, 71%) or using a linear elastic equation (2/7, 29%). Alternatively, they were modeled separately as nucleus pulposus and annulus fibrosus. Techniques for the nucleus pulposus included a Mooney-Rivlin model (3/6, 50%), Ogden second-order variant (1/6, 17%), and fluid elements (2/6, 33%). The annulus fibrosus included a Mooney-Rivlin model (2/6, 33%), Ogden second-order variant (1/6, 17%), Ogden third-order variant (1/6, 17%), and linear elastic equation (2/6, 33%).

A total of 3 source articles were found for IVD, and 1 was shortlisted (Multimedia Appendix 6) [90]. The choice of material laws and values of those who directly cited these prioritized source articles are listed in Table 10.

Table 10. Extracted material equations for the intervertebral disc.

Study, year	Pathology	Reference	Law	Variant	E ^a (MPa)	ν ^b	α ^c	μ ^d (MPa)	D ^e (MPa ⁻¹)
Greaves et al [41], 2008	SCI ^f	Spilker et al [90], 1986	Linear elastic	— ^g	3.4	—	—	1.2 ^h	—

^aE: Young modulus.

^bν: Poisson ratio. Where missing, the value of ν was assumed to be 0.45.

^cα: material exponent parameter.

^dμ: ground shear hyperelastic modulus.

^eD: compressibility constant.

^fSCI: spinal cord injury.

^gNot available.

^hValues in italics are input based on the identity for isotropic materials, $D=3(1-2\nu)/(\mu\{1+\nu\})$, and for linear elastic, $\mu=E/(2\{1+\nu\})$.

Findings From the FEA Studies of DCM

Of the DCM models, 60% (9/15) specifically focused on ossification of the posterior longitudinal ligament (OPLL), a specific subtype of DCM.

Stress and Static Cord Compression

A total of 8 models explored the relationship between the amount of static spinal cord compression and spinal cord stress. Kato et al [56] and Kim et al [21] used parametric models of the spinal cord to explore the implications of OPLL (anterior compression at 2 adjacent vertebrae). The model was constrained posteriorly, reflecting the lamina. They found that the stress increased with increasing cord compression, with an apparent

exponential relationship. Minimal stress was detected at <40% but dramatically increased at ≥50%. This relationship was replicated by Nishida et al [91] using posterior compression, by Liang et al [48] simulating a disk prolapse, and in a multisegmental model of OPLL by Khuyagbaatar et al [52,57]. Furthermore, it was replicated in cervical spondylosis by Levy et al [18] (Figure 3 [18,21,52,57]).

Maximal stress was observed in the gray matter and, to a lesser extent, in the lateral and posterior funiculus. Nishida et al [91] observed differences in the stress distribution at low compression rates depending on the spinal cord level related to differing morphology; however, beyond a compression rate of 30%, this was consistent (Figure 4).

Okazaki et al [50] explored the implications of spinal cord aging using a parametric model of the spinal cord. The model was given white and gray matter properties based on a young or aged bovine spinal cord specimen. They observed that stress

increased under a low amount of anterior compression in the aged spinal cord and was more widely distributed throughout the gray matter and white matter. In contrast, the gray matter was unaffected in the young specimen.

Figure 3. Spinal cord compression and spinal cord stress in degenerative cervical myelopathy models. For models tabulating the von Mises stress at different measures of static compression or canal stenosis (n=4) [18,21,52,57], the values were plotted on a line graph with a line of best fit representing the average value (blue).

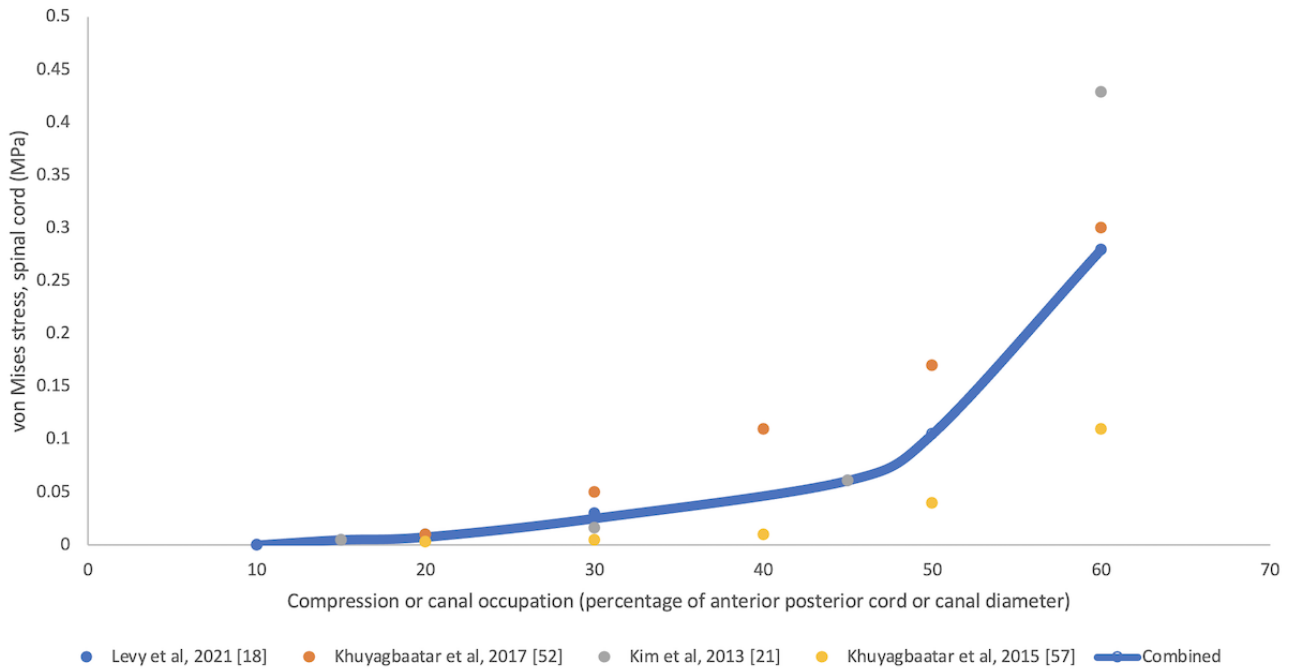
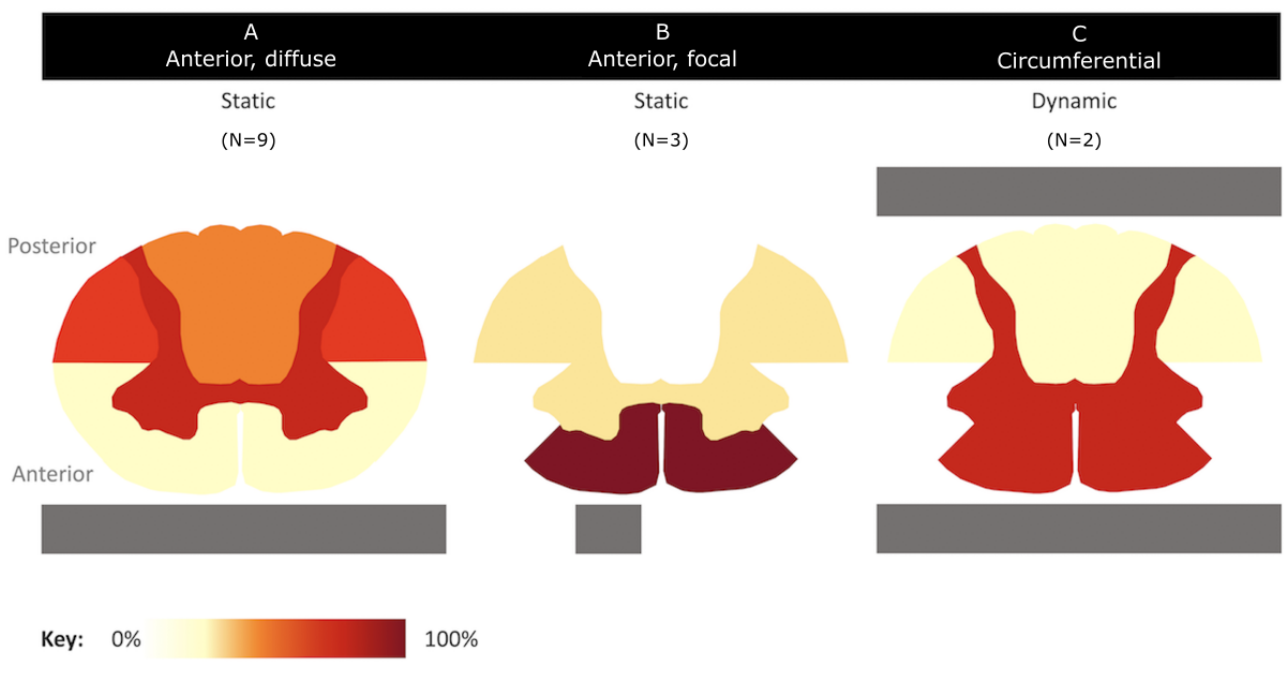


Figure 4. Spinal cord compression and location of spinal cord stress in degenerative cervical myelopathy models. The spinal cord was partitioned, per hemicord, as gray matter and anterior, anterolateral, posterolateral, and posterior white matter. For each study, reporting the cross-sectional distribution of Von Mises stress (n=12) and the location of stress that fell within the top 30% of measured stress was noted. These frequencies were aggregated by compression pattern and displayed for (A) anterior diffuse and static, (B) anterior focal and static, and (C) circumferential and dynamic distribution and location of stress as relative proportions.



Stress and Dynamic Cord Compression

Nishida et al [22] used a parametric model to explore the implications of ligamentum flavum buckling in neck extension in the context of cervical stenosis. For this, the spinal cord was restricted posteriorly by the ligamentum flavum and then anteriorly, either by a central curvature (representing a disk prolapse) or a flat lateral or flat cross-sectional constraint (representing the ligament). The amount of ligamentum flavum buckling was measured using a kinematic MRI. Spinal cord stress was observed in all scenarios and was maximal using the flat cross-sectional constraint.

Later, Nishida et al [54] used a parametric model of OPLL to demonstrate that while dynamic and static compression alone could stress the spinal cord, they could also act together, although it was unclear whether this was additive or multiplicative. In dynamic compression alone, stress was more restricted to gray matter.

Stress and Shape of Cord Compression

Khuyagbaatar et al [57] and Kim et al [21] did not identify any difference in OPLL shape or type with respect to observed spinal cord stress. Furthermore, in the study by Nishida et al [22], the distribution of stress was broadly comparable across the three scenarios affecting the gray matter and anterior and posterolateral aspects of the white matter tracts. In unilateral compression only, the ipsilateral gray matter was affected. Levy et al [18] explored gradually increasing anterior diffuse (broad-based disk), anterior lateral, and circumferential compression using a static multilevel model. Different phenotypes of stress were observed, including peak stress, point of onset, and rate of increase. The highest stress was observed with an anterior diffuse or circumferential compression (Figure 4).

Stress and Surgical Decompression

Khuyagbaatar et al [39] used a multisegmental model to explore the implications of hemilaminectomy, laminectomy, and laminoplasty on spinal cord stress following a 1-, 2-, 3-, or 4-level posterior decompression for continuous OPLL. Stress remained elevated following hemilaminectomy but was low and equivalent between laminectomy and laminoplasty. The postoperative deformity was not modeled.

Nishida et al [55] used a parametric model to explore the implications of alignment following posterior decompression for OPLL. They demonstrated that although stress decreased significantly following decompression, it slightly increased in the anterior funiculus, increasing in the gray matter and posterolateral funiculi with progressive deformity. They subsequently replicated this in a separate analysis [51], demonstrating that kyphosis and increased mobility after decompression would elevate the observed stress.

Khuyagbaatar et al [49,52] explored the effects of laminectomy and laminoplasty, respectively, for the treatment of OPLL using a multisegmental static compression model. They demonstrated that all procedures reduced spinal cord stress significantly (>90%), whether in lordotic (K Line positive) or kyphotic deformity (K Line negative) [94]. However, stress was elevated

within the exiting C5 nerve root following laminectomy if there was a kyphotic deformity and lateral-type OPLL following laminoplasty. In both instances, the amount of nerve root stress was related to the amount of anterior compression.

Stoner et al [20] used a multisegmental dynamic model (C2-T1) to explore the implications of multilevel C4-7 cervical spondylosis (anterior disk prolapses and osteophyte formation) treated with C4-7 anterior cervical discectomy and fusion (ACDF), laminoplasty, or ACDF with laminectomy. Notably, all procedures caused stress to increase at adjacent levels above those of healthy controls. However, a stand-alone ACDF caused increased stress within the spinal cord at C3 to a level above that of the preoperative DCM model in flexion.

Where possible, these were aggregated, demonstrating that the spinal cord tolerated significant compression before stress increased exponentially (Figure 3 [18,21,52,57]). Aggregating the distributions of stress observed across studies, based on the nature of compression, demonstrated differing stress distributions (Figure 4). For static and diffuse anterior compression, the bilateral posterior white matter and gray matter were the most affected. For static and focal compression, the anterior white matter and, to a lesser extent, the gray matter were most affected. This was observed bilaterally despite a focal or lateral element. For circumferential compression in a dynamic model, the bilateral gray matter and posterior white matter were the most affected.

Stress and Tissue Injury

Notably, although differential patterns of stress were observed throughout these DCM models, the levels remained relatively low (<0.5 MPa). DCM FEA models did not explore the relationship between the observed stress and tissue injury.

Discussion

Overview

FEA is a promising technique used in DCM, although there remain uncertainties regarding the ideal approach and its clinical interpretation. This review highlights the numerous decisions investigators must make when performing FEA, which can affect findings and underpin the need for a systematic approach, as applied in this study. On the basis of current evidence, we have shortlisted our preferred material property choices for a DCM model and conclude that a distinction between gray and white matter is preferable.

Principal Findings and Comparison to Prior Work

A total of 15 studies were identified applying FEA to investigate DCM. The insights from these studies broadly align with the current evidence base. First, the spinal cord can tolerate some compression. This is in keeping with clinical practice, where asymptomatic spinal cord compression is far more common [3], and the amount of cord compression is a poor surrogate for disease severity or progression [1]. Second, the movement of the subaxial cervical spine can augment the stress on the spinal cord. This is in keeping with clinical practice, including the concept of dynamic injury and the proposed role of flexion/extension MRI or electrophysiology [95-98]. Finally,

it demonstrated the significant effectiveness of decompression surgery, regardless of the technique, and the comparatively minor gains of using one technique over the other. This is in keeping with clinical practice, where high-quality comparisons of anterior versus posterior surgery are equivalent, and currently, there is no strong evidence that routine stabilization (eg, instrumented fusion vs laminoplasty vs laminectomy or ACDF vs ACDF with a plate) is required [99-102], all pointing toward the need for a personalized surgical approach [15].

Furthermore, although more nuanced findings were proposed by the identified FEA studies and this would require *in vivo* corroboration, the application of FEA in DCM appears well founded overall. More widely, it also seems potentially valuable and timely. The pathobiology of DCM is poorly understood, with its investigation being among the top 10 global research priorities [1]. Current preclinical models have many limitations. For example, common recent models use an expandable polymer inserted behind the spinal cord and within the canal to cause cervical myelopathy. Therefore, this does not model anterior compression, nor does it truly represent a chronic injury mechanism. Furthermore, in clinical practice, clinical decisions are based on imperfect tools [103]. For example, structural MRI in a supine position defines the nature of degenerative changes but not if, where, or how an SCI occurs. FEA could change this, particularly given the parallel advances in the automatic segmentation of MRI [12].

Furthermore, while this review highlights that FEA is a versatile technique, investigators must make many decisions regarding how it is applied. These decisions can alter the findings and, therefore, must be carefully considered. At this stage, there seem to be only a few pervasive insights. First, it seems prudent to model the white matter and gray matter separately. Ichihara et al [73] demonstrated that these structures have differing material properties, and how they are defined alters the observed stress and strain. Furthermore, these structures age differently, as shown by Ozawa et al [76]. Histological studies of DCM have shown differing disease features among the white matter and gray matter, with the gray matter being the focus of more significant cellular changes [9]. Moreover, aging is an important factor in DCM, associated with greater disease severity, a greater rate of progression, and poorer response to treatment [104]. There are also early indicators that accelerating aging is a pathological process [1]. Therefore, the observation that the gray matter was unaffected in the younger spinal cord specimen is noteworthy [34,45,50].

Second, while some models have chosen to use linear elastic equations, time-independent hyperelastic models more closely reflected the known material properties of the spinal cord. These, or simply tabulated data, were generally adopted by DCM studies and supported by a single study that evaluated different approaches [38]. Conceptually, taking a more faithful approach to modeling the spinal cord material properties is likely to be more applicable to DCM and its etiology, as contrasted with traumatic SCI, spinal cord stress may be below the limits for tissue injury (eg, asymptomatic spinal cord compression), and above (eg, DCM). It is worth noting that none of these approaches considers the impact of repetitive injuries, and it is likely that time dependence in modeling is relevant [1]. Given

the timeline of DCM pathogenesis (years), this is likely beyond the normal material scales.

Finally, similar to DCM, as the stresses involved are well below the elastic limit of the bone, the vertebrae can be modeled simply as rigid bodies. The critical aspect for bones is instead the way that their geometry and movement affect the loading on the soft tissues.

However, there remain many uncertainties for further evaluation. These include the role of spinal cord oscillation, the appropriateness of the reference material properties for DCM, and the relationship between the measured stress and tissue injury. First, no studies specifically consider spinal cord oscillations [105]. The spinal cord oscillates cranio-caudally with heart rate. Recent imaging studies have indicated that this increased in the context of symptomatic stenosis, the nature of which may correlate with clinical measures of disease severity [106,107]. Spinal cord oscillation would likely result in a shear force on the spinal cord.

Second, it is uncertain how applicable the material properties are to DCM. Most elements are based on young healthy tissue references. In contrast, the ligaments and disks, for example, in DCM, are often degenerated and calcified, and, as aforementioned, the structure of the spinal cord is also recognized to change with age.

However, most importantly, none of these studies have specifically explored how the measured stress is related to tissue injury. Bridging this gap is critical, not only to fully confirm the appropriateness of FEA for DCM but also to guide its clinical interpretation [108]. All biological systems will have some baseline stress or strain; therefore, establishing disease thresholds will be critical to its development. The parallel development of *in vivo* techniques to measure tissue injury can complement this, for example, microstructural MRI and the less developed but promising serum and CSF biomarkers; however, this requires further prospective study.

Limitations

This study has some limitations. First, the search strategy focused on FEA models of the spinal cord and used citations to identify the source articles for all anatomical elements. Consequently, relevant source articles on the behavior of anatomical elements may have been missed. This is more likely for elements that were further removed from the spinal cord, such as the IVD, and experiments published more recently. This was a pragmatic decision based on the fact that existing investigators would likely have the best perspective on the literature, that this is a small research field, and that detailed biomechanical data on elements such as the IVD were unlikely to be so relevant. Consistent decisions across different research groups and findings across source articles would endorse this. Furthermore, due to the nature of our synthesis, we were unable to update our search. Although this may result in the omission of newer FEA articles, we believe that our review provides a useful approach for future investigators aiming to use FEA in DCM. Second, the methods used to shortlist source articles represent a framework we developed for the purpose of building a DCM FEA model. Again, the popularity of the shortlisted

articles across research groups provides some external validation, but it is possible that different investigators would reach different conclusions. For this reason, all source articles are listed in [Multimedia Appendix 6](#), with their respective direct object identifiers. Third, this review aggregates data from a range of different experimental approaches and aims. Therefore, the analysis is largely qualitative, adhering to the SWiM guidelines [26]. Consequently, some conclusions, such as the relationship between the nature of spinal cord compression and stress distribution, remain tentative.

Conclusions

FEA has significant potential to help unlock uncertainties around the pathophysiology of DCM and inform clinical care. Currently,

the application of FEA to DCM remains in its infancy. This review has adopted an intensive and iterative approach to help future investigators use FEA in DCM, including the aggregation of experimental data reporting on material properties and how they have been interpreted thus far. While single recommendations have been made, they have their limitations. The choice of material properties will influence the model performance, and investigators should consider their decisions carefully, particularly as new evidence emerges. More broadly, the methodology used in this review may be relevant to future updates and other clinical FEA initiatives when selecting material properties.

Acknowledgments

This study aligns with the AO Spine Research Objectives and Common Data Elements for Degenerative Cervical Myelopathy (RECODE-DCM) James Lind Alliance top research priorities, selected by people living and working with degenerative cervical myelopathy. This includes “biological basis,” and, to a lesser extent, “individualizing surgery” and “imaging and electrophysiology.” VN is supported by an NIHR Rosetrees Advanced Trust Fellowship. BMD was supported by a National Institute for Health Research Clinical Doctoral Research Fellowship. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, Rosetrees Trust or the Department of Health and Social Care. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

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

Authors' Contributions

BMD designed the study, developed the search strategy, conducted the searches, screened the retrieved papers, extracted relevant information, and drafted the manuscript. SS contributed to paper screening and data extraction. ARF contributed to the writing of the subsequent drafts of this paper. VJFN and MPFS contributed throughout the project, starting from conceptualization to study design, search strategy development, and editing subsequent drafts of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. [\[DOCX File, 108 KB - biomedeng_v9i1e48146_app1.docx\]](#)

Multimedia Appendix 2

Search strategy. [\[DOCX File, 13 KB - biomedeng_v9i1e48146_app2.docx\]](#)

Multimedia Appendix 3

The quality assessment tool developed by the authors. [\[DOCX File, 13 KB - biomedeng_v9i1e48146_app3.docx\]](#)

Multimedia Appendix 4

Comparison of modeling decisions. [\[DOCX File, 274 KB - biomedeng_v9i1e48146_app4.docx\]](#)

Multimedia Appendix 5

Comparison of chosen equation and reference material property study.

[[DOCX File , 936 KB - biomedeng_v9i1e48146_app5.docx](#)]

Multimedia Appendix 6

Material properties of other anatomical elements.

[[DOCX File , 1052 KB - biomedeng_v9i1e48146_app6.docx](#)]

Multimedia Appendix 7

Interactive network files.

[[ZIP File \(Zip Archive\), 561 KB - biomedeng_v9i1e48146_app7.zip](#)]

Multimedia Appendix 8

Higher-resolution version of [Figure 2](#).

[[PNG File , 5013 KB - biomedeng_v9i1e48146_app8.png](#)]

References

1. Davies BM, Mowforth O, Gharooni AA, Tetreault L, Nouri A, Dhillon RS, et al. A new framework for investigating the biological basis of degenerative cervical myelopathy [AO spine RECODE-DCM research priority number 5]: mechanical stress, vulnerability and time. *Global Spine J* 2022 Feb 17;12(1_suppl):78S-96S [[FREE Full text](#)] [doi: [10.1177/21925682211057546](https://doi.org/10.1177/21925682211057546)] [Medline: [35174728](https://pubmed.ncbi.nlm.nih.gov/35174728/)]
2. Davies BM, Munro C, Khan DZ, Fitzpatrick SM, Hilton B, Mowforth OD, et al. Outcomes of degenerative cervical myelopathy from the perspective of persons living with the condition: findings of a semistructured interview process with partnered internet survey. *Global Spine J* 2022 Apr 18;12(3):432-440 [[FREE Full text](#)] [doi: [10.1177/2192568220953811](https://doi.org/10.1177/2192568220953811)] [Medline: [33203262](https://pubmed.ncbi.nlm.nih.gov/33203262/)]
3. Smith SS, Stewart ME, Davies BM, Kotter MR. The prevalence of asymptomatic and symptomatic spinal cord compression on magnetic resonance imaging: a systematic review and meta-analysis. *Global Spine J* 2021 May 24;11(4):597-607 [[FREE Full text](#)] [doi: [10.1177/2192568220934496](https://doi.org/10.1177/2192568220934496)] [Medline: [32677521](https://pubmed.ncbi.nlm.nih.gov/32677521/)]
4. Davies BM, Mowforth OD, Smith EK, Kotter MR. Degenerative cervical myelopathy. *BMJ* 2018 Feb 22;360:k186 [[FREE Full text](#)] [doi: [10.1136/bmj.k186](https://doi.org/10.1136/bmj.k186)] [Medline: [29472200](https://pubmed.ncbi.nlm.nih.gov/29472200/)]
5. Fehlings MG, Tetreault LA, Riew KD, Middleton JW, Aarabi B, Arnold PM, et al. A clinical practice guideline for the management of patients with degenerative cervical myelopathy: recommendations for patients with mild, moderate, and severe disease and nonmyelopathic patients with evidence of cord compression. *Global Spine J* 2017 Sep;7(3 Suppl):70S-83S [[FREE Full text](#)] [doi: [10.1177/2192568217701914](https://doi.org/10.1177/2192568217701914)] [Medline: [29164035](https://pubmed.ncbi.nlm.nih.gov/29164035/)]
6. Fehlings MG, Ibrahim A, Tetreault L, Albanese V, Alvarado M, Arnold P, et al. A global perspective on the outcomes of surgical decompression in patients with cervical spondylotic myelopathy. *Spine* 2015;40(17):1322-1328. [doi: [10.1097/brs.0000000000000988](https://doi.org/10.1097/brs.0000000000000988)]
7. Oh T, Lafage R, Lafage V, Protosaltis T, Challier V, Shaffrey C, et al. Comparing quality of life in cervical spondylotic myelopathy with other chronic debilitating diseases using the short form survey 36-health survey. *World Neurosurg* 2017 Oct;106:699-706. [doi: [10.1016/j.wneu.2016.12.124](https://doi.org/10.1016/j.wneu.2016.12.124)] [Medline: [28065875](https://pubmed.ncbi.nlm.nih.gov/28065875/)]
8. Davies BM, Phillips R, Clarke D, Furlan JC, Demetriades AK, Milligan J, et al. Establishing the socio-economic impact of degenerative cervical myelopathy is fundamental to improving outcomes [AO spine RECODE-DCM research priority number 8]. *Global Spine J* 2022 Feb;12(1_suppl):122S-129S [[FREE Full text](#)] [doi: [10.1177/21925682211039835](https://doi.org/10.1177/21925682211039835)] [Medline: [35174730](https://pubmed.ncbi.nlm.nih.gov/35174730/)]
9. Badhiwala JH, Ahuja CS, Akbar MA, Witiw CD, Nassiri F, Furlan JC, et al. Degenerative cervical myelopathy - update and future directions. *Nat Rev Neurol* 2020 Feb 23;16(2):108-124. [doi: [10.1038/s41582-019-0303-0](https://doi.org/10.1038/s41582-019-0303-0)] [Medline: [31974455](https://pubmed.ncbi.nlm.nih.gov/31974455/)]
10. Martin AR, De Leener B, Cohen-Adad J, Cadotte DW, Nouri A, Wilson JR, et al. Can microstructural MRI detect subclinical tissue injury in subjects with asymptomatic cervical spinal cord compression? A prospective cohort study. *BMJ Open* 2018 Apr 13;8(4):e019809 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2017-019809](https://doi.org/10.1136/bmjopen-2017-019809)] [Medline: [29654015](https://pubmed.ncbi.nlm.nih.gov/29654015/)]
11. Tetreault L, Kopjar B, Côté P, Arnold P, Fehlings MG. A clinical prediction rule for functional outcomes in patients undergoing surgery for degenerative cervical myelopathy: analysis of an international prospective multicenter data set of 757 subjects. *J Bone Joint Surg Am* 2015 Dec 16;97(24):2038-2046. [doi: [10.2106/JBJS.O.00189](https://doi.org/10.2106/JBJS.O.00189)] [Medline: [26677238](https://pubmed.ncbi.nlm.nih.gov/26677238/)]
12. Martin AR, Tetreault L, Nouri A, Curt A, Freund P, Rahimi-Movaghar V, et al. Imaging and electrophysiology for degenerative cervical myelopathy [AO spine RECODE-DCM research priority number 9]. *Global Spine J* 2022 Feb 19;12(1_suppl):130S-146S [[FREE Full text](#)] [doi: [10.1177/21925682211057484](https://doi.org/10.1177/21925682211057484)] [Medline: [34797993](https://pubmed.ncbi.nlm.nih.gov/34797993/)]
13. Henderson FC, Geddes JF, Vaccaro AR, Woodard E, Berry KJ, Benzel EC. Stretch-associated injury in cervical spondylotic myelopathy: new concept and review. *Neurosurgery* 2005 May;56(5):1101-13; discussion 1101. [Medline: [15854260](https://pubmed.ncbi.nlm.nih.gov/15854260/)]
14. Nouri A, Martin AR, Tetreault L, Nater A, Kato S, Nakashima H, et al. MRI analysis of the combined prospectively collected AOSpine North America and international data: the prevalence and spectrum of pathologies in a global cohort of patients

- with degenerative cervical myelopathy. *Spine (Phila Pa 1976)* 2017 Jul 15;42(14):1058-1067. [doi: [10.1097/BRS.0000000000001981](https://doi.org/10.1097/BRS.0000000000001981)] [Medline: [27861250](https://pubmed.ncbi.nlm.nih.gov/27861250/)]
15. Rodrigues-Pinto R, Montenegro TS, Davies BM, Kato S, Kawaguchi Y, Ito M, et al. Optimizing the application of surgery for degenerative cervical myelopathy [AO spine RECODE-DCM research priority number 10]. *Global Spine J* 2022 Feb 17;12(1_suppl):147S-158S [FREE Full text] [doi: [10.1177/21925682211062494](https://doi.org/10.1177/21925682211062494)] [Medline: [35174733](https://pubmed.ncbi.nlm.nih.gov/35174733/)]
 16. Jones CF, Clarke EC. Engineering approaches to understanding mechanisms of spinal column injury leading to spinal cord injury. *Clin Biomech (Bristol, Avon)* 2019 Apr;64:69-81. [doi: [10.1016/j.clinbiomech.2018.03.019](https://doi.org/10.1016/j.clinbiomech.2018.03.019)] [Medline: [29625748](https://pubmed.ncbi.nlm.nih.gov/29625748/)]
 17. Jones AC, Wilcox RK. Finite element analysis of the spine: towards a framework of verification, validation and sensitivity analysis. *Med Eng Phys* 2008 Dec;30(10):1287-1304. [doi: [10.1016/j.medengphy.2008.09.006](https://doi.org/10.1016/j.medengphy.2008.09.006)] [Medline: [18986824](https://pubmed.ncbi.nlm.nih.gov/18986824/)]
 18. Lévy S, Baucher G, Roche PH, Evin M, Callot V, Arnoux PJ. Biomechanical comparison of spinal cord compression types occurring in Degenerative Cervical Myelopathy. *Clin Biomech (Bristol, Avon)* 2021 Jan;81:105174. [doi: [10.1016/j.clinbiomech.2020.105174](https://doi.org/10.1016/j.clinbiomech.2020.105174)] [Medline: [33279293](https://pubmed.ncbi.nlm.nih.gov/33279293/)]
 19. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018 Nov 19;18(1):143 [FREE Full text] [doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)] [Medline: [30453902](https://pubmed.ncbi.nlm.nih.gov/30453902/)]
 20. Stoner KE, Abode-Iyamah KO, Fredericks DC, Viljoen S, Howard MA, Grosland NM. A comprehensive finite element model of surgical treatment for cervical myelopathy. *Clin Biomech (Bristol, Avon)* 2020 Apr;74:79-86. [doi: [10.1016/j.clinbiomech.2020.02.009](https://doi.org/10.1016/j.clinbiomech.2020.02.009)] [Medline: [32145673](https://pubmed.ncbi.nlm.nih.gov/32145673/)]
 21. Kim YH, Khuyagbaatar B, Kim K. Biomechanical effects of spinal cord compression due to ossification of posterior longitudinal ligament and ligamentum flavum: a finite element analysis. *Med Eng Phys* 2013 Sep;35(9):1266-1271. [doi: [10.1016/j.medengphy.2013.01.006](https://doi.org/10.1016/j.medengphy.2013.01.006)] [Medline: [23419995](https://pubmed.ncbi.nlm.nih.gov/23419995/)]
 22. Nishida N, Kato Y, Imajo Y, Kawano S, Taguchi T. Biomechanical analysis of cervical spondylotic myelopathy: the influence of dynamic factors and morphometry of the spinal cord. *J Spinal Cord Med* 2013 Jul 19;35(4):256-261. [doi: [10.1179/2045772312y.0000000024](https://doi.org/10.1179/2045772312y.0000000024)]
 23. Li Z, Liu H, Yang M, Zhang W. A biomechanical analysis of four anterior cervical techniques to treating multilevel cervical spondylotic myelopathy: a finite element study. *BMC Musculoskelet Disord* 2021 Mar 15;22(1):278 [FREE Full text] [doi: [10.1186/s12891-021-04150-7](https://doi.org/10.1186/s12891-021-04150-7)] [Medline: [33722229](https://pubmed.ncbi.nlm.nih.gov/33722229/)]
 24. Sunstein CR, Kahneman D, Sibony O. *Noise: A Flaw in Human Judgment*. New York, NY: HarperCollins Publishers; May 18, 2021.
 25. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol* 2010 Sep;25(9):603-605. [doi: [10.1007/s10654-010-9491-z](https://doi.org/10.1007/s10654-010-9491-z)] [Medline: [20652370](https://pubmed.ncbi.nlm.nih.gov/20652370/)]
 26. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 2020 Jan 16;368:l6890 [FREE Full text] [doi: [10.1136/bmj.l6890](https://doi.org/10.1136/bmj.l6890)] [Medline: [31948937](https://pubmed.ncbi.nlm.nih.gov/31948937/)]
 27. Yu QQ, Liu SQ, Wang JJ, Xu ML, Zhang WX, Cheng LM, et al. Effects of a contusion load on spinal cord with different curvatures. *Comput Methods Biomech Biomed Engin* 2021 Sep;24(12):1302-1309. [doi: [10.1080/10255842.2021.1884232](https://doi.org/10.1080/10255842.2021.1884232)] [Medline: [33586540](https://pubmed.ncbi.nlm.nih.gov/33586540/)]
 28. Beauséjour MH, Wagnac E, Arnoux PJ, Thiong JM, Petit Y. Numerical investigation of spinal cord injury after flexion-distraction injuries at the cervical spine. *J Biomech Eng* 2022 Jan 01;144(1):011011. [doi: [10.1115/1.4052003](https://doi.org/10.1115/1.4052003)] [Medline: [34369552](https://pubmed.ncbi.nlm.nih.gov/34369552/)]
 29. Jannesar S, Salegio EA, Beattie MS, Bresnahan JC, Sparrey CJ. Correlating tissue mechanics and spinal cord injury: patient-specific finite element models of unilateral cervical contusion spinal cord injury in non-human primates. *J Neurotrauma* 2021 Mar 15;38(6):698-717 [FREE Full text] [doi: [10.1089/neu.2019.6840](https://doi.org/10.1089/neu.2019.6840)] [Medline: [33066716](https://pubmed.ncbi.nlm.nih.gov/33066716/)]
 30. Zhu R, Chen YH, Yu QQ, Liu SQ, Wang JJ, Zeng ZL, et al. Effects of contusion load on cervical spinal cord: a finite element study. *Math Biosci Eng* 2020 Jan 16;17(3):2272-2283 [FREE Full text] [doi: [10.3934/mbe.2020120](https://doi.org/10.3934/mbe.2020120)] [Medline: [32233534](https://pubmed.ncbi.nlm.nih.gov/32233534/)]
 31. Arhptsov K, Marom G. Numerical models of spinal cord trauma: the effect of cerebrospinal fluid pressure and epidural fat on the results. *J Neurotrauma* 2021 Aug 01;38(15):2176-2185. [doi: [10.1089/neu.2021.0065](https://doi.org/10.1089/neu.2021.0065)] [Medline: [33971729](https://pubmed.ncbi.nlm.nih.gov/33971729/)]
 32. Bailly N, Diotalevi L, Beauséjour MH, Wagnac É, Mac-Thiong JM, Petit Y. Numerical investigation of the relative effect of disc bulging and ligamentum flavum hypertrophy on the mechanism of central cord syndrome. *Clin Biomech (Bristol, Avon)* 2020 Apr;74:58-65. [doi: [10.1016/j.clinbiomech.2020.02.008](https://doi.org/10.1016/j.clinbiomech.2020.02.008)] [Medline: [32145670](https://pubmed.ncbi.nlm.nih.gov/32145670/)]
 33. Sparrey CJ, Salegio EA, Camisa W, Tam H, Beattie MS, Bresnahan JC. Mechanical design and analysis of a unilateral cervical spinal cord contusion injury model in non-human primates. *J Neurotrauma* 2016 Jun 15;33(12):1136-1149 [FREE Full text] [doi: [10.1089/neu.2015.3974](https://doi.org/10.1089/neu.2015.3974)] [Medline: [26670940](https://pubmed.ncbi.nlm.nih.gov/26670940/)]
 34. Sparrey CJ, Manley GT, Keaveny TM. Effects of white, grey, and pia mater properties on tissue level stresses and strains in the compressed spinal cord. *J Neurotrauma* 2009 Apr;26(4):585-595 [FREE Full text] [doi: [10.1089/neu.2008.0654](https://doi.org/10.1089/neu.2008.0654)] [Medline: [19292657](https://pubmed.ncbi.nlm.nih.gov/19292657/)]

35. Lam CJ, Assinck P, Liu J, Tetzlaff W, Oxland TR. Impact depth and the interaction with impact speed affect the severity of contusion spinal cord injury in rats. *J Neurotrauma* 2014 Dec 15;31(24):1985-1997 [[FREE Full text](#)] [doi: [10.1089/neu.2014.3392](https://doi.org/10.1089/neu.2014.3392)] [Medline: [24945364](#)]
36. Yan YB, Qi W, Wu ZX, Qiu TX, Teo EC, Lei W. Finite element study of the mechanical response in spinal cord during the thoracolumbar burst fracture. *PLoS One* 2012;7(9):e41397 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0041397](https://doi.org/10.1371/journal.pone.0041397)] [Medline: [23028426](#)]
37. Li XF, Dai LY. Acute central cord syndrome: injury mechanisms and stress features. *Spine (Phila Pa 1976)* 2010 Sep 01;35(19):E955-E964. [doi: [10.1097/BRS.0b013e3181c94cb8](https://doi.org/10.1097/BRS.0b013e3181c94cb8)] [Medline: [20543769](#)]
38. Jannesar S, Nadler B, Sparrey CJ. The transverse isotropy of spinal cord white matter under dynamic load. *J Biomech Eng* 2016 Sep 01;138(9). [doi: [10.1115/1.4034171](https://doi.org/10.1115/1.4034171)] [Medline: [27428053](#)]
39. Khuyagbaatar B, Kim K, Man Park W, Hyuk Kim Y. Biomechanical behaviors in three types of spinal cord injury mechanisms. *J Biomech Eng* 2016 Aug 01;138(8). [doi: [10.1115/1.4033794](https://doi.org/10.1115/1.4033794)] [Medline: [27276391](#)]
40. Khuyagbaatar B, Kim K, Hyuk Kim Y. Effect of bone fragment impact velocity on biomechanical parameters related to spinal cord injury: a finite element study. *J Biomech* 2014 Aug 22;47(11):2820-2825. [doi: [10.1016/j.jbiomech.2014.04.042](https://doi.org/10.1016/j.jbiomech.2014.04.042)] [Medline: [24891036](#)]
41. Greaves CY, Gadala MS, Oxland TR. A three-dimensional finite element model of the cervical spine with spinal cord: an investigation of three injury mechanisms. *Ann Biomed Eng* 2008 Mar;36(3):396-405. [doi: [10.1007/s10439-008-9440-0](https://doi.org/10.1007/s10439-008-9440-0)] [Medline: [18228144](#)]
42. Czyz M, Scigala K, Jarmundowicz W, Beidziński R. The biomechanical analysis of the traumatic cervical spinal cord injury using finite element approach. *Acta Bioeng Biomech* 2008;10(1):43-54. [Medline: [18634353](#)]
43. Maikos JT, Qian Z, Metaxas D, Shreiber DI. Finite element analysis of spinal cord injury in the rat. *J Neurotrauma* 2008 Jul;25(7):795-816. [doi: [10.1089/neu.2007.0423](https://doi.org/10.1089/neu.2007.0423)] [Medline: [18627257](#)]
44. Scifert J, Totoribe K, Goel V, Huntzinger J. Spinal cord mechanics during flexion and extension of the cervical spine: a finite element study. *Pain Phys* 2002;5(4):394-400. [doi: [10.36076/ppj.2002/5/394](https://doi.org/10.36076/ppj.2002/5/394)]
45. Fournely M, Petit Y, Wagnac E, Evin M, Arnoux PJ. Effect of experimental, morphological and mechanical factors on the murine spinal cord subjected to transverse contusion: a finite element study. *PLoS One* 2020 May 11;15(5):e0232975 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0232975](https://doi.org/10.1371/journal.pone.0232975)] [Medline: [32392241](#)]
46. Persson C, Summers J, Hall RM. The importance of fluid-structure interaction in spinal trauma models. *J Neurotrauma* 2011 Jan;28(1):113-125. [doi: [10.1089/neu.2010.1332](https://doi.org/10.1089/neu.2010.1332)] [Medline: [21047151](#)]
47. Wilcox RK, Allen DJ, Hall RM, Limb D, Barton DC, Dickson RA. A dynamic investigation of the burst fracture process using a combined experimental and finite element approach. *Eur Spine J* 2004 Oct;13(6):481-488 [[FREE Full text](#)] [doi: [10.1007/s00586-003-0625-9](https://doi.org/10.1007/s00586-003-0625-9)] [Medline: [14714241](#)]
48. Liang D, Tu GJ, Han YX, Guo DW. Accurate simulation of the herniated cervical intervertebral disc using controllable expansion: a finite element study. *Comput Methods Biomech Biomed Eng* 2021 Jun;24(8):897-904. [doi: [10.1080/10255842.2020.1857745](https://doi.org/10.1080/10255842.2020.1857745)] [Medline: [33331162](#)]
49. Khuyagbaatar B, Kim K, Purevsuren T, Lee SH, Kim YH. Biomechanical effects on cervical spinal cord and nerve root following laminoplasty for ossification of the posterior longitudinal ligament in the cervical spine: a comparison between open-door and double-door laminoplasty using finite element analysis. *J Biomech Eng* 2018 Jul 01;140(7). [doi: [10.1115/1.4039826](https://doi.org/10.1115/1.4039826)] [Medline: [29677281](#)]
50. Okazaki T, Kanchiku T, Nishida N, Ichihara K, Sakuramoto I, Ohgi J, et al. Age-related changes of the spinal cord: a biomechanical study. *Exp Ther Med* 2018 Mar 24;15(3):2824-2829 [[FREE Full text](#)] [doi: [10.3892/etm.2018.5796](https://doi.org/10.3892/etm.2018.5796)] [Medline: [29599828](#)]
51. Nishida N, Kanchiku T, Kato Y, Imajo Y, Suzuki H, Yoshida Y, et al. Cervical ossification of the posterior longitudinal ligament: factors affecting the effect of posterior decompression. *J Spinal Cord Med* 2017 Jan;40(1):93-99 [[FREE Full text](#)] [doi: [10.1080/10790268.2016.1140392](https://doi.org/10.1080/10790268.2016.1140392)] [Medline: [26788904](#)]
52. Khuyagbaatar B, Kim K, Park WM, Kim YH. Biomechanical investigation of post-operative C5 palsy due to ossification of the posterior longitudinal ligament in different types of cervical spinal alignment. *J Biomech* 2017 May 24;57:54-61. [doi: [10.1016/j.jbiomech.2017.03.019](https://doi.org/10.1016/j.jbiomech.2017.03.019)] [Medline: [28427722](#)]
53. Khuyagbaatar B, Kim K, Park WM, Kim YH. Effect of posterior decompression extent on biomechanical parameters of the spinal cord in cervical ossification of the posterior longitudinal ligament. *Proc Inst Mech Eng H* 2016 Jun;230(6):545-552. [doi: [10.1177/0954411916637383](https://doi.org/10.1177/0954411916637383)] [Medline: [26951839](#)]
54. Nishida N, Kanchiku T, Kato Y, Imajo Y, Yoshida Y, Kawano S, et al. Cervical ossification of the posterior longitudinal ligament: biomechanical analysis of the influence of static and dynamic factors. *J Spinal Cord Med* 2015 Sep;38(5):593-598 [[FREE Full text](#)] [doi: [10.1179/2045772314Y.0000000221](https://doi.org/10.1179/2045772314Y.0000000221)] [Medline: [24964955](#)]
55. Nishida N, Kanchiku T, Kato Y, Imajo Y, Yoshida Y, Kawano S, et al. Biomechanical analysis of cervical myelopathy due to ossification of the posterior longitudinal ligament: effects of posterior decompression and kyphosis following decompression. *Exp Ther Med* 2014 May;7(5):1095-1099 [[FREE Full text](#)] [doi: [10.3892/etm.2014.1557](https://doi.org/10.3892/etm.2014.1557)] [Medline: [24940393](#)]

56. Kato Y, Kanchiku T, Imajo Y, Kimura K, Ichihara K, Kawano S, et al. Biomechanical study of the effect of degree of static compression of the spinal cord in ossification of the posterior longitudinal ligament. *J Neurosurg Spine* 2010 Mar;12(3):301-305. [doi: [10.3171/2009.9.SPINE09314](https://doi.org/10.3171/2009.9.SPINE09314)] [Medline: [20192631](https://pubmed.ncbi.nlm.nih.gov/20192631/)]
57. Khuyagbaatar B, Kim K, Park WM, Kim YH. Influence of sagittal and axial types of ossification of posterior longitudinal ligament on mechanical stress in cervical spinal cord: a finite element analysis. *Clin Biomech (Bristol, Avon)* 2015 Dec;30(10):1133-1139. [doi: [10.1016/j.clinbiomech.2015.08.013](https://doi.org/10.1016/j.clinbiomech.2015.08.013)] [Medline: [26351002](https://pubmed.ncbi.nlm.nih.gov/26351002/)]
58. Henao J, Labelle H, Arnoux PJ, Aubin CE. Biomechanical simulation of stresses and strains exerted on the spinal cord and nerves during scoliosis correction maneuvers. *Spine Deform* 2018 Jan;6(1):12-19. [doi: [10.1016/j.jspd.2017.04.008](https://doi.org/10.1016/j.jspd.2017.04.008)] [Medline: [29287811](https://pubmed.ncbi.nlm.nih.gov/29287811/)]
59. Henao J, Aubin CE, Labelle H, Arnoux PJ. Patient-specific finite element model of the spine and spinal cord to assess the neurological impact of scoliosis correction: preliminary application on two cases with and without intraoperative neurological complications. *Comput Methods Biomech Biomed Eng* 2016;19(8):901-910. [doi: [10.1080/10255842.2015.1075010](https://doi.org/10.1080/10255842.2015.1075010)] [Medline: [26324393](https://pubmed.ncbi.nlm.nih.gov/26324393/)]
60. Bertram CD. Evaluation by fluid/structure-interaction spinal-cord simulation of the effects of subarachnoid-space stenosis on an adjacent syrinx. *J Biomech Eng* 2010 Jun;132(6):061009. [doi: [10.1115/1.4001165](https://doi.org/10.1115/1.4001165)] [Medline: [20887034](https://pubmed.ncbi.nlm.nih.gov/20887034/)]
61. Bertram CD, Bilston LE, Stoodley MA. Tensile radial stress in the spinal cord related to arachnoiditis or tethering: a numerical model. *Med Biol Eng Comput* 2008 Jul;46(7):701-707. [doi: [10.1007/s11517-008-0332-0](https://doi.org/10.1007/s11517-008-0332-0)] [Medline: [18347831](https://pubmed.ncbi.nlm.nih.gov/18347831/)]
62. Kato Y, Kataoka H, Ichihara K, Imajo Y, Kojima T, Kawano S, et al. Biomechanical study of cervical flexion myelopathy using a three-dimensional finite element method. *J Neurosurg Spine* 2008 May;8(5):436-441. [doi: [10.3171/SPI/2008/8/5/436](https://doi.org/10.3171/SPI/2008/8/5/436)] [Medline: [18447689](https://pubmed.ncbi.nlm.nih.gov/18447689/)]
63. Kato Y, Kanchiku T, Imajo Y, Ichihara K, Kawano S, Hamanama D, et al. Flexion model simulating spinal cord injury without radiographic abnormality in patients with ossification of the longitudinal ligament: the influence of flexion speed on the cervical spine. *J Spinal Cord Med* 2009;32(5):555-559 [FREE Full text] [doi: [10.1080/10790268.2009.11754557](https://doi.org/10.1080/10790268.2009.11754557)] [Medline: [20025151](https://pubmed.ncbi.nlm.nih.gov/20025151/)]
64. Mitsuhashi N, Fujieda K, Tamura T, Kawamoto S, Takagi T, Okubo K. BodyParts3D: 3D structure database for anatomical concepts. *Nucleic Acids Res* 2009 Jan 01;37(Database issue):D782-D785 [FREE Full text] [doi: [10.1093/nar/gkn613](https://doi.org/10.1093/nar/gkn613)] [Medline: [18835852](https://pubmed.ncbi.nlm.nih.gov/18835852/)]
65. Frostell A, Hakim R, Thelin EP, Mattsson P, Svensson M. A review of the segmental diameter of the healthy human spinal cord. *Front Neurol* 2016 Dec 23;7:238. [doi: [10.3389/fneur.2016.00238](https://doi.org/10.3389/fneur.2016.00238)] [Medline: [28066322](https://pubmed.ncbi.nlm.nih.gov/28066322/)]
66. Barker JB, Cronin DS, Chandrashekar N. High rotation rate behavior of cervical spine segments in flexion and extension. *J Biomech Eng* 2014 Dec;136(12):121004. [doi: [10.1115/1.4028107](https://doi.org/10.1115/1.4028107)] [Medline: [25070575](https://pubmed.ncbi.nlm.nih.gov/25070575/)]
67. Onan OA, Heggeness MH, Hipp JA. A motion analysis of the cervical facet joint. *Spine (Phila Pa 1976)* 1998 Feb 15;23(4):430-439. [doi: [10.1097/00007632-199802150-00005](https://doi.org/10.1097/00007632-199802150-00005)] [Medline: [9516697](https://pubmed.ncbi.nlm.nih.gov/9516697/)]
68. Grauer JN, Panjabi MM, Cholewicki J, Nibu K, Dvorak J. Whiplash produces an S-shaped curvature of the neck with hyperextension at lower levels. *Spine (Phila Pa 1976)* 1997 Nov 01;22(21):2489-2494. [doi: [10.1097/00007632-199711010-00005](https://doi.org/10.1097/00007632-199711010-00005)] [Medline: [9383854](https://pubmed.ncbi.nlm.nih.gov/9383854/)]
69. Moroney SP, Schultz AB, Miller JA, Andersson GB. Load-displacement properties of lower cervical spine motion segments. *J Biomech* 1988;21(9):769-779 [FREE Full text] [doi: [10.1016/0021-9290\(88\)90285-0](https://doi.org/10.1016/0021-9290(88)90285-0)] [Medline: [3053721](https://pubmed.ncbi.nlm.nih.gov/3053721/)]
70. Stoner KE, Abode-Iyamah KO, Magnotta VA, Howard MA, Grosland NM. Measurement of in vivo spinal cord displacement and strain fields of healthy and myelopathic cervical spinal cord. *J Neurosurg Spine* 2019 Mar 22;31(1):53-59. [doi: [10.3171/2018.12.SPINE18989](https://doi.org/10.3171/2018.12.SPINE18989)] [Medline: [30901756](https://pubmed.ncbi.nlm.nih.gov/30901756/)]
71. Ozawa H, Matsumoto T, Ohashi T, Sato M, Kokubun S. Mechanical properties and function of the spinal pia mater. *J Neurosurg Spine* 2004 Jul;1(1):122-127. [doi: [10.3171/spi.2004.1.1.0122](https://doi.org/10.3171/spi.2004.1.1.0122)] [Medline: [15291032](https://pubmed.ncbi.nlm.nih.gov/15291032/)]
72. Ichihara K, Taguchi T, Sakuramoto I, Kawano S, Kawai S. Mechanism of the spinal cord injury and the cervical spondylotic myelopathy: new approach based on the mechanical features of the spinal cord white and gray matter. *J Neurosurg* 2003 Oct;99(3 Suppl):278-285. [doi: [10.3171/spi.2003.99.3.0278](https://doi.org/10.3171/spi.2003.99.3.0278)] [Medline: [14563145](https://pubmed.ncbi.nlm.nih.gov/14563145/)]
73. Ichihara K, Taguchi T, Shimada Y, Sakuramoto I, Kawano S, Kawai S. Gray matter of the bovine cervical spinal cord is mechanically more rigid and fragile than the white matter. *J Neurotrauma* 2001 Mar;18(3):361-367. [doi: [10.1089/08977150151071053](https://doi.org/10.1089/08977150151071053)] [Medline: [11284555](https://pubmed.ncbi.nlm.nih.gov/11284555/)]
74. Bilston LE, Thibault LE. The mechanical properties of the human cervical spinal cord. *In Vitro. Ann Biomed Eng* 1995 Sep;24(S1):67-74. [doi: [10.1007/bf02770996](https://doi.org/10.1007/bf02770996)]
75. Tunturi AR. Elasticity of the spinal cord, pia, and denticulate ligament in the dog. *J Neurosurg* 1978 Jun;48(6):975-979. [doi: [10.3171/jns.1978.48.6.0975](https://doi.org/10.3171/jns.1978.48.6.0975)] [Medline: [660249](https://pubmed.ncbi.nlm.nih.gov/660249/)]
76. Ozawa H, Matsumoto T, Ohashi T, Sato M, Kokubun S. Comparison of spinal cord gray matter and white matter softness: measurement by pipette aspiration method. *J Neurosurg* 2001 Oct;95(2 Suppl):221-224. [doi: [10.3171/spi.2001.95.2.0221](https://doi.org/10.3171/spi.2001.95.2.0221)] [Medline: [11599840](https://pubmed.ncbi.nlm.nih.gov/11599840/)]
77. Kimpara H, Nakahira Y, Iwamoto M, Miki K, Ichihara K, Kawano SI, et al. Investigation of anteroposterior head-neck responses during severe frontal impacts using a brain-spinal cord complex FE model. *Stapp Car Crash J* 2006 Nov;50:509-544. [doi: [10.4271/2006-22-0019](https://doi.org/10.4271/2006-22-0019)] [Medline: [17311175](https://pubmed.ncbi.nlm.nih.gov/17311175/)]

78. Hong JY, Suh SW, Park SY, Modi HN, Rhyu IJ, Kwon S, et al. Analysis of dural sac thickness in human spine-cadaver study with confocal infrared laser microscope. *Spine J* 2011 Dec;11(12):1121-1127. [doi: [10.1016/j.spinee.2011.11.001](https://doi.org/10.1016/j.spinee.2011.11.001)] [Medline: [22172494](https://pubmed.ncbi.nlm.nih.gov/22172494/)]
79. Zarzur E. Mechanical properties of the human lumbar dura mater. *Arq Neuropsiquiatr* 1996 Sep;54(3):455-460. [doi: [10.1590/s0004-282x1996000300015](https://doi.org/10.1590/s0004-282x1996000300015)] [Medline: [9109991](https://pubmed.ncbi.nlm.nih.gov/9109991/)]
80. Persson C, Evans S, Marsh R, Summers JL, Hall RM. Poisson's ratio and strain rate dependency of the constitutive behavior of spinal dura mater. *Ann Biomed Eng* 2010 Mar 20;38(3):975-983. [doi: [10.1007/s10439-010-9924-6](https://doi.org/10.1007/s10439-010-9924-6)] [Medline: [20087767](https://pubmed.ncbi.nlm.nih.gov/20087767/)]
81. Wilcox RK, Bilston LE, Barton DC, Hall RM. Mathematical model for the viscoelastic properties of dura mater. *J Orthop Sci* 2003 May;8(3):432-434. [doi: [10.1007/s10776-003-0644-9](https://doi.org/10.1007/s10776-003-0644-9)] [Medline: [12768491](https://pubmed.ncbi.nlm.nih.gov/12768491/)]
82. Tubbs RS, Salter G, Grabb PA, Oakes WJ. The denticulate ligament: anatomy and functional significance. *J Neurosurg* 2001 Apr;94(2 Suppl):271-275. [doi: [10.3171/spi.2001.94.2.0271](https://doi.org/10.3171/spi.2001.94.2.0271)] [Medline: [11302630](https://pubmed.ncbi.nlm.nih.gov/11302630/)]
83. Bloomfield I, Johnston IH, Bilston LE. Effects of proteins, blood cells and glucose on the viscosity of cerebrospinal fluid. *Pediatr Neurosurg* 1998 May 3;28(5):246-251. [doi: [10.1159/000028659](https://doi.org/10.1159/000028659)] [Medline: [9732257](https://pubmed.ncbi.nlm.nih.gov/9732257/)]
84. Brydon HL, Hayward R, Harkness W, Bayston R. Physical properties of cerebrospinal fluid of relevance to shunt function. 1: the effect of protein upon CSF viscosity. *Br J Neurosurg* 1995 Jul 06;9(5):639-644. [doi: [10.1080/02688699550040927](https://doi.org/10.1080/02688699550040927)] [Medline: [8561936](https://pubmed.ncbi.nlm.nih.gov/8561936/)]
85. Przybylski GJ, Carlin GJ, Patel PR, Woo SL. Human anterior and posterior cervical longitudinal ligaments possess similar tensile properties. *J Orthop Res* 1996 Nov;14(6):1005-1008. [doi: [10.1002/jor.1100140623](https://doi.org/10.1002/jor.1100140623)] [Medline: [8982146](https://pubmed.ncbi.nlm.nih.gov/8982146/)]
86. Yoganandan N, Kumaresan S, Pintar FA. Geometric and mechanical properties of human cervical spine ligaments. *J Biomech Eng* 2000 Dec;122(6):623-629. [doi: [10.1115/1.1322034](https://doi.org/10.1115/1.1322034)] [Medline: [11192384](https://pubmed.ncbi.nlm.nih.gov/11192384/)]
87. Yoganandan N, Pintar F, Butler J, Reinartz J, Sances AJ, Larson SJ. Dynamic response of human cervical spine ligaments. *Spine (Phila Pa 1976)* 1989 Oct;14(10):1102-1110. [doi: [10.1097/00007632-198910000-00013](https://doi.org/10.1097/00007632-198910000-00013)] [Medline: [2588060](https://pubmed.ncbi.nlm.nih.gov/2588060/)]
88. Kulkarni VA, Massie JB, Zauner F, Murphy M, Akeson WH. Novel biomechanical quantification methodology for lumbar intraforaminal spinal nerve adhesion in a laminectomy and disc injury rat model. *J Neurosci Methods* 2007 Oct 15;166(1):20-23. [doi: [10.1016/j.jneumeth.2007.06.025](https://doi.org/10.1016/j.jneumeth.2007.06.025)] [Medline: [17689664](https://pubmed.ncbi.nlm.nih.gov/17689664/)]
89. Singh A, Lu Y, Chen C, Cavanaugh JM. Mechanical properties of spinal nerve roots subjected to tension at different strain rates. *J Biomech* 2006;39(9):1669-1676. [doi: [10.1016/j.jbiomech.2005.04.023](https://doi.org/10.1016/j.jbiomech.2005.04.023)] [Medline: [15996674](https://pubmed.ncbi.nlm.nih.gov/15996674/)]
90. Spilker RL, Jakobs DM, Schultz AB. Material constants for a finite element model of the intervertebral disk with a fiber composite annulus. *J Biomech Eng* 1986 Feb;108(1):1-11. [doi: [10.1115/1.3138575](https://doi.org/10.1115/1.3138575)] [Medline: [3959546](https://pubmed.ncbi.nlm.nih.gov/3959546/)]
91. Nishida N, Kanchiku T, Imajo Y, Suzuki H, Yoshida Y, Kato Y, et al. Stress analysis of the cervical spinal cord: impact of the morphology of spinal cord segments on stress. *J Spinal Cord Med* 2016 Feb 25;39(3):327-334. [doi: [10.1177/2045772315y.0000000012](https://doi.org/10.1177/2045772315y.0000000012)]
92. Persson C, Evans S, Marsh R, Summers JL, Hall RM. Poisson's ratio and strain rate dependency of the constitutive behavior of spinal dura mater. *Ann Biomed Eng* 2010 Mar;38(3):975-983. [doi: [10.1007/s10439-010-9924-6](https://doi.org/10.1007/s10439-010-9924-6)] [Medline: [20087767](https://pubmed.ncbi.nlm.nih.gov/20087767/)]
93. Jones CF, Kroeker SG, Crompton PA, Hall RM. The effect of cerebrospinal fluid on the biomechanics of spinal cord: an ex vivo bovine model using bovine and physical surrogate spinal cord. *Spine* 2008 Aug 1;33(17):E580-E588. [doi: [10.1097/brs.0b013e31817ecc57](https://doi.org/10.1097/brs.0b013e31817ecc57)]
94. Taniyama T, Hirai T, Yamada T, Yuasa M, Enomoto M, Yoshii T, et al. Modified K-line in magnetic resonance imaging predicts insufficient decompression of cervical laminoplasty. *Spine* 2013;38(6):496-501. [doi: [10.1097/brs.0b013e318273a4f7](https://doi.org/10.1097/brs.0b013e318273a4f7)]
95. Nouri A, Tetreault L, Singh A, Karadimas SK, Fehlings MG. Degenerative cervical myelopathy: epidemiology, genetics, and pathogenesis. *Spine* 2015;40(12):E675-E693. [doi: [10.1097/brs.0000000000000913](https://doi.org/10.1097/brs.0000000000000913)]
96. Gondar R, Nouri A, Jannelli G, Schaller K, Tessitore E. Does spondylolisthesis affect severity and outcome of degenerative cervical myelopathy? A systematic review and meta-analysis. *Global Spine J* 2021 Sep 16;11(7):1134-1141 [FREE Full text] [doi: [10.1177/2192568220960452](https://doi.org/10.1177/2192568220960452)] [Medline: [33063537](https://pubmed.ncbi.nlm.nih.gov/33063537/)]
97. Park D, Kim BH, Cho J, Yang JW, Yang DH, Kim MS, et al. Diagnostic role of flexion-extension central motor conduction time in cervical spondylotic myelopathy. *Spine (Phila Pa 1976)* 2021 Nov 15;46(22):1564-1571 [FREE Full text] [doi: [10.1097/BRS.0000000000003706](https://doi.org/10.1097/BRS.0000000000003706)] [Medline: [32991514](https://pubmed.ncbi.nlm.nih.gov/32991514/)]
98. Kolcun JP, Chieng LO, Madhavan K, Wang MY. The role of dynamic magnetic resonance imaging in cervical spondylotic myelopathy. *Asian Spine J* 2017 Dec;11(6):1008-1015 [FREE Full text] [doi: [10.4184/asj.2017.11.6.1008](https://doi.org/10.4184/asj.2017.11.6.1008)] [Medline: [29279758](https://pubmed.ncbi.nlm.nih.gov/29279758/)]
99. Ghogawala Z, Terrin N, Dunbar MR, Breeze JL, Freund KM, Kanter AS, et al. Effect of ventral vs dorsal spinal surgery on patient-reported physical functioning in patients with cervical spondylotic myelopathy: a randomized clinical trial. *JAMA* 2021 Mar 09;325(10):942-951 [FREE Full text] [doi: [10.1001/jama.2021.1233](https://doi.org/10.1001/jama.2021.1233)] [Medline: [33687463](https://pubmed.ncbi.nlm.nih.gov/33687463/)]
100. Wang J, Wo J, Wen J, Zhang L, Xu W, Wang X. Laminoplasty versus laminectomy with fusion for treatment of multilevel cervical compressive myelopathy: an updated meta-analysis. *Postgrad Med J* 2022 Sep 01;98(1163):680-688. [doi: [10.1136/postgradmedj-2020-139667](https://doi.org/10.1136/postgradmedj-2020-139667)] [Medline: [37062984](https://pubmed.ncbi.nlm.nih.gov/37062984/)]
101. Lao L, Zhong G, Li X, Qian L, Liu Z. Laminoplasty versus laminectomy for multi-level cervical spondylotic myelopathy: a systematic review of the literature. *J Orthop Surg Res* 2013 Dec 01;8(1):45. [doi: [10.1186/1749-799x-8-45](https://doi.org/10.1186/1749-799x-8-45)]

102. Cheung ZB, Gidumal S, White S, Shin J, Phan K, Osman N, et al. Comparison of anterior cervical discectomy and fusion with a stand-alone interbody cage versus a conventional cage-plate technique: a systematic review and meta-analysis. *Global Spine J* 2019 Jun 17;9(4):446-455 [FREE Full text] [doi: [10.1177/2192568218774576](https://doi.org/10.1177/2192568218774576)] [Medline: [31218204](https://pubmed.ncbi.nlm.nih.gov/31218204/)]
103. Akter F, Yu X, Qin X, Yao S, Nikrouz P, Syed YA, et al. The pathophysiology of degenerative cervical myelopathy and the physiology of recovery following decompression. *Front Neurosci* 2020 Apr 30;14:138 [FREE Full text] [doi: [10.3389/fnins.2020.00138](https://doi.org/10.3389/fnins.2020.00138)] [Medline: [32425740](https://pubmed.ncbi.nlm.nih.gov/32425740/)]
104. Grodzinski B, Durham R, Mowforth O, Stubbs D, Kotter MR, Davies BM. The effect of ageing on presentation, management and outcomes in degenerative cervical myelopathy: a systematic review. *Age Ageing* 2021 May 05;50(3):705-715. [doi: [10.1093/ageing/afaa236](https://doi.org/10.1093/ageing/afaa236)] [Medline: [33219816](https://pubmed.ncbi.nlm.nih.gov/33219816/)]
105. Mikulis DJ, Wood ML, Zerdoner OA, Poncelet BP. Oscillatory motion of the normal cervical spinal cord. *Radiology* 1994 Jul;192(1):117-121. [doi: [10.1148/radiology.192.1.8208922](https://doi.org/10.1148/radiology.192.1.8208922)] [Medline: [8208922](https://pubmed.ncbi.nlm.nih.gov/8208922/)]
106. Hupp M, Pfender N, Vallotton K, Rosner J, Friedl S, Zipser CM, et al. The restless spinal cord in degenerative cervical myelopathy. *Am J Neuroradiol* 2021 Feb 04;42(3):597-609. [doi: [10.3174/ajnr.a6958](https://doi.org/10.3174/ajnr.a6958)]
107. Vavasour IM, Meyers SM, MacMillan EL, Madler B, Li DK, Rauscher A, et al. Increased spinal cord movements in cervical spondylotic myelopathy. *Spine J* 2014 Oct 01;14(10):2344-2354. [doi: [10.1016/j.spinee.2014.01.036](https://doi.org/10.1016/j.spinee.2014.01.036)] [Medline: [24462810](https://pubmed.ncbi.nlm.nih.gov/24462810/)]
108. Koliass AG, Honeybul S. *Traumatic Brain Injury: Science, Practice, Evidence and Ethics*. Cham, Switzerland: Springer; 2021.

Abbreviations

ACDF: anterior cervical discectomy and fusion

CSF: cerebrospinal fluid

DCM: degenerative cervical myelopathy

FEA: finite element analysis

IVD: intervertebral disk

MRI: magnetic resonance imaging

OPLL: ossification of the posterior longitudinal ligament

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

SCI: spinal cord injury

SWiM: Synthesis Without Meta-Analysis

Edited by T Leung; submitted 13.04.23; peer-reviewed by M Arab-Zozani, A Perez Sanpablo; comments to author 28.08.23; revised version received 31.10.23; accepted 15.02.24; published 28.03.24.

Please cite as:

Davies B, Schaefer S, Rafati Fard A, Newcombe V, Sutcliffe M

Finite Element Analysis for Degenerative Cervical Myelopathy: Scoping Review of the Current Findings and Design Approaches, Including Recommendations on the Choice of Material Properties

JMIR Biomed Eng 2024;9:e48146

URL: <https://biomedeng.jmir.org/2024/1/e48146>

doi: [10.2196/48146](https://doi.org/10.2196/48146)

PMID: [38875683](https://pubmed.ncbi.nlm.nih.gov/38875683/)

©Benjamin Davies, Samuel Schaefer, Amir Rafati Fard, Virginia Newcombe, Michael Sutcliffe. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 28.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Review

Trends in South Korean Medical Device Development for Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder: Narrative Review

Yunah Cho¹, MPH; Sharon L Talboys², MPH, PhD

¹Division of Public Health, Department of Family and Preventive Medicine, University of Utah Asia Campus, Incheon, Republic of Korea

²Division of Public Health, Department of Family and Preventive Medicine, University of Utah School of Medicine, Salt Lake City, UT, United States

Corresponding Author:

Yunah Cho, MPH

Division of Public Health, Department of Family and Preventive Medicine

University of Utah Asia Campus

119-3 Songdomunhwa-ro, Yeonsu-gu

Incheon, 21985

Republic of Korea

Phone: 82 032 626 6901

Fax: 82 032 626 6010

Email: yunah.cho@utah.edu

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are among the most prevalent mental disorders among school-aged youth in South Korea and may play a role in the increasing pressures on teachers and school-based special education programming. A lack of support for special education; tensions between teachers, students, and parents; and limited backup for teacher absences are common complaints among Korean educators. New innovations in technology to screen and treat ADHD and ASD may offer relief to students, parents, and teachers through earlier and efficient diagnosis; access to treatment options; and ultimately, better-managed care and expectations.

Objective: This narrative literature review provides an account of medical device use and development in South Korea for the diagnosis and management of ADHD and ASD and highlights research gaps.

Methods: A narrative review was conducted across 4 databases (PubMed, Korean National Assembly Library, Scopus, and PsycINFO). Journal articles, dissertations, and government research and development reports were included if they discussed medical devices for ADHD and ASD. Only Korean or English papers were included. Resources were excluded if they did not correspond to the research objective or did not discuss at least 1 topic about medical devices for ADHD and ASD. Journal articles were excluded if they were not peer reviewed. Resources were limited to publications between 2013 and July 22, 2024.

Results: A total of 1794 records about trends in Korean medical device development were categorized into 2 major groups: *digital therapeutics* and *traditional therapy*. Digital therapeutics resulted in 5 subgroups: *virtual reality and artificial intelligence*, *machine learning and robot*, *gaming and visual contents*, *eye-feedback and movement intervention*, and *electroencephalography and neurofeedback*. Traditional therapy resulted in 3 subgroups: *cognitive behavioral therapy and working memory*; *diagnosis and rating scale*; and *musical, literary therapy, and mindfulness-based stress reduction*. Digital therapeutics using artificial intelligence, machine learning, and electroencephalography technologies account for the biggest portions of development in South Korea, rather than traditional therapies. Most resources, 94.15% (1689/1794), were from the Korean National Assembly Library.

Conclusions: Limitations include small sizes of populations to conclude findings in many articles, a lower number of articles discussing medical devices for ASD, and a majority of articles being dissertations. Emerging digital medical devices and those integrated with traditional therapies are important solutions to reducing the prevalence rates of ADHD and ASD in South Korea by promoting early diagnosis and intervention. Furthermore, their application will relieve pressures on teachers and school-based special education programming by providing direct supporting resources to students with ADHD or ASD. Future development of medical devices for ADHD and ASD is predicted to heavily rely on digital technologies, such as those that sense people's behaviors, eye movement, and brainwaves.

(JMIR Biomed Eng 2024;9:e60399) doi:[10.2196/60399](https://doi.org/10.2196/60399)

KEYWORDS

ADHD; attention-deficit/hyperactivity disorder; ASD; autism spectrum disorder; medical device; digital therapeutics

Introduction

Background

Attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are some of the most prevalent mental disorders among school-aged youth in South Korea. Insufficient support for those with ADHD or ASD affects their delayed improvement, and this circumstance may play a role in the increasing pressures on teachers and school-based special education programming. As teachers are the second most important people who impact children's early diagnosis and intervention [1], teachers are under increasing pressure in South Korea from parents and substandard special education resources, leading them to protest [2]. The protests were prompted by the news of a teacher who resorted to suicide over excessive complaints from demanding parents while also trying to manage students [3]. Sadly, this tragedy was followed by several more incidents of teacher suicides [4]. The lack of support for special education; tensions between teachers, students, and parents; and the lack of backup for teacher absences are common complaints among Korean educators [4]. New innovations in technology to screen and treat ADHD and ASD may offer some relief to students, parents, and teachers through earlier and efficient diagnosis; access to treatment options; and ultimately, better-managed care and expectations.

Prevalence of ADHD and ASD

ADHD is recognized by an ongoing pattern of inattention and hyperactivity-impulsivity that interferes with development or functioning [5]. ASD is defined as a developmental and neurological disorder that affects how people communicate with others, interact, behave, and learn [6]. The number of patients with ADHD in South Korea has consistently increased, and the total number has risen by 250% in 2022 [7]. Among this entire population, people aged between 0 and 19 years accounted for the majority of cases, ranging from 57% to 85% from 2018 to 2022 [8]. The prevalence of ASD in 2021 was 12.8%, which translated to roughly 32,000 individuals [9]. The rate has increased by 4.3% since 2010 [9]. According to the database of registered people with developmental disabilities in June 2021, the Ministry of Health and Welfare of South Korea announced that 56.7% of the population with ASD were young individuals aged between 0 and 19 years [10].

Objective

To set up improved special education systems for young people with ADHD or ASD, experts claim that innovational medical devices for ADHD and ASD are crucial to treating them in a timely and proper manner [11,12]. While diverse types of medical devices exist, including devices for assessment, screening, and training, few studies have examined the use of these medical devices in South Korea or trends in the development of new devices for ADHD and ASD in South Korea. This study provides a review of the literature focused on gaps in the research related to medical device use and

development in South Korea for the diagnosis and management of ADHD and ASD.

Methods

Search Strategy

A narrative review was conducted to examine the trends in Korean medical device development focusing on medical equipment for ADHD and ASD, using the National Assembly Library, PubMed, Scopus, and PsycINFO. The detailed search terms were presented in the *Search Strategies* section in [Multimedia Appendix 1](#). Data and studies were retrieved and reviewed after screening years and language. Key search terms included: *ADHD*, *ASD* or *autism*, *early*, *diagnosis*, *treatment*, *screening*, *medical device*, *intervention*, and *training*. The list of references from the 4 databases was cross-checked to identify duplicates.

Eligibility Criteria

Journal articles and dissertations were included if they discussed diverse types of medical devices for ADHD and ASD, were peer reviewed, and were published in 2013 or later. Government research and development project reports were also included if they discussed relevant topics and were published in 2013 or later. Only Korean or English papers were included. The expected outcome from the included sources was updated information on Korean medical equipment for ADHD and ASD and an emphasis on examining the trends in Korean medical equipment for ADHD and ASD. Non-peer-reviewed interview articles were also included.

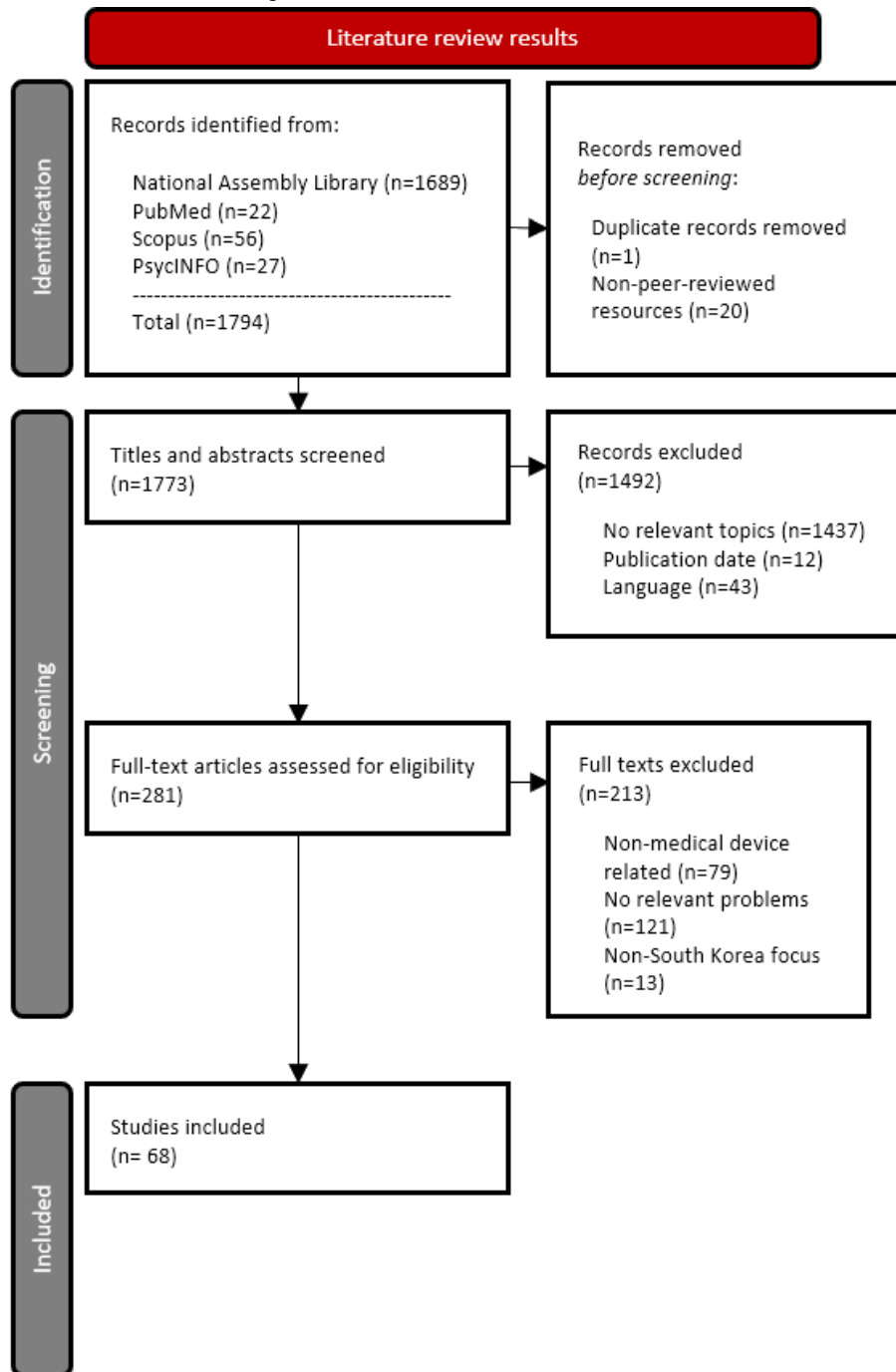
Resources were excluded if they did not correspond to the research objective or did not discuss at least 1 topic about medical devices for ADHD and ASD in the title or abstract. Journal articles were excluded if they were not peer reviewed, published before 2013, or written other than Korean or English. Detailed eligibility criteria for study inclusion are described in [Multimedia Appendix 1](#).

Results

Selection of Sources of Evidence

The search identified 94.15% (1689/1794) records through the National Assembly Library, 1.23% (22/1794) records through PubMed, 3.12% (56/1794) records through Scopus, and 1.5% (27/1794) records through PsycINFO ([Figure 1](#)). Of the total 1794 records, 1 (0.1%) duplicate record was found, and 20 (1.1%) records were removed after non-peer-reviewed resources were screened. Of the remaining 1773 full-text records, 84.1% (n=1492) were excluded based on their content. Among the remaining 281 records, 213 were excluded: 37.1% (79/213) of resources were unrelated to medical devices, 56.8% (121/213) were irrelevant problems, and 6.1% (13/213) were non-South Korea focused. Thus, 24.1% (68/281) of records were included in this paper. [Figure 1](#) depicts a flow diagram describing the selection of sources of evidence.

Figure 1. Flow diagram of narrative review describing the selection of sources of evidence.



Synthesis of Results

Overview

After a review of the records, 9 categories were developed post hoc to describe trends in Korean medical device development (Figure 2). The 9 groups included digital therapeutics; virtual reality (VR) and artificial intelligence (AI); machine learning and robot; gaming and visual contents; eye-feedback and

movement intervention; electroencephalography and neurofeedback; cognitive behavioral therapy (CBT) and working memory; diagnosis and rating scale; and musical, literary therapy, and mindfulness-based stress reduction (MBSR). These 9 groups were recategorized into 2 big groups: digital therapeutics and traditional therapy.

Table 1 summarizes the selected resources on the trends in Korean medical device development for ADHD and ASD.

Figure 2. Nine groups of the trends in Korean medical device development for attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD). MBSR: mindfulness-based stress reduction.

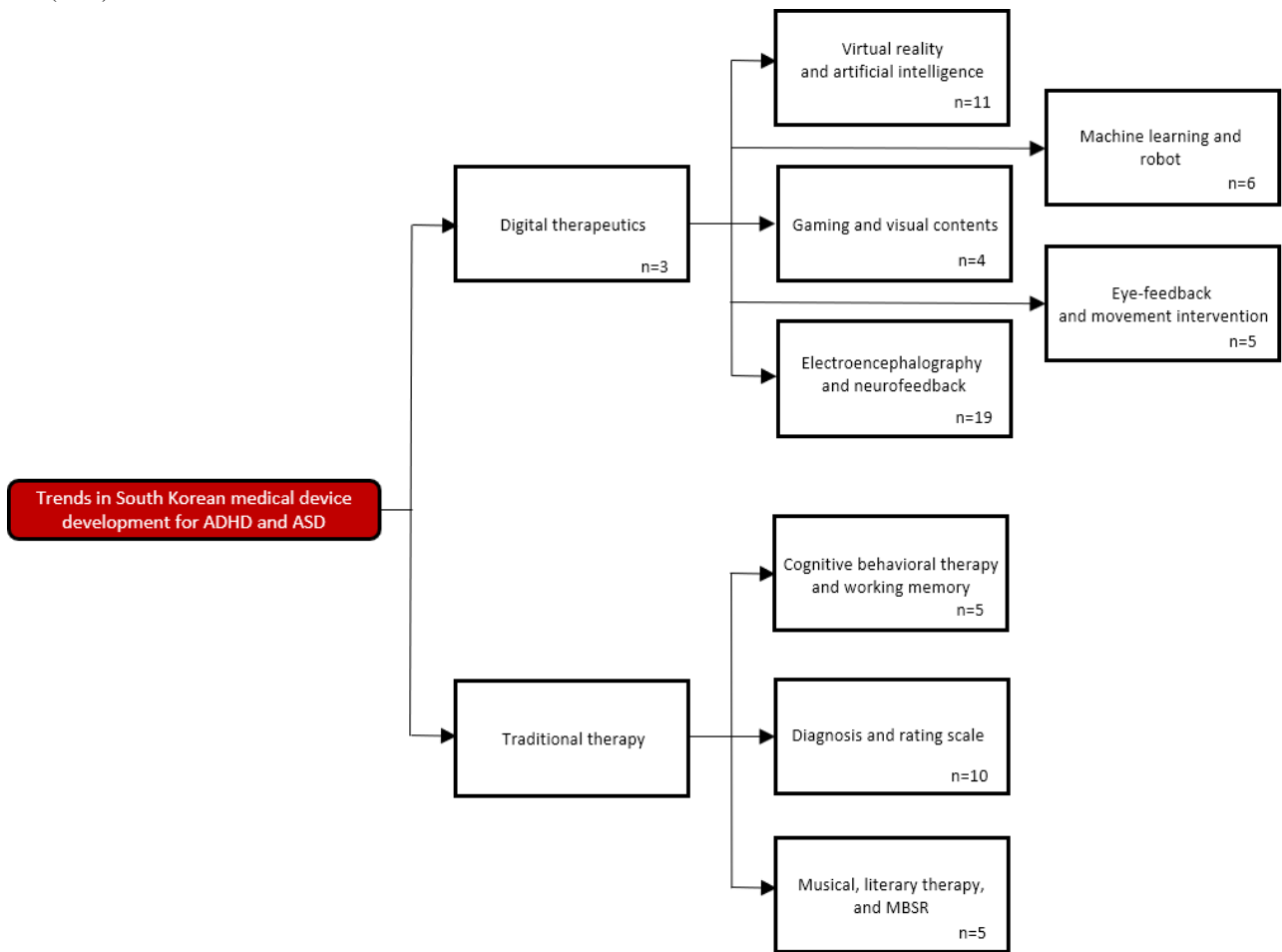


Table 1. Overview of the selected resources.

Study, year	Title	Country	Type
Choi [13], 2022	Study of textbooks on the principles of digital therapeutics to respond to ADHD and digital drama	South Korea	Doctoral dissertation
Lee [14], 2022	The efficacy of digital therapeutics for the treatment of attention deficit hyperactivity: a systematic review and meta-analysis	South Korea	Master's thesis
Son et al [15], 2023	Current status and outlook for digital therapeutics	South Korea	Research report
Rashid et al [16], 2024	Power of alignment: exploring the effect of face alignment on ASD diagnosis using facial images	Malaysia	Peer-reviewed article
Kim [17], 2019	To strive for the universalization of virtual reality therapy programs	South Korea	Interview
Korea Electronics Technology Institute [18], 2020	VR/AR platform technology based on bio-signal for mental health of kids/silver generation	South Korea	Government R and D ^a project report
Megerian et al [19], 2022	Evaluation of an artificial intelligence-based medical device for diagnosis for autism spectrum disorder	United States	Peer-reviewed article
Son [20], 2022	Towards standardizing attention-deficit/hyperactivity disorder diagnosis- a virtual reality, artificial intelligence application	South Korea	Master's thesis
Park et al [21], 2019	Design and implementation of VR-based life care contents for attention deficit hyperactivity disorder (ADHD)	South Korea	Peer-reviewed article
Ryu [22], 2022	Implications of VR-based psychotherapeutic effects for ADHD and CD among adolescents	South Korea	Peer-reviewed article
Ryu and Hwang [23], 2021	Artificial intelligence analysis of biosignals for automated detection and automated diagnosis of ADHD and CD	South Korea	Peer-reviewed article
Voss et al [24], 2019	Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder a randomized clinical trial	United States	Peer-reviewed article
Yoo [25], 2020	virtual reality based digital therapeutics system for diagnosing attention-deficit hyperactivity disorder	South Korea	Master's thesis
Yonsei University Office of Research Affairs [26], 2019	Development of mobile VR neurocognitive battery and establishment of database, implementation of AI-based early diagnosis/prevention system for cognitive control vulnerable groups utilizing digital representation modeling	South Korea	Government R and D project report
Imbiriba et al [27], 2023	Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism	United States	Peer-reviewed article
Kim et al [28], 2016	Exploring the applicability of Tele-presence robot intervention for at-risk children with ADHD	South Korea	Peer-reviewed article
Lee [29], 2022	Development of a contract-less sensing system and a classifier using deep learning for robot-based ADHD screening	South Korea	Doctoral dissertation
Lee et al [30], 2021	Development of a machine-learning predictive model for first-grade children at risk for ADHD	South Korea	Peer-reviewed article
Shin et al [31], 2018	Exploring the performance difference on the active based task with a robot for ADHD screening	South Korea	Peer-reviewed article
Yeom [32], 2018	Supervised classification of childhood ADHD using robot-assisted tests	South Korea	Master's thesis
Jung [33], 2022	ADHD can be treated like playing a game in daily life	South Korea	Interview
Lee and Lim [34], 2018	A study on the effect of communication functional board game on self-control, self-esteem, family function and peer relationship of ADHD children	South Korea	Peer-reviewed article
Park [35], 2019	To improve the concentration of ADHD children study on functional games	South Korea	Master's thesis
Sungkyunkwan University Cooperation Center [36], 2020	A study on the therapeutic applications of digital games	South Korea	Government R and D project report
Kim [37], 2018	Development of movement intervention visualization contents to improve behavior of ASD and ADHD	South Korea	Master's thesis

Study, year	Title	Country	Type
Kim [38], 2019	The characteristic of attentional networks in sluggish cognitive tempo: the effect of eye-feedback training on orienting attention in individuals with SCT	South Korea	Doctoral dissertation
Sandbank and Cascio [39], 2018	Using a motion-tracking device to facilitate motion control in children with ASD for neuroimaging	United States	Peer-reviewed article
Yoo and Kim [40], 2015	A preliminary study on the development of the focus reaction time tests	South Korea	Peer-reviewed article
Yoo et al [41], 2024	Development of an innovative approach using portable eye tracking to assist ADHD screening: a machine learning study	South Korea	Peer-reviewed article
Alhassan et al [42], 2023	Energy-efficient EEG-based scheme for autism spectrum disorder detection using wearable sensors	United States	Peer-reviewed article
Bhattacharyya et al [43], 2022	Integration of electroencephalogram (EEG) and motion tracking sensors for objective measure of attention-deficit hyperactivity disorder (MAHD) in preschoolers	United States	Peer-reviewed article
Hong et al [44], 2013	Development of brain imaging diagnosis and brain-based training programs for ADHD students	South Korea	Government R and D project report
Hong et al [45], 2014	Validation of the effectiveness of brain-based training programs for ADHD students	South Korea	Government R and D project report
Kang [46], 2013	Brain music as a potential tool for diagnosing attention-deficit/hyperactivity disorder (ADHD)	South Korea	Master's thesis
Kim [47], 2017	The effects of neurofeedback training and executive function improvement programs on attention and brain function quotient of elementary school children	South Korea	Master's thesis
Kim [48], 2021	Machine learning-based EEG classification for assisting the diagnosis of ADHD in children	South Korea	Peer-reviewed article
Kim [49], 2022	Deep learning approach on the improvement of diagnosing ADHD with fMRI	South Korea	Master's thesis
Kim et al [50], 2015	Clinical significance for neurofeedback training of children with attention-deficit/hyperactivity disorder	South Korea	Peer-reviewed article
Kim et al [51], 2022	The classification scheme of ADHD for children based on the CNN model	South Korea	Peer-reviewed article
Lee [52], 2013	The effects of the neurofeedback training on the attention in adolescents with autism spectrum disorder	South Korea	Master's thesis
Lee [53], 2020	Effects of neurofeedback brain wave training on the attention concentration and language development of children delayed in language development	South Korea	Master's thesis
Lee [54], 2022	The effect of EEG training through neurofeedback on attention and pragmatic language ability in children with ADHD prone language delay	South Korea	Master's thesis
Nam [55], 2016	Effect of neurofeedback based robotic invention education of attention ability of ADHD children	South Korea	Peer-reviewed article
Nam and Mun [56], 2015	Development of neurofeedback based robotic invention education program for ADHD children	South Korea	Peer-reviewed article
Ryu [57], 2015	Effects of neurofeedback training on EEG, continuous performance task, and ADHD symptoms in ADHD in ADHD-prone college students	South Korea	Master's thesis
Ryu [58], 2021	A study on the clinical usefulness of EEG and QEEG measurements for the diagnostic criteria of ADHD	South Korea	Peer-reviewed article
Siddharth et al [59], 2019	A wearable multi-model biosensing system toward real-world applications	United States	Peer-reviewed article
Yun and Kwack [60], 2015	The treatment effect of neurofeedback training on executive function in attention-deficit hyperactivity disorder	South Korea	Peer-reviewed article
An et al [61], 2016	Cognitive behavioral therapy for college students with ADHD tendencies	South Korea	Peer-reviewed article

Study, year	Title	Country	Type
Hong et al [62], 2015	Development of working memory training program for ADHD children and effectiveness verification	South Korea	Peer-reviewed article
Chang and Park [63], 2020	Development and application of the working memory improvement program for children with ADHD in the first grade elementary school	South Korea	Peer-reviewed article
Lee [64], 2019	The effects of self-monitoring cognitive functions training program on the attention-concentration ability and the hyperactivity of the children with ADHD tendency	South Korea	Peer-reviewed article
Park et al [65], 2015	Effects of cognitive behavioral therapy on attention deficit hyperactivity disorder among school-aged children in Korea	South Korea	Peer-reviewed article
Kang et al [66], 2015	Development of Korean adult ADHD rating scale	South Korea	Peer-reviewed article
Kim [67], 2016	(The) clinical utility of K-CBCL 6-18 in diagnosing ADHD: focused on children with psychological disorder in Child Welfare Institution	South Korea	Master's thesis
Lee [68], 2015	Current status and future improvement of the Korea ADHD rating scale-IV (K-ARS-IV)	South Korea	Peer-reviewed article
Lee [69], 2017	A review on the diagnosis of ADHD for special education	South Korea	Peer-reviewed article
Lee [70], 2020	A review of diagnosis and evaluation procedure for the child and adolescent with attention deficit hyperactivity disorder	South Korea	Peer-reviewed article
Lee et al [71], 2015	Clinical utility of the Korean version of CBCL6-18 in the diagnosis of attention-deficit hyperactivity disorder	South Korea	Peer-reviewed article
Lee et al [72], 2016	The guideline of diagnosis and treatment of attention-deficit hyperactivity disorder: developed by ADHD Translational Research Center	South Korea	Peer-reviewed article
Lee et al [1], 2014	A study on agreement between parent's and teacher's ratings according to ADHD screening	South Korea	Peer-reviewed article
National Research Foundation of Korea [73], 2016	Success in quantifying the level of attention and concentration through meditation and exercise [electronic data]: expected to be used in diagnostic tests for ADHD, depression, and dementia in children	South Korea	News release
Park [74], 2015	Clinical application of advanced test of attention as a diagnostic tool in children with attention-deficit/hyperactivity disorder	South Korea	Doctoral dissertation
Cho [75], 2023	Development of rhythm-based music intervention protocols through timing control in children with ADHD	South Korea	Doctoral dissertation
Choi [76], 2019	Development of a music program for improvement of the mental concentration and human relationship using Carl Orff's pedagogics: centered about the ADHD students	South Korea	Master's thesis
Kim [77], 2016	The effects of literary therapy program based on SST by using picture cards on ADHD of adolescents for EBD	South Korea	Peer-reviewed article
Kim [78], 2016	Effects of mindfulness-based stress reduction (MBSR) program on attention, perceived stress, and anxiety on attention-deficit/hyperactivity disorder (ADHD) prone university students	South Korea	Master's thesis
Son [79], 2022	A study on development of diagnostic assessment tools of music therapy in children with attention deficit hyperactivity disorder	South Korea	Doctoral dissertation

^aR and D: research and development.

Digital Therapeutics

Digital therapeutics is a broad category that refers to high-quality software, that is, digital technologies, including AI, VR, augmented reality (AR), apps, and wearable devices, that provide evidence-based therapeutic interventions to patients to prevent, manage, or treat medical disorders or diseases [15]. Digital therapeutics is one of the promising methods of intervention, treatments, and diagnosis for ADHD and ASD in South Korea. Two dissertations [13,14] and 1 research report [15] described digital therapeutics. Several specific types of

digital therapeutics, such as AI, machine learning, and VR, will be further discussed in detail in other groups later.

The digital health care market is consistently growing in South Korea, especially since the COVID-19 pandemic has facilitated web-based treatments and health communication. The South Korean government has suggested diverse policies to promote the digital health care industry. In April of 2022, 10 digital therapies were approved by the Ministry of Food and Drug Safety of the Republic of Korea to conduct clinical trials [15]. However, none of them have been approved as a medical device [15]. Although the Korean digital therapeutics industry is in a beginning stage of development, the digital therapeutics market

is highly promising in South Korea given the policy support and attention from the South Korean government.

For instance, digital therapeutics is applied to treatments and education for students with ADHD [13,14]. Digital therapeutics is emphasized as a new treatment approach for children and adolescents with ADHD. A substantial improvement was found in groups using digital therapeutics compared with control groups [14]. Educational materials about digital therapeutics were also highlighted for elementary school teachers educating students with ADHD. The use of Korean medical devices for ADHD and ASD reflects the attention to educating teachers about the importance and functions of digital therapeutics, as well as its direct application to those with ADHD or ASD.

VR and AI

VR and AI are categorized into digital therapeutics. Both technologies are promising tools that recent studies have highlighted their potential [19,20]. VR and AI were mentioned in 11 resources. Two government research and development project reports [18,26], 6 peer-reviewed articles [10-15,18-23,26], 2 dissertations [16,20,24,25], and 1 interview [17] described medical devices for ADHD and ASD using VR and AI.

Although the fundamental treatment method for ADHD is medication, behavioral problems are treated by CBT [17]. However, traditional CBT has limited accessibility in clinical settings due to additional time to visit hospitals, health care personnel, and relevant resources [17]. In this circumstance, VR technology has a big advantage in solving these limitations by enabling real-time simulations and virtual training [17]. VR-based diagnosis of ADHD is also considered to have accurate and objective results given that the model is based on VR settings, while the traditional diagnosis relies on verbal interviews [25]. Furthermore, virtual social interactions allow repetitive practice for anger recognition, anger regulation, and social problem solving [22]. This VR-based training helps control their aggressive and impulsive behaviors [22]. Some recent studies also indicated that VR-based interventions for ADHD can prevent potential crime, especially for young people with ADHD [22,23]. An AI-automated diagnosis system for diagnostic classification and automated detection based on the biosignals of ADHD was introduced for the prediction, suppression, and prevention of adolescent recidivism [23].

To enhance the efficiency of treatment and diagnosis of ADHD and ASD, VR technologies are often integrated with AI to predict, analyze, and define different types of data from patients and users [19,20]. For example, a device using a gradient-boosted decision tree algorithm was evaluated to test the accuracy of its AI-based software when health care providers diagnose ASD in children aged between 18 and 72 months [19]. The study found that an increased number of children with ASD were able to be diagnosed in a primary care setting, potentially promoting early intervention and treatment [19]. In addition, the combination of a deep learning algorithm and facial images is a novel approach in the diagnosis of ASD [16]. Given that ASD is marked by impaired neurological development, the human face provides insights into brain structure and function [16]. Consequently, facial features can serve as an important

biomarker for diagnosing ASD [16]. This idea is also applied to a wearable digital intervention that promoted emotion recognition and facial engagement [24]. Researchers found that children with ASD who wore superpower glasses showed significant improvements on socialization by providing social cues and detecting facial expressions [24].

Following the trends of the combined technologies with VR and AI, research and government research and development projects have studied possible medical device models for people with ADHD or ASD [18,21,25,26]. The Korea Electronics Technology Institute conducted a project to develop VR and AR platform technology based on biosignals for the mental health of kids and silver generation [18]. In total, 7 types of VR, 6 types of AR mental health content, a mental-care cloud platform, wireless transfer technology for 4K video streaming, and systems to measure and analyze biosignals were developed throughout the project [18]. These VR and AR technologies can be broadly applied to mental health VR and AR platform services at mental-health clinics in South Korea [18]. This application also positively impacts the Korean web content industry beyond the Korean medical device industry.

Medical VR and AI technologies were interweaved with IT and life care content markets [21]. Regarding ADHD treatments, contents and systems using immersive and vivid exposure in virtual settings have been actively tried [21]. The system virtually provides the actual circumstances where people with ADHD can be trained through sensory, cognitive, and linguistic simulations [21]. With a similar purpose, the Office of Research Affairs at Yonsei University conducted a project to develop mobile VR neuropsychological batteries and an AI-based database of early diagnosis and promotion systems using digital phenotypic modeling [26]. The developed device was based on a multilayer platform integrating emotions, social ability, and neurological information [26]. Both devices target vulnerable populations with limited access to traditional treatments for ADHD.

The overall trends of Korean medical devices for ADHD and ASD concentrate on improving the current conditions of the medical device application and its use [25,26]. While most of the traditional approaches require personnel, physical resources, and travel time, VR and AI-based medical devices minimize the requirements.

Machine Learning and Robot

Machine learning and robot-based medical devices for ADHD and ASD are also included in the category of digital therapeutics. They were found in 6 resources. Two dissertations [29,32] and 4 journal articles [27,28,30,31] addressed its trends.

The importance of early diagnosis is highlighted in many studies about ADHD and ASD [30,41]. To facilitate early diagnosis and ADHD screening, machine learning and robot-based technologies are used as a promising tool. A machine learning predictive model is one of the solutions to increase the accuracy of ADHD prediction [30]. As a longitudinal predictive model, several types of machine learning analysis were applied to predict the future and classify findings, such as supervised learning, random forest, gradient boosting, and neural network

models [30]. This model identified that children who showed specific risk indicators during infancy and early childhood are likely to be diagnosed as being at risk for ADHD when entering elementary schools [30]. Similarly, machine learning and a wearable biosensor help to predict imminent aggressive behavior in inpatient young people with ASD [27]. In addition, several robot-based ADHD screening devices have been tested, such as a contactless sensing system, a deep learning–based classifier, storyboard content for children, and an automated childhood ADHD classifier [29,31,32]. The contactless sensing system, for instance, quantitatively measures the movements of children with ADHD [29]. These devices automatically detect and analyze behavioral reactions, and identify results based on collected data [29,31,32].

Furthermore, machine learning and robot-based devices are also applied to interventions. Remote robot–based interventions are effective in enhancing the level of concentration and encouraging positive learning attitudes among children with severe ADHD symptoms [28]. They recognize a robot as a peer, a good behavioral model, and a learning helper [28].

The overall observations and findings imply that robot-based models are relatively more attractive in younger ages. Machine learning systems also have a higher effectiveness and accuracy of screening.

Gaming and Visual Contents

Many types of gaming can be a part of digital therapeutics, depending on their medium. Given that a gaming approach has less rejection than others [35], its use is actively discussed in the recent medical device development for ADHD and ASD. Gaming was mentioned in 4 resources. One dissertation [35], 1 interview [33], 1 journal article [34], and 1 research report [36] found their trends in Korean medical device development for ADHD and ASD.

Gaming is applied to various fields today, not just as an entertainment tool [36]. The research found that gaming helps people relieve negative emotions and improves symptoms [36]. The development of gaming items was motivated by one of the limitations that psychiatrists' diagnoses rely on subjective individual decisions [33]. A gaming device, AttnKare, made by Hippo T&C, is complex equipment that uses a VR test and measures eye movements and patience [33]. The AI in the device analyzes the collected information and makes individual diagnoses [33].

The cognitive rehabilitation field recently uses computer technology, focusing on basic cognitive function, memory, problem-solving ability, and perception of space and time [35]. This new digital model is personalized to different individuals [35]. Serious games in this field, defined as education-purposed games with entertaining functions [36], are a promising method that can result in easier and faster positive outcomes both in education and treatment [35]. For example, when comparing responses from 2 groups using a communication-functional board game or a traditional board game, those who used a communication-functional board game showed a better score in self-control, self-esteem, family function, and peer relationship [34].

Eye-Feedback and Movement Intervention

Eye-feedback and movement technologies are often found in ADHD screening devices. Because eye movements are linked to brain areas with neuropsychological functions, such as response inhibition, selective attention, and working memory, their impairments lead to the primary traits of ADHD [41]. Eye-feedback and movement intervention also have a complex relationship with the categories above, such as VR and AI, gaming, and machine learning. The information about this category was found in 5 resources. Two dissertations [37,38] and 3 journal articles [39-41] described medical devices using eye-feedback and movement intervention.

Using a screening model for ADHD with eye-tracking features and machine learning, 33 eye-tracking features were identified to distinguish children with ADHD from developing children [41]. Eye-tracking characteristics have the potential to serve as a reliable marker for compromised neurobiological function in individuals undergoing ADHD screening [41]. The focus reaction time tests were identified as a valid tool for diagnosing children with ADHD [40]. Given that visual materials tend to be eye-catching and vision accounts for 80% of human recognition [37], visual content can also play an important role in developing interventions for ADHD using eye movements. For example, eye-feedback training improves sluggish cognitive tempo, one of the symptoms of ADHD that shows a lack of energy, slowness in behavior or thinking, and drowsiness [38]. A motion-training system with real-time visual feedback also facilitated motion control in children with ASD [39].

Electroencephalography and Neurofeedback

This category discusses electroencephalography and neurofeedback. Both concepts are relevant to digital technologies, including gaming and machine learning [49]. Electroencephalography, a recording of the brain's electrical activity, measures brainwaves. Neurofeedback is used to modify brainwaves by providing stimulus in neurofeedback training, which is considered a promising physiological approach for the diagnosis and interventions of neurological disorders, such as ASD and ADHD [42,45,47,48,52-56,60]. This topic was mentioned in 19 resources. Two research reports [44,45], 10 journal articles [42,43,48,50,51,55,56,58-60], and 7 dissertations [46,47,49,52-54,57] discussed electroencephalography-based medical devices.

Wearable wireless systems and sensing systems are new potential solutions for diagnosing ASD and ADHD by collecting physiological indicators [42,43,59]. Electroencephalography can detect the abnormalities of the neural system related to ASD and ADHD [42,43,59]. The research found that ADHD can be diagnosed by sounds derived from brainwaves, using (1) ADHD diagnosing algorithms developed by electroencephalography brainwaves with several mathematical methods, eyes-open, and resting-state brainwaves, and (2) a sonification algorithm to convert brainwaves to musical sounds [46].

Convolutional neural network (CNN) is another emerging idea to automatically extract electroencephalography features for medical diagnosis [48,49]. CNN is a neural network modeled after the functioning of the visual cortex for processing data

that contains spatial information [49]. Recent research explores deep learning-based devices using CNN to effectively classify electroencephalography signals [48]. A deep learning-based approach using functional magnetic resonance imaging (fMRI) was also another recent discussion [49]. While previous trials covered the entire brain area to identify ADHD, the recent study suggests examining specific brain portions related to the classification of ADHD using the deep learning system by demonstrating a higher level of accuracy [49].

Neurofeedback is another key topic in interventions for ADHD and ASD. Neurofeedback training is a form of self-regulation therapy for brainwaves, using the concept of operant conditioning [52]. During brainwave measurement, patients receive visual or auditory feedback on cortical activity to normalize brain function by inhibiting or reinforcing specific frequency ranges of brainwaves [52]. Neurofeedback enables them to receive real-time feedback on their brainwave states and engage in training to regulate brainwaves as desired [52]. Many studies claim that neurofeedback training positively impacts children with ADHD [47,52-54,60].

Neurofeedback training positively impacted children with ASD by improving their attention and abnormal brainwaves [52]. Students who received neurofeedback training showed increased scores in memorizing numbers and matching colors, numbers, and words [52]. Furthermore, neurofeedback training can be applied to those with ADHD [47,60]. Recent research has reported that 30% of people with ADHD with executive function deficits and inhibitory deficits cannot be treated both by medication and CBT [50,60]. Neurofeedback training is suggested as one of the promising alternative solutions of medication to improve executive functions, inhibition, and working memory [47,50,60]. Moreover, delayed language development and communication ability among children with ADHD can be improved by neurofeedback training [53,54]. In fact, parents having children with ADHD have reported positive outcomes after using neurofeedback training [50]. These trends imply that neurofeedback models can be more effective when they are integrated with different digital items, such as VR, gaming, and AI [54].

As an example of complex medical devices for interventions for ADHD, research suggested a robotic intervention education using neurofeedback [55]. In this program, students with ADHD were encouraged to craft a robot and control its movements using brainwave signals [55]. This program aimed to enhance the level of concentration as well as treatment of ADHD with a children-level storyline [56]. The satisfaction was evaluated positively, while a general operation process had a few comments on further development [56]. In addition, CNN is also used to diagnose ADHD in young children <16 years who are too immature to perform self-diagnosis or use medical equipment [51]. Gaming content is used to increase the objectivity and accuracy of ADHD diagnosis and collected electroencephalography data are classified based on the CNN model [51].

With a similar context of education using neurofeedback, the Korea Institute of Curriculum and Evaluation conducted a 2-year project to design and implement brain-based training for children

with ADHD [44,45]. Neuroeducation was applied to the project to explore neuroeducational research tools, including electroencephalography, positron emission tomography, and fMRI [44]. fMRI was especially highlighted to indirectly measure brain activity status by quantifying cerebral blood volume, cerebral blood flow, and blood oxygen saturation [44]. The training program, named Korea Institute of Curriculum and Evaluation Working memory Enhancement Program, involves altering brain function through interaction with the environment, which leads to improved cognitive functions [44,45]. The Korea Institute of Curriculum and Evaluation Working memory Enhancement Program showed positive outcomes among children with ADHD in a clinical trial by enhancing cognitive abilities and demotivating behavioral problems [45].

Traditional Therapy

Traditional therapy mostly does not use medical devices. The 3 categories under traditional therapy examined the trends in traditional treatments and diagnosis for ADHD and ASD.

CBT and Working Memory

CBT and working memory fundamentally aim to improve cognitive ability as well as attention deficits and impulsive behaviors [61,63,64]. CBT focuses on a behavioral intervention [61], while working memory refers to a cognitive function that involves temporarily holding or manipulating information for a short period [62,63]. They were found in 5 journal articles [61-65].

CBT can be a more effective intervention for adults with ADHD than children with ADHD because adults relatively have a higher cognitive ability and reflective thinking [61]. CBT demotivated people to think about ADHD and think negatively, while knowledge of ADHD was increased [61]. CBT can also be developed as a self-monitoring cognitive training program to help children with ADHD regulate and monitor their thoughts and behaviors during task execution [64]. This approach focuses on individual behavioral problems as well as individual thinking processes, which can be applied to diverse treatments and research on ADHD [65].

A working memory training program is another method to reduce impulsive behaviors and hyperactivities [63]. Given that delivery forms of information and cognitive ability are correlated with one another, previous research findings indicated that delivery forms of information influence outcomes of working memory training programs [62]. This statement implies that a better performance is presented when performing a preferred delivery form of information [62], meaning that understanding a target population's preference for communication matters in working memory training.

Diagnosis and Rating Scales

While many studies discuss the recent trends in medical devices for ADHD and ASD, mostly focusing on digital technologies, traditional methodologies of diagnosis and rating scales are still discussed to update the standards and guidelines. The diagnosis and rating scales were examined in 10 resources. One news release [73], 7 journal articles [1,66,68-72], and 2 dissertations

[67,74] addressed the current trends in diagnosis and rating scales of ADHD and ASD.

ADHD diagnosis should be systematically approached through diagnostic algorithms to make safe and accurate decisions [70]. Multiple factors, including age, gender, and individual perceptions of ADHD, need to be considered, and the diagnostic decision needs to be based on the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)* [70]. The *International Classification of Diseases, 10th Revision (ICD-10)* is also discussed [69,70]. Although a few differences are presented between *DSM-5* and *ICD-10*, both models focus on attention deficits, hyperactivities, academic and social difficulties, and impulsive behaviors [70]. ADHD diagnosis usually refers to the *DSM-5*, while public health statistics and materials are based on *ICD-10* [69]. *ICD-10* has more strict standards of ADHD diagnosis than *DSM-5* by recognizing all 3 categories: attention deficits, hyperactivity, and impulsions [69].

In addition to *DSM-5* and *ICD-10*, the Children Behavior Check List (CBCL) is a self-report assessment scale developed to evaluate various emotional and behavioral problems of children and adolescents through reports from parents or close adults in their environment [67]. In South Korea, the US version of the CBCL 4-18 in 1991 was standardized and first introduced as the Korean version of CBCL, and the Korean version of the CBCL 6-18 is the recent version for parents [67,71].

ADHD screening and evaluation were performed in in-person interviews at hospitals, mental-health centers, and school counseling offices [72]. Two interview tools are used: Diagnostic Interview Schedule for Children-IV and Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL-K) [72]. While Diagnostic Interview Schedule for Children-IV is a structured interview tool that can be administered by general people, K-SADS-PL-K is a semistructured tool that should be administered by trained evaluators [72]. With K-SADS-PL-K, a recent study tried the advanced test of attention, consisting of visual tests and auditory tests that present target and nontarget stimuli at regular intervals, and participants were instructed to respond only to the target stimuli [74]. However, the accuracy of distinguishing a group with ADHD from another group without ADHD was not high, which suggests limitations in using the advanced test of attention as a diagnostic tool for confirmation [74].

Given that ADHD symptoms tend to be presented at an early age, parents' and teachers' knowledge and perception of ADHD greatly impact their children's diagnosis and intervention [1]. Interestingly, the ratings of parents and teachers about symptoms of children with ADHD had no significant correlations, and parents' ratings and DISC were not matched [1]. By contrast, the rating of teachers was consistent and showed a high correlation with DISC [1]. These findings imply that DISC and the rating of teachers are more reliable and consistent than the rating of parents [1].

In case childhood ADHD may persist into adulthood, the Korean Adult ADHD Rating Scale was developed for monitoring and screening treatment of adults with ADHD [66]. Inattention was

recorded as the most general symptom of ADHD in adulthood [66]. The Korean Adult ADHD Rating Scale is expected to effectively rate difficulty in emotional control and disorganization, such as inattention, hyperactivity, and impulsivity [66]. This rating scale was also suggested to extend its range of use to adolescents, embrace gender differences, and identify screening and rating scales, respectively [68]. In this light, traditional rating scales are consistently developed. For example, one of the recent rating scales is a tactile stimulation distribution device to quantify exercise and MBSR [67]. The details of MBSR and other traditional therapies are discussed in the last category.

Musical, Literary Therapy, and MBSR

MBSR, musical, and literary therapy described here were developed to increase the effectiveness of screening and intervention for children with ADHD. They were addressed in 5 resources. One journal article [77] and 4 dissertations [75,76,78,79] discussed how they were recently shaped.

The tactile stimulation distribution device was motivated by mindfulness, MBSR, and CBT and scientifically demonstrated a level of concentration of subjects [78]. The quantified data of stimulation were compared with the cognitive outcomes of subjects [78]. The correlative data were referred to as concentration, and the opposed data were considered a distraction [78]. This logic was also supported by left- and right-brain activities [78]. In fact, an MBSR-based program showed a significant improvement in reducing inattention, stress, and anxiety in college students with ADHD [78].

Another approach to intervention for children with ADHD is literary therapy based on social skills training [78]. The program was designed to train them to improve a social relationship between peers and adults and engage in group activities at home and school [78]. Using photo cards and photo books to inspire their imagination and creativity, the general symptoms of ADHD decreased with a significant improvement in emotional and mental stability [78].

Musical therapy is also used for screening and intervention of ADHD. Screening ADHD using musical therapy aims to strategize a plan of treatment by understanding individual conditions and the goals of treatments [79]. This screening is essential to comprehend how musical reactions can be used to improve symptoms when music attracts clients' changes [79]. While musical therapy screening is designed for a broad understanding of individuals, interventions using musical therapy have a specific purpose to target specific symptoms. A rhythm-based musical intervention was developed to enhance timing control in children with ADHD [75]. The protocol contributed to controlling motor timing and perceived timing using a metronome, guiding a proper speed of response to suggested stimuli in the environment [75]. Carl Orff's pedagogics, focusing on improvisatory performance with observation, imitation, exploration, and music literacy, is another type of intervention using musical therapy [76]. This program required small group activities, which encouraged social interaction with different individuals [76]. Furthermore, improvisatory work improved inattention, and imitating musical expression demotivated hyperactivities [76]. These findings

indicate that musical therapy is a highly effective method both for screening and intervention in children with ADHD.

Discussion

Principal Findings

This study conducted a review of the literature to reduce gaps in the research related to medical device use and development in South Korea for the diagnosis and management of ADHD and ASD. The trends in Korean medical device development for ADHD and ASD are categorized into 2 major groups with 8 subgroups in total. Digital therapeutics using AI, machine learning, and electroencephalography technologies account for the biggest portions of development in South Korea, rather than traditional therapies. Given that both ADHD and ASD are neurological disorders, emerging medical device technologies especially focus on electroencephalography and neurofeedback. Different types of digital models are combined or applied to understand brain activities and brainwaves.

In this vein, future development of medical devices for ADHD and ASD is predicted to heavily rely on digital technologies. As digital medical devices are emerging trends in South Korea, they can also be integrated with traditional therapies. For example, the rhythm-based musical intervention can be applied to a gaming device for ADHD, which can also detect particular brainwaves and provide real-time neurofeedback. Recent research has reported that traditional therapies, including musical features and MBSR, have succeeded in screening and intervention for ADHD and ASD. Understanding their strengths and integration with digital medical devices will double the effectiveness of screening and intervention outcomes.

However, this growing transformation is faster than people's perception of their development. To follow the trends and learn digital literacy for new digital medical devices, training programs about up-to-date digital devices for ADHD and ASD are recommended, especially for parents and teachers to relieve tension in school. The active application of digital devices in school settings is also expected to enable early diagnosis and treatment for students with ADHD or ASD. Because parents and teachers are primary and secondly important people for children with ADHD or ASD [1], education for them is essential to implementing new medical devices into routine care in the real world.

In addition to the application of digital devices, traditional therapies are used for children with ADHD or ASD in school settings. While digital therapeutics is a promising tool today, traditional therapies have still demonstrated their efficacy in screening and interventions. The research presented real-world case studies of the applications that showed positive outcomes and high reliability [75,76,78,79]. Extending this idea, future

research could discuss the potential efficacy of integrating digital therapeutics and traditional therapies for the diagnosis and interventions of ADHD and ASD. Furthermore, potential ethical dilemmas associated with the use of medical devices for these conditions are another important topic to study. Understanding the negative effects and limitations of different types of devices in clinical settings will also guide the direction of future development of medical devices for ADHD and ASD.

Limitations

The first limitation of this study is that many resources had small population sizes to conclude their findings, which makes it hard to generalize the outcomes. To define the accurate trends in Korean medical device development for ADHD and ASD, additional studies conducted with larger populations should be examined. Second, a lower number of records specifically discussed medical devices for ASD, while most of the selected resources focused on ADHD. The results had to focus more on devices for ADHD than ASD. Further research on medical devices for ASD should be studied to understand the need for medical devices for ASD. These studies expect to promote early diagnosis and interventions, which lead to reduced prevalence rates for both ADHD and ASD. Third, given several emerging medical device areas, most of the selected resources were dissertations. They helped understand the recent trends in medical devices for ADHD and ASD; however, peer-reviewed journal articles are required in the future to examine in-depth trends in specific medical devices for ADHD and ASD. Fourth, the limited number of databases were used, especially only 1 Korean database was explored. Fifth, search terms are difficult to truly replicate the same search in the different languages. Further research is recommended to conduct Korean-focused medical devices by directly communicating with Korean medical device companies and relevant experts to reduce the language gaps. The results from this paper will help guide future works.

Conclusions

In conclusion, this study aims to provide significant insight to understand the recent trends in Korean medical device development, focusing on medical devices for ADHD and ASD. Emerging digital medical devices and those integrated with traditional therapies are some of the important solutions to reducing the prevalence rates of ADHD and ASD in South Korea by promoting early diagnosis and intervention. Furthermore, their application will relieve pressures on teachers and school-based special education programming by providing direct supporting resources to students with ADHD or ASD. Educating parents and teachers about the trends in relevant medical devices also matters in further responses to their children. Further research is recommended to focus on medical devices for ASD given that the number of current studies discuss those for ADHD rather than ASD.

Acknowledgments

The authors would like to acknowledge Dr. Hye-Joo Kwon, Associate Professor, University of Utah Asia Campus, who reviewed the manuscript and provided feedback for the improvement. This study was supported by the University of Utah Asia Campus Undergraduate Research Opportunity Grant.

Authors' Contributions

YC conceived of the research topic, conducted the search and record review, and wrote background, methods, results, and discussion. SLT contributed to the background, methods, and discussion and conducted secondary analysis and auditing of search methods and resource inclusion.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies and eligibility criteria.

[[DOCX File, 27 KB - biomedeng_v9i1e60399_app1.docx](#)]

References

1. Lee YE, Bae SM, Sunjoo H, Lee LM. A study on agreement between parent's and teacher's ratings according to ADHD screening. *Res Emot Behav Disord* 2014;30(4):387-400 [[FREE Full text](#)]
2. Kim KD. "All school assistants, don't get involved today". Teachers hold a large rally at 2 p.m. to 'improve teaching authority'. *Pen & Mike*. URL: <https://www.pennmike.com/news/articleView.html?idxno=67022> [accessed 2024-04-29]
3. "Even though I was born in '00." Teacher in Seoi-cho committed suicide, new faculty member lost his life due to 'true parents'. *Whisper the News*. URL: <https://www.newssocdak.com/news/articleView.html?idxno=10534> [accessed 2024-04-29]
4. Choi J. "I can't even get sick" ... school special education with only one teacher in mind. *Hyundai*. URL: <https://www.yna.co.kr/view/AKR20230731104000061> [accessed 2023-07-31]
5. Attention-deficit/hyperactivity disorder. National Institute of Mental Health (NIMH). URL: <https://tinyurl.com/2vxjwayd> [accessed 2024-04-29]
6. Autism spectrum disorder. National Institute of Mental Health (NIMH). URL: <https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd> [accessed 2024-04-29]
7. ADHD treatment status by region. HIRA Bigdata Open portal. 2023 Jul 31. URL: <https://tinyurl.com/nhzdsvhs> [accessed 2024-04-29]
8. Number of patients with activity and attention disorder (F900). National Health Insurance Service. URL: <https://www.data.go.kr/data/15123392/fileData.do> [accessed 2024-04-29]
9. Announcement of the results of the 2021 survey on the status of persons with developmental disabilities. Ministry of Health and Welfare of South Korea Division of Services for Persons with Disabilities. URL: <https://tinyurl.com/mwtcp38c> [accessed 2022-09-06]
10. Main indicators of the 2021 survey on the status of people with developmental disabilities. Ministry of Health and Welfare of South Korea Division of Services for Persons with Disabilities. 2022 Mar 17. URL: <https://tinyurl.com/mwtcp38c> [accessed 2024-04-29]
11. Lee S, Lee SJ, Yoon S. Demand for the establishment of a support system for early diagnosis of autism spectrum disorder and linkage to education: focusing on parents' experiences and perceptions: a study of support system for facilitating early diagnosis and intervention for young children with autism spectrum disorders: based on the experiences and the perceptions of parents. *J Korean Assoc Persons Autism* 2013;13(1):167-199 [[FREE Full text](#)]
12. Skutle A, Bu ET, Jellestad FK, van Emmerik-van Oortmerssen K, Dom G, Verspreet S, et al. Early developmental, temperamental and educational problems in 'substance use disorder' patients with and without ADHD. Does ADHD make a difference? *Addict Behav Rep* 2015 Dec;2:13-18 [[FREE Full text](#)] [doi: [10.1016/j.abrep.2015.03.001](https://doi.org/10.1016/j.abrep.2015.03.001)] [Medline: [29531989](https://pubmed.ncbi.nlm.nih.gov/29531989/)]
13. Choi ES. A study on the principles of digital therapeutics for ADHD and digital drama response. *Korea Soc Inform Technol* 2019;1-7 [[FREE Full text](#)]
14. Lee D. (The) efficacy of digital therapeutics for the treatment of attention deficit hyperactivity: a systematic review and meta-analysis. *Chung-Ang University*. 2019. URL: <https://library.cau.ac.kr/search/i-discovery/9539955?type=biblios-list-view> [accessed 2024-04-29]
15. Son J, Yang S, Jung I. Current status and outlook for digital therapeutics. *Korea Insurance Research Institute*. 2023. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=MONO12023000037043> [accessed 2024-04-29]
16. Rashid MM, Alam MS. Power of alignment: exploring the effect of face alignment on ASD diagnosis using facial images. *IJUM Eng J* 2024 Jan 01;25(1):317-327. [doi: [10.31436/ijumej.v25i1.2838](https://doi.org/10.31436/ijumej.v25i1.2838)]
17. Kim JJ. To strive for the universalization of virtual reality therapy programs. *Smart Device Trends Mag* 2019;37(1):42-45 [[FREE Full text](#)]
18. Korea Electronics Technology Institute. Safe VR/AR platform technology based on vital signs for the mental health of the kids/silver generation. *Korea Creative Content Agency*. 2020. URL: <https://tinyurl.com/3jx3burk> [accessed 2024-04-29]
19. Megerian JT, Dey S, Melmed RD, Coury DL, Lerner M, Nicholls CJ, et al. Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. *NPJ Digit Med* 2022 May 05;5(1):57 [[FREE Full text](#)] [doi: [10.1038/s41746-022-00598-6](https://doi.org/10.1038/s41746-022-00598-6)] [Medline: [35513550](https://pubmed.ncbi.nlm.nih.gov/35513550/)]

20. Son H. Towards standardizing attention-deficit/hyperactivity disorder diagnosis- a virtual reality, artificial intelligence application. Sungkyunkwan University. 2019. URL: <https://tinyurl.com/279rs794> [accessed 2024-04-29]
21. Park JW, Jin CS, Jeong WJ, Oh SH. Design and implementation of VR-based life care contents for attention deficit hyperactivity disorder (ADHD). *J King Comput* 2019;15(5):84-91 [FREE Full text] [doi: [10.23019/kingpc.15.5.201910.008](https://doi.org/10.23019/kingpc.15.5.201910.008)]
22. Ryu C. Implications of VR-based psychotherapeutic effects for ADHD and CD among adolescents. *Korean Assoc Addict Crime* 2022;12(1):25-55. [doi: [10.26606/kaac.2022.12.1.2](https://doi.org/10.26606/kaac.2022.12.1.2)]
23. Ryu C, Hwang Y. Artificial intelligence analysis of bio-signals for automated detection/automated diagnosis of ADHD and CD. *Asian Forum Correct* 2021;15(3):87-118. [doi: [10.46626/affc.2021.15.3.4](https://doi.org/10.46626/affc.2021.15.3.4)]
24. Voss C, Schwartz J, Daniels J, Kline A, Haber N, Washington P, et al. Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial. *JAMA Pediatr* 2019 May 01;173(5):446-454 [FREE Full text] [doi: [10.1001/jamapediatrics.2019.0285](https://doi.org/10.1001/jamapediatrics.2019.0285)] [Medline: [30907929](https://pubmed.ncbi.nlm.nih.gov/30907929/)]
25. Yoo S. Virtual reality based digital therapeutics system for diagnosing attention-deficit hyperactivity disorder. Sungkyunkwan University. 2020. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT12021000021390> [accessed 2024-04-29]
26. Yonsei University Office of Research Affairs. Development of mobile VR neurocognitive battery and establishment of database, implementation of AI-based early diagnosis/prevention system for cognitive control vulnerable groups utilizing digital representation modeling. Korea Creative Content Agency. 2019. URL: <https://tinyurl.com/37puskfx> [accessed 2024-04-29]
27. Imbiriba T, Demirkaya A, Singh A, Erdogmus D, Goodwin MS. Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism. *JAMA Netw Open* 2023 Dec 01;6(12):e2348898 [FREE Full text] [doi: [10.1001/jamanetworkopen.2023.48898](https://doi.org/10.1001/jamanetworkopen.2023.48898)] [Medline: [38127348](https://pubmed.ncbi.nlm.nih.gov/38127348/)]
28. Kim S, Kim J, Lee H, Kim S, Shin Y, Kim C. Exploring the applicability of tele-presence robot intervention for at-risk children with ADHD. *J Spec Educ Rehabil Sci* 2016 Sep 30;55(3):415-434. [doi: [10.15870/jsers.2016.09.55.3.415](https://doi.org/10.15870/jsers.2016.09.55.3.415)]
29. Lee D. Development of a contactless sensing system and a classifier using deep learning for robot-based ADHD screening. Gwangju Institute of Science and Technology. URL: <https://tinyurl.com/4wy2h4jb> [accessed 2024-04-29]
30. Lee D, Jang H, Kim H, Bae J, Park J. Development of a machine-learning predictive model for first-grade children at risk for ADHD. *Korea J Childcare Educ* 2019;17(5):83-103 [FREE Full text]
31. Shin YH, Lee HS, Kim SB, Noh HJ, Kang SK, Choi MT, et al. Exploring the performance difference on the active based task with a robot for ADHD screening. *J Emot Behav* 2018 Mar 31;34(1):113-133. [doi: [10.33770/jebd.34.1.7](https://doi.org/10.33770/jebd.34.1.7)]
32. Yeom J. Supervised classification of childhood ADHD using robot-assisted tests. Sungkyunkwan University. 2018. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT1201933474> [accessed 2024-04-29]
33. Jung T. ADHD can be treated like playing a game in daily life. *Nation Econ* 2019;376(3):44-45 [FREE Full text]
34. Lee Y, Lim J. Effects of communicative functional board games on self-control, self-esteem, family functioning, and peer relationships in children with ADHD: a study on the effect of communication functional board game on self-control, self-esteem, family function and peer relationship of ADHD children Dong-Eui Lee. *J Public Policy* 2018;35(1):99-134 [FREE Full text]
35. Park W. To improve the concentration of ADHD children study on functional games. Daegu University. 2019. URL: <https://tinyurl.com/5n98xv5b> [accessed 2024-04-29]
36. Sungkyunkwan University Cooperation Center. A study on the therapeutic applications of digital games. Korea Creative Content Agency. 2020. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=MONO12021000024297> [accessed 2024-04-29]
37. Kim E. Development of movement intervention visualization contents to improve behavior of ASD and ADHD. Inha University. 2018. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT1201851176> [accessed 2024-04-29]
38. Kim G. The characteristic of attentional networks in sluggish cognitive tempo: the effect of eye-feedback training on orienting attention in individuals with SCT. Chung-Ang University. 2019. URL: <https://tinyurl.com/yvxbjkpx> [accessed 2024-04-29]
39. Sandbank M, Cascio C. Using a motion-tracking device to facilitate motion control in children with ASD for neuroimaging. *Dev Neurorehabil* 2019 Aug 06;22(6):365-375. [doi: [10.1080/17518423.2018.1502831](https://doi.org/10.1080/17518423.2018.1502831)] [Medline: [30081715](https://pubmed.ncbi.nlm.nih.gov/30081715/)]
40. Yoo J, Kim G. A preliminary study on the development of the focus reaction time tests. *J Dev Disabil* 2015;43(3):59-74 [FREE Full text]
41. Yoo JH, Kang C, Lim JS, Wang B, Choi C, Hwang H, et al. Development of an innovative approach using portable eye tracking to assist ADHD screening: a machine learning study. *Front Psychiatry* 2024 Feb 15;15:1337595 [FREE Full text] [doi: [10.3389/fpsyt.2024.1337595](https://doi.org/10.3389/fpsyt.2024.1337595)] [Medline: [38426003](https://pubmed.ncbi.nlm.nih.gov/38426003/)]
42. Alhassan S, Soudani A, Almusallam M. Energy-efficient EEG-based scheme for autism spectrum disorder detection using wearable sensors. *Sensors (Basel)* 2023 Feb 16;23(4):2228 [FREE Full text] [doi: [10.3390/s23042228](https://doi.org/10.3390/s23042228)] [Medline: [36850829](https://pubmed.ncbi.nlm.nih.gov/36850829/)]
43. Bhattacharyya N, Singh S, Banerjee A, Ghosh R, Sinha O, Das N, et al. Integration of electroencephalogram (EEG) and motion tracking sensors for objective measure of attention-deficit hyperactivity disorder (MAHD) in pre-schoolers. *Rev Sci Instrum* 2022 May 01;93(5):054101. [doi: [10.1063/5.0088044](https://doi.org/10.1063/5.0088044)] [Medline: [35649790](https://pubmed.ncbi.nlm.nih.gov/35649790/)]

44. Hong S, Lee M, Jin G. Development of brain imaging diagnosis and brain-based training programs for ADHD students. Korea Institute of Curriculum and Evaluation. 2013. URL: <https://tinyurl.com/msemnmh9> [accessed 2024-04-29]
45. Hong S, Lee M, Jin G. Validation of the effectiveness of brain-based training programs for ADHD students. Korea Institute of Curriculum and Evaluation. 2014. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=MONO1201543162> [accessed 2024-04-29]
46. Kang GM. Brain music as a potential tool for diagnosing attention-deficit/hyperactivity disorder (ADHD). Advanced Institute of Science & Technology. 2013. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT1201669483> [accessed 2024-04-29]
47. Kim MA. The effects of neurofeedback training and executive function improvement programs on attention and brain function quotient of elementary school children. Daegu National University of Education. 2017. URL: <https://tinyurl.com/52y76wjc> [accessed 2024-04-29]
48. Kim MG. Machine learning-based EEG classification for assisting the diagnosis of ADHD in children. *J Korea Multimed* 2021;24(10):1336-1345 [FREE Full text]
49. Kim T. Deep learning approach on the improvement of diagnosing ADHD with fMRI. Hanyang University. 2019. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT12022000042521> [accessed 2024-04-29]
50. Kim JI, Yoon S, Oh HK, Lee S. Clinical significance for neurofeedback training of children with attention-deficit/hyperactivity disorder. *J Korean Neuropsychiatr Assoc* 2015;54(1):62. [doi: [10.4306/jknpa.2015.54.1.62](https://doi.org/10.4306/jknpa.2015.54.1.62)]
51. Kim DH, Park SM, Kim DH. The classification scheme of ADHD for children based on the CNN model. *Korea Electron Telecommun Soc* 2022;17(5):809-814. [doi: [10.13067/JKIECS.2022.17.5.809](https://doi.org/10.13067/JKIECS.2022.17.5.809)]
52. Lee Y. The effects of the neurofeedback training on the attention in adolescents with autism spectrum disorders. Hanyang University. 2013. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT1201356458> [accessed 2024-04-29]
53. Lee HR. Effects of neurofeedback brain wave training on the attention concentration and language development of children delayed in language development. Daegu University. 2019. URL: <https://tinyurl.com/329uth8m> [accessed 2024-04-29]
54. Lee MR. The effect of EEG training through neuro feedback on attention and pragmatic language ability in children with ADHD prone language delay. Daegu University. 2019. URL: <https://tinyurl.com/4kdn8uh7> [accessed 2024-04-29]
55. Nam H. Effect of neurofeedback based robotic invention education on attention ability of ADHD children. *Asia Pac J Multimed Ser Art Human Sociol* 2016 Jun 30;6(6):273-283. [doi: [10.14257/ajmahs.2016.06.33](https://doi.org/10.14257/ajmahs.2016.06.33)]
56. Nam H, Mun Y. Development of neurofeedback based robotic invention education program for ADHD children. *Asia Pac J Multimed Ser Art Human Sociol* 2015 Dec 31;5(6):429-438. [doi: [10.14257/ajmahs.2015.12.18](https://doi.org/10.14257/ajmahs.2015.12.18)]
57. Ryu M. Effects of neurofeedback training on EEG, continuous performance task, and ADHD symptoms in ADHD-prone college students. Jeonbuk National University. 2015. URL: <https://tinyurl.com/yc4z4s7w> [accessed 2024-04-29]
58. Ryu C. A study on the clinical usefulness of EEG and QEEG measurements for the diagnostic criteria of ADHD. *Korean Assoc Addict Crime Rev* 2021;11(3):23-45. [doi: [10.26606/kaac.2021.11.3.2](https://doi.org/10.26606/kaac.2021.11.3.2)]
59. Siddharth, Patel AN, Jung T, Sejnowski TJ. A wearable multi-modal bio-sensing system towards real-world applications. *IEEE Trans Biomed Eng* 2019 Apr;66(4):1137-1147. [doi: [10.1109/tbme.2018.2868759](https://doi.org/10.1109/tbme.2018.2868759)]
60. Yun SM, Kwack YS. The treatment effect of neuro feedback training on executive function in attention-deficit hyperactivity disorder. *J Korean Acad Child Adolesc Psychiatry* 2015 Mar 31;26(1):45-51. [doi: [10.5765/jkacp.2015.26.1.45](https://doi.org/10.5765/jkacp.2015.26.1.45)]
61. An JS, Kim JM, Jung HM. Cognitive behavioral therapy for college students with ADHD tendencies. *Korean J Health Psychol* 2016 Dec;21(4):699-718. [doi: [10.17315/kjhp.2016.21.4.002](https://doi.org/10.17315/kjhp.2016.21.4.002)]
62. Hong S, Lee M, Jung E, Kim C, Son Y. Development of working memory training program for ADHD children and effectiveness verification. *J Curric Eval* 2015 Nov 30;18(3):209-232. [doi: [10.29221/jce.2015.18.3.209](https://doi.org/10.29221/jce.2015.18.3.209)]
63. Chang YI, Park HY. Development and application of the working memory improvement program for children with ADHD in the first grade elementary school. *J Oflearn Cent Curric Instr* 2020 Jul 30;20(14):879-904. [doi: [10.22251/jlcci.2020.20.14.879](https://doi.org/10.22251/jlcci.2020.20.14.879)]
64. Lee KH. The effects of self-monitoring cognitive functions training program on the attention-concentration ability and the hyperactivity of the children with ADHD tendency. *J Emot Behav Disord* 2019 Dec 31;35(4):59-75. [doi: [10.33770/jebd.35.4.4](https://doi.org/10.33770/jebd.35.4.4)]
65. Park W, Park S, Hwang S. [Effects of cognitive behavioral therapy on attention deficit hyperactivity disorder among school-aged children in Korea: a meta-analysis]. *J Korean Acad Nurs* 2015 Apr;45(2):169-182. [doi: [10.4040/jkan.2015.45.2.169](https://doi.org/10.4040/jkan.2015.45.2.169)] [Medline: [25947179](https://pubmed.ncbi.nlm.nih.gov/25947179/)]
66. Kang T, Kim J, Bahn GH, Song SH, Kim J, Kim J, et al. Development of Korean adult ADHD rating scale. *J Korean Acad Child Adolesc Psychiatry* 2015 Dec 31;26(4):295-310. [doi: [10.5765/jkacp.2015.26.4.295](https://doi.org/10.5765/jkacp.2015.26.4.295)]
67. Kim S. (The)clinical utility of K-CBCL 6-18 in diagnosing ADHD: focused on children with psychological disorder in child welfare institution. Sookmyung Women's University. 2016. URL: <https://tinyurl.com/4nu96dj6> [accessed 2024-04-29]
68. Lee SH. Current Status and Future Improvement of the Korean ADHD Rating Scale-IV. (K-ARS-IV.). *Korean Soc Behav Disord* 2015;31(4):227-259 [FREE Full text]
69. Lee SH. A review on the diagnosis of ADHD for special education. *J Emot Behav Disord* 2017 Dec 31;33(4):283-311. [doi: [10.33770/jebd.33.4.14](https://doi.org/10.33770/jebd.33.4.14)]

70. Lee S. A review of diagnosis and evaluation procedure for the child and adolescent with attention deficit hyperactivity disorder. *Korean Soc Emot Behav* 2020 Sep 30;36(3):1-24. [doi: [10.33770/jebd.36.3.1](https://doi.org/10.33770/jebd.36.3.1)]
71. Lee SJ, Shin MS, Kim BY, Yoon H, Shin Y, Kim Y, et al. Clinical utility of the Korean version of CBCL6-18 in the diagnosis of attention-deficit hyperactivity disorder. *Korean J Psychol Assoc* 2015 Nov;34(4):829-850 [FREE Full text] [doi: [10.15842/kjcp.2015.34.4.001](https://doi.org/10.15842/kjcp.2015.34.4.001)]
72. Lee S, Choi J, Kim K, Kim JW, Kim S, Kang T, et al. The guideline of diagnosis and treatment of attention-deficit hyperactivity disorder: developed by ADHD translational research center. *J Korean Acad Child Adolesc Psychiatry* 2016 Dec 31;27(4):236-266. [doi: [10.5765/jkacap.2016.27.4.236](https://doi.org/10.5765/jkacap.2016.27.4.236)]
73. National Research Foundation of Korea. Succeeded in quantifying the level of attention span through meditation and exercise [electronic resources]: expected to be used for diagnostic tests for ADHD, depression, and dementia in children. Korea Creative Content Agency. 2016. URL: <https://tinyurl.com/37tkjjid> [accessed 2024-04-29]
74. Park J. Clinical application of advanced test of attention as a diagnostic tool in children with attention-deficit/hyperactivity disorder. University of Ulsan. 2015. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT1201575064> [accessed 2024-04-29]
75. Cho M. Development of rhythm-based music intervention protocols through timing control in children with ADHD. Kosin University. 2019. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT12023000051273> [accessed 2024-04-29]
76. Choi Y. Development of a music program that improves attention and interpersonal relationships using the Carl Orff teaching method: focusing on a program for ADHD students. Chungnam National University. 2019. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT1201947448> [accessed 2024-04-29]
77. Kim SB. The effects of literary therapy program based on SST by using picture cards on ADHD of adolescents for EBD. *J Spec Educ Rehab Sci* 2016 Jun 30;55(2):1. [doi: [10.15870/jsers.2016.06.55.2.1](https://doi.org/10.15870/jsers.2016.06.55.2.1)]
78. Kim B. Effects of mindfulness-based stress reduction (MBSR) program on attention, perceived stress, and anxiety on attention-deficit/hyperactivity disorder (ADHD) prone university students. Jeonbuk National University. 2016. URL: <https://tinyurl.com/4xzh85cs> [accessed 2024-04-29]
79. Son M. A study on development of diagnostic assessment tools of music therapy in children with attention deficit hyperactivity disorder. Dong University. 2019. URL: <https://dl.nanet.go.kr/search/searchInnerDetail.do?controlNo=KDMT12022000052529> [accessed 2024-04-29]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder

AI: artificial intelligence

AR: augmented reality

ASD: autism spectrum disorder

CBCL: Children Behavior Check List

CBT: cognitive behavioral therapy

CNN: convolutional neural network

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition

fMRI: functional magnetic resonance imaging

ICD-10: International Classification of Diseases, 10th Revision

K-SADS-PL-K: Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version

MBSR: mindfulness-based stress reduction

VR: virtual reality

Edited by T Leung, A Coristine; submitted 09.05.24; peer-reviewed by AV Fial; comments to author 18.07.24; revised version received 06.08.24; accepted 04.09.24; published 15.10.24.

Please cite as:

Cho Y, Talboys SL

Trends in South Korean Medical Device Development for Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder: Narrative Review

JMIR Biomed Eng 2024;9:e60399

URL: <https://biomedeng.jmir.org/2024/1/e60399>

doi: [10.2196/60399](https://doi.org/10.2196/60399)

PMID:

©Yunah Cho, Sharon L Talboys. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 15.10.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Sacroiliac Joint Dysfunction in Endurance Runners Using Wearable Technology as a Clinical Monitoring Tool: Systematic Review

Stuart Evans¹, BA, HBsc, PhD

School of Education, La Trobe University, Melbourne, Australia

Corresponding Author:

Stuart Evans, BA, HBsc, PhD

School of Education

La Trobe University

Plenty Road

Bundoora

Melbourne, 3086

Australia

Email: stuart.evans@latrobe.edu.au

Abstract

Background: In recent years, researchers have delved into the relationship between the anatomy and biomechanics of sacroiliac joint (SIJ) pain and dysfunction in endurance runners to elucidate the connection between lower back pain and the SIJ. However, the majority of SIJ pain and dysfunction cases are diagnosed and managed through a traditional athlete-clinician arrangement, where the athlete must attend regular in-person clinical appointments with various allied health professionals. Wearable sensors (wearables) are increasingly serving as a clinical diagnostic tool to monitor an athlete's day-to-day activities remotely, thus eliminating the necessity for in-person appointments. Nevertheless, the extent to which wearables are used in a remote setting to manage SIJ dysfunction in endurance runners remains uncertain.

Objective: This study aims to conduct a systematic review of the literature to enhance our understanding regarding the use of wearables in both in-person and remote settings for biomechanical-based rehabilitation in SIJ dysfunction among endurance runners. In addressing this issue, the overarching goal was to explore how wearables can contribute to the clinical diagnosis (before, during, and after) of SIJ dysfunction.

Methods: Three online databases, including PubMed, Scopus, and Google Scholar, were searched using various combinations of keywords. Initially, a total of 4097 articles were identified. After removing duplicates and screening articles based on inclusion and exclusion criteria, 45 articles were analyzed. Subsequently, 21 articles were included in this study. The quality of the investigation was assessed using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) evidence-based minimum set of items for reporting in systematic reviews.

Results: Among the 21 studies included in this review, more than half of the investigations were literature reviews focusing on wearable sensors in the diagnosis and treatment of SIJ pain, wearable movement sensors for rehabilitation, or a combination of both for SIJ gait analysis in an intelligent health care setting. As many as 4 (19%) studies were case reports, and only 1 study could be classified as fully experimental. One paper was classified as being at the "pre" stage of SIJ dysfunction, while 6 (29%) were identified as being at the "at" stage of classification. Significantly fewer studies attempted to capture or classify actual SIJ injuries, and no study directly addressed the injury recovery stage.

Conclusions: SIJ dysfunction remains underdiagnosed and undertreated in endurance runners. Moreover, there is a lack of clear diagnostic or treatment pathways using wearables remotely, despite the availability of validated technology. Further research of higher quality is recommended to investigate SIJ dysfunction in endurance runners and explore the use of wearables for rehabilitation in remote settings.

(*JMIR Biomed Eng* 2024;9:e46067) doi:[10.2196/46067](https://doi.org/10.2196/46067)

KEYWORDS

sacroiliac; sacroiliac dysfunction; sacroiliac wearables; sensors; injury management

Introduction

Physical activity, exercise, and sport are increasingly promoted as part of a healthy lifestyle. However, increased participation in physical activity and sport specialization may raise the risk of injury [1]. Running remains one of the most prevalent forms of physical activity, attracting individuals of all capability and ability levels to engage in this form of cardiovascular exercise. However, the burden of running-related injuries and their potential impact on quality of life and societal costs call for research and effective interventions in all the areas associated with sports injury, namely, prevention, assessment, and recovery [2,3]. One of the most overlooked sources of lower back pain (LBP) in endurance runners is injury to the sacroiliac joints (SIJs) [4].

The SIJs are the largest axial joints in the body and sit between the sacrum and pelvic bones on either side. The SIJs connect the spine to the pelvis and facilitate load transfer from the lumbar spine to the lower extremities. Specifically, the SIJs sit between the iliac's articular surface and the sacral auricular surface. Therefore, the SIJ supports the torso and upper body muscular areas to dampen the impact of ambulation as the SIJ can experience forces of shearing, torsion, rotation, and tension when running. To improve and promote efficiency in running while focusing on injury prevention, allied health professionals are exploring different preventative, monitoring, and rehabilitative methods.

Numerous investigations have been undertaken to identify the factors contributing to the management of SIJ dysfunction and the underlying biomechanical mechanisms responsible for pain [3,4]. One consideration is using wearable sensor technology for clinical monitoring. In this regard, wearable sensors (wearables) incorporate a broad range of advances in microelectromechanical systems [5], electrocardiogram [6], electromyogram [7], and electroencephalogram-based neural sensing platforms [8]. As injuries such as SIJ dysfunction can require frequent monitoring, the continuousness of patient/athlete monitoring for timely intervention and rehabilitation seems essential. Wearables present an opportunity to measure the biomechanical parameters of SIJ dysfunction in a continuous, real-time, and noninvasive manner by leveraging electronics packaging technology. It has been conveyed that by leveraging this technology, more time for engagement, continuity of experience, and dynamic data for decision-making for both athletes and clinicians will endure [9]. While remote and ambulatory monitoring are growing needs in the health care environment [10], the efficacy surrounding wearables in remote monitoring relative to SIJ dysfunction remains largely unknown. This is despite the acknowledgment that remote monitoring provides increased data volume and can promote improved athlete performance [11] and accelerate the patient/athlete

rehabilitation processes [12]. Furthermore, an apparent limitation of existing research is that there has been a focus on the effectiveness of wearables on running performance metrics that generally do not consider ongoing rehabilitative considerations [13]. Strategies for the prevention of [14] and recovery from [3] SIJ injury have been proposed, alongside models of injury causation [15] and injury factors [16] (eg, intrinsic vs extrinsic; modifiable vs not modifiable). In turn, this has the potential to help monitor compliance, quality, and progress of movement performance when an injury-prevention or return-to-activity program is implemented [17]. Clinicians and allied health professionals often focus on exploring various training methods for preventive and rehabilitative measures. However, they rarely evaluate these methods in conjunction with biomechanical parameters and their impact on SIJ dysfunction. Thus, there is a need for evidence-based information on how wearables could be used for rehabilitation purposes in a remote setting when SIJ dysfunction is considered.

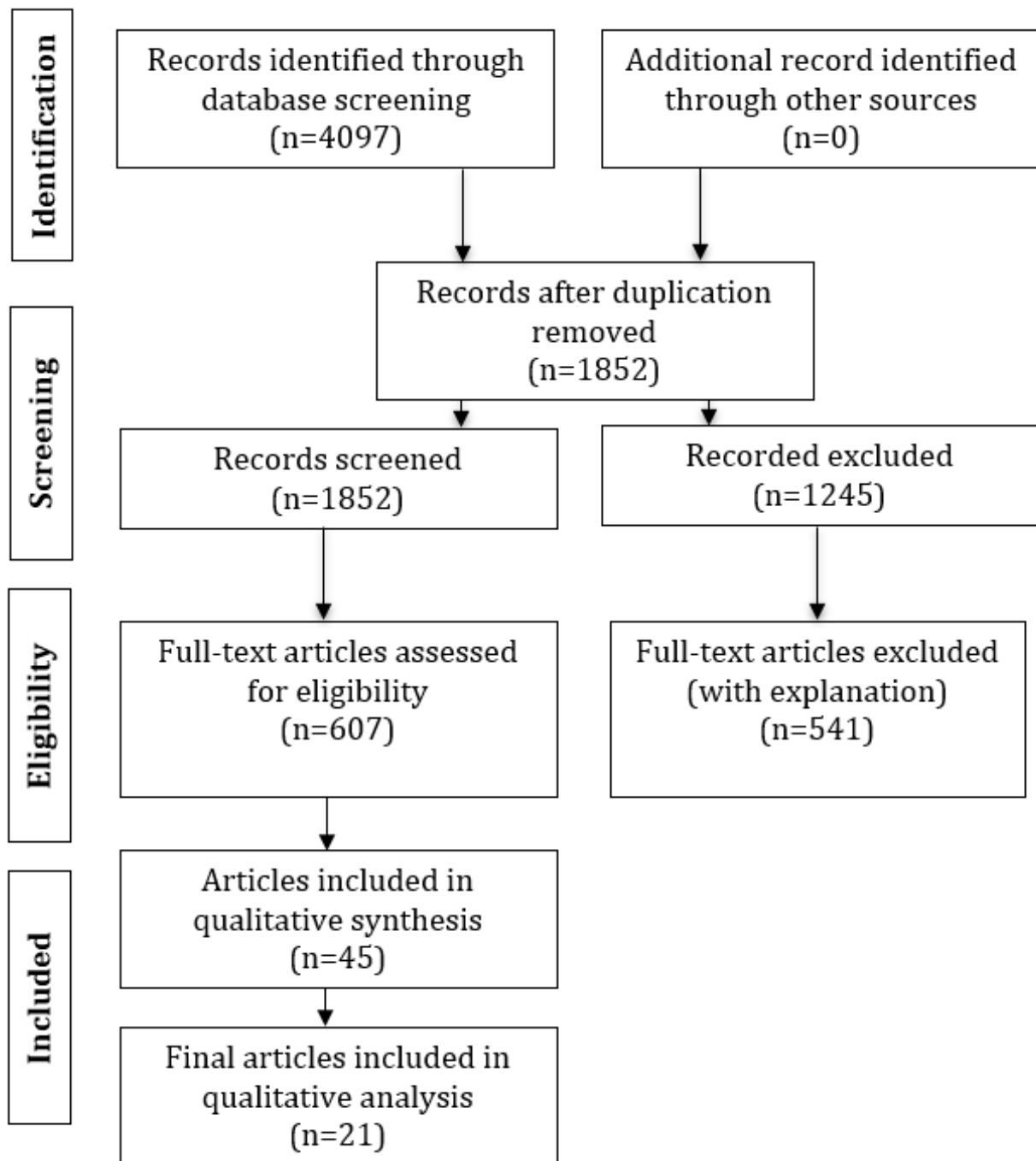
To maintain pace with the rapidly evolving field of wearables in endurance runners, this review provides an update on the state of the literature with a particular focus on literature published in the past 10 years. Case studies illustrate the use of wearable data in the development or monitoring of running programs. For the purposes of this review, a "wearable device" was operationally defined as a device that can be attached to the runner, shoe, or garment, or is a smartphone app. Thus, the purpose of this study was to systematically review the literature and gain a better understanding of the use of wearables in both in-person and remote settings for rehabilitation of SIJ dysfunction in endurance runners. Addressing this issue, the overall goal was to investigate how wearables can contribute to the clinical diagnosis (before, at, and after) of SIJ dysfunction.

Methods

Study Design

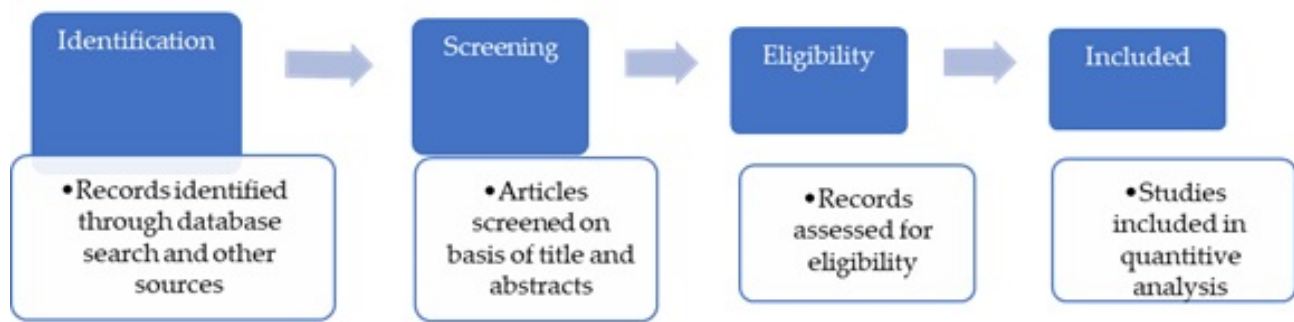
The design and reporting of this review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; [Figure 1](#) and [Multimedia Appendix 1](#) [18]) 2020 statement [18]. The general search strategy ([Multimedia Appendix 2](#)) and search terms are described in [Table 1](#). Articles published up to October 1, 2022, were reviewed.

Thereafter, the selection process consisted of the following steps using the PRISMA guidelines ([Figure 2](#)): (1) an initial title screening for relevant articles was performed once the searched database results had been combined and duplicates had been removed; (2) both the titles and abstracts of the selected articles were then reviewed (a review of the full text was completed if it was not clear from the title or abstract whether the study met the review criteria); and (3) the full texts and selected articles were read based on the inclusion/exclusion criteria.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.**Table 1.** Systematic search strategy and key terms used.

Search strategy	Key terms ^a
Wearable technology	“Wearable Biomechanics” OR “Wearable Technology” OR “Wearable Devices” OR “Wearable Sensors Biomechanics” OR “IMU” OR “Inertial Sensor” OR “Inertial Measurement Unit” OR “Gyroscope” OR “Magnetometer” OR Accelerometer* OR “Pressure insoles” OR “Remote Wearables”
Running gait	“Running Biomechanics” OR “Endurance Running” OR “Run” OR “Jog” OR “Running over 5 km” OR “Endurance Runners” OR “Long Distance Runners” OR “Athletics”
Sacroiliac joint	“SIJ pain” OR “SIJ rehabilitation” OR “SIJ dysfunction” OR “SIJ injury prevention” or “SIJ management”

^aTITLE-ABS-KEY was used as the search strategy.

Figure 2. Steps in the selection process.

A systematic search was conducted to identify potentially relevant papers in the following scientific databases: PubMed, Scopus, and Google Scholar. The focus of this review was on journal articles published in English that described the use of wearable technology to analyze, quantify, and emphasize the use of wearables for remote monitoring of SIJ dysfunction and rehabilitation in endurance runners. This extends to endurance runners undergoing rehabilitation for SIJ dysfunction (ie, had been diagnosed) or the ongoing management of SIJ dysfunction in previously diagnosed endurance runners (ie, rehabilitation). For this search strategy, an endurance runner was considered as someone partaking in regular running-related events (eg, recreational, fun runs) or competitive events (eg, competition, professional, elite). An endurance runner was classified as an athlete running more than 5 km in a single session, either during repeated trials or in studies that classified participants as endurance runners. In line with the main objective, inclusion and exclusion criteria were established to help eliminate studies that were not aligned with the research questions. An

independent coder reviewed subsequent abstracts yielded from the search strategy and then the full articles for study selection. The review screened for information inclusive of health record and research systems including design, functionality, implementation, applications (remote and in-person settings) outcomes, and benefits. The search included articles published between 2000 and 2022. A manual review of the reference section of selected articles was then performed to identify relevant studies missed in the electronic search. Only English language articles were reviewed (Table 2).

Inclusion and Exclusion Criteria

A summary of the inclusion and exclusion criteria is presented in Table 2.

Although no restriction was imposed on the types of wearable technology used in SIJ dysfunction, the search terms were primarily focused on wearable inertial sensors and inertial measurement unit (IMU) devices (Table 3).

Table 2. Summary of inclusion and exclusion criteria.

Study characteristic	Inclusion criteria	Exclusion criteria
Communication type	<ul style="list-style-type: none"> Journal and conference proceedings. 	<ul style="list-style-type: none"> Letters, short communications, technical notes, and other non-peer-reviewed literature. Non-evidence-based guidelines, letters to the editor, and expert opinion papers.
Injury classification	<ul style="list-style-type: none"> Before, during, or after the clinical diagnosis of sacroiliac joint dysfunction which included or incorporated the use of wearables as a viable method of evaluating sacroiliac joint motion. 	<ul style="list-style-type: none"> Articles reporting exclusively on activity monitoring from global navigation satellite systems and injury surveillance without biomechanical measurements.
Classification of wearable	<ul style="list-style-type: none"> Accelerometer, gyroscope, magnetometer, or a combination of these (inertial measurement unit), foot/shoe insoles (pressure mapping). 	<ul style="list-style-type: none"> Temperature sensors, pulse oximeters, pressure sensors, correlated glycemc measurement sensors, biosensitivity techniques, smartphone apps and related sensors, rehabilitation, and monitoring ambulator-based sensors.
Defined running gait outcome measure	<ul style="list-style-type: none"> Spatiotemporal (global outcomes of the running gait cycle): running velocity, acceleration of the center of mass, distance, displacement, ground contact time, step length, step frequency (cadence), stance time, and flight time were included. Kinematics (description of segmental or joint movement, generally in the 3 cardinal planes, namely, sagittal, coronal [frontal], and transverse planes, without consideration for forces). Kinetic (the action of forces in producing or changing motion): for example, ground reaction force, peak pressure, center of pressure, braking, impulse, time to peak pressure, pressure time integral, loads, force time integral, and contact area. 	<ul style="list-style-type: none"> Studies aiming to determine running power or economy were excluded as well as studies investigating walking gait variability or regularity. Studies evaluating robotic systems, exoskeletons, prosthetics, and virtual reality environments were excluded. Studies investigating the use of biofeedback or gait retraining (ie, nonnatural running gait) and studies involving the use of altered weight conditions (eg, wearable resistance, antigravity treadmills, or water-based protocols). Computer algorithms; machine learning or statistical approaches; and those using robotic systems, exoskeletons, prosthetics, and virtual reality environments.
Participant	<ul style="list-style-type: none"> Age >18 years, male and female. Endurance running included runners regularly completing over 5 km in training or competitive situations. The endurance runner was partaking in regular running-related events (eg, recreational, fun runs) or competitive-based events (eg, competition, professional, elite). The runner was classified as an athlete running more than 5 km in a single protocol session, either during repeated trials or in studies that classified participants as endurance runners. 	<ul style="list-style-type: none"> Age <18 years. Endurance runners not regularly completing over 5 km in training or competitive situations. Studies done on animals and cadavers.

Table 3. Comparative overview of wearable sensor modalities used in running today.

Category	Technology	Capabilities	Pros	Cons
Soft tissue injury prevention	Surface electromyogram	Identifies muscle recruitment and potential weaknesses	Small, wireless, and provides live data	Low signal-to-noise ratio
Workload management and athletic performance	GPS, inertial measurement unit, and accelerometers	Distance, velocity, acceleration, deceleration, mediolateral movement, work, power, dehydration, fatigue, athletic performance, detecting gait parameters	Good range of data points	No biometric data and GPS can be pricey
Cardiac health	Electrocardiogram/photoplethysmography sensors	Heart rate, sleep rate, heart rate variability, respiration, muscle oxygen saturation, atrial fibrillation, stress levels, respiration rates, blood volume, and body temperature	Accurate and cost-effective	Price point

^aThe table presents a comparative overview of common wearable sensors currently available rather than the components used for analysis (ie, some studies used an inertial measurement unit, but only analyzed data from 1 element of the unit).

Study Classification and Assessment

The selected studies reported multiple feature domains, including (1) strength of evidence, time setting, and primary scope; (2) study characterization in terms of experimental conditions, setting (running field based and running laboratory based using treadmills), and age of endurance runners tested; and (3) characteristics of the technologies and types of wearable device and measures used relative to SIJ dysfunction. The author also defined and assessed (4) the Injury-research Readiness Level (IrRL) relative to SIJ dysfunction.

Selection Process: Strength of Evidence, Time Setting, and Scope

The strength of evidence for each article was assessed across 3 main categories, ordered in decreasing strength based on the experimental design used: experimental, that is, meeting the requirements of endurance running and SIJ dysfunction at or after clinical diagnosis and injury; randomized controlled trials; quasi-experimental, that is, including manipulation of the experimental conditions under which participants performed endurance running, but lacking random assignment or group comparison; and observational, that is, without assessing the effects of an intervention, and only describing participant behavior [19]. A separate class was used for studies looking exclusively at the validation of new equipment or methods. Literature reviews on wearables combined with synergies in remote settings or endurance running-related SIJ injuries were included and assessed by the primary author.

Classification and Characterization of SIJ Dysfunction

Studies were required to classify and characterize the diagnosis of SIJ dysfunction. Therefore, akin to Preatoni et al [20], an “at/post” classification was used to express the chronological relationship between the experimental data collected and the SIJ dysfunction in endurance runners. Thus, studies were classified as the at category if they were identifying and classifying SIJ injury factors, diagnosis, or underlying mechanisms, and therefore, attempted to capture or track SIJ injury occurrences in endurance running (eg, cohort studies with biomechanical screening and in-field injury events that

referenced use of wearable technology). Studies were classified as post if the data collection was performed after the SIJ injurious event, that is, during the SIJ recovery phase with the aim focused on rehabilitation techniques in both field-based and laboratory environments where the endurance runner had received a clinical diagnosis of SIJ dysfunction. The post classification was also used for studies that assessed the likelihood of SIJ injury or a greater magnitude of dysfunction. For clarity, studies that examined endurance runners who had returned to full running activity (eg, comparisons between healthy individuals and those with a history of a specific or existing SIJ injury) were classified as pre because they were not centered on the recovery process that goes from injury occurrence (or medical intervention, if relevant) to being able to return to full running activity.

The characterization of studies was based on the following categories: (1) studies analyzing preexisting running-related SIJ dysfunction using wearable technologies to monitor running biomechanics in both a field-based and laboratory setting for the purpose of clinical management and clinical management in a remote setting (pre); (2) studies assessing endurance running-related SIJ dysfunction or injury factors or injury risk using wearable technologies to monitor running biomechanics after SIJ dysfunction has been formally diagnosed and classified as in the acute stage of injury in both field-based and laboratory settings for the purpose of clinical management (at); (3) studies assessing ongoing running-related SIJ injury factors or injury risk using wearable technologies to monitor running biomechanics after SIJ dysfunction has been formally diagnosed and classified as in the chronic stage of injury in both field-based and laboratory settings for the purpose of clinical management or management in a remote setting (post); and (4) studies attempting to establish injury threshold criteria from a biomechanical perspective, studies characterizing protective wearable devices, and studies focusing on post-SIJ injury monitoring or return-to-run assessment using wearables. Validation and literature review studies were classified according to the primary aim for which the method or tool tested had been devised, as stated by the authors.

To describe the experimental conditions, information was extracted about the settings in which data were collected (ie, laboratory vs field based). Specifically, studies were labeled as field based if wearable-obtained data were acquired during a scheduled running training event, a simulated running training event, or a running competition in a specific setting. Conversely, investigations carried out within a laboratory or in the field but using wearable technologies were labeled accordingly. The stage of SIJ dysfunction addressed by the study was then classified as either chronic (caused by overuse) or acute (resulting from specific events), following the criteria outlined by Bahr et al [15]. Furthermore, annotated classification of the endurance runner was addressed by each study (ie, recreationally active, trained/developmental, highly trained/national level, elite/international, world-class, or not specified/insufficient data to be classified) [21]. The risk of bias was assessed by the primary author.

Injury-Research Readiness Level

Building on the System Readiness Level framework by Sauser et al [22], an IrRL was modeled to capture the maturity, functionality, and readiness of the studies aiming to contribute to preventing, assessing, or recovering from SIJ dysfunction. According to the System Readiness Level model, technology and system development follow similar maturation paths, whereby technology is inserted into a system and interacts via a proposed architecture. Knowing about the system components and their integration is important, and this knowledge allows a classification of the system as being in its research, development, or deployment stage [23]. In the context of SIJ dysfunction and endurance running-related injuries, for this review, a method is deemed mature for deployment only when it relies on measuring wearable tools that are characterized by high ecological validity (ie, fully wearable and unobtrusive or markerless), can be applied directly in the field, is supported by validation studies against an established gold standard, or when validation is not practicable but adheres to standardized experimental procedures. Specifically, the biomechanical quantities pertaining to the SIJ should demonstrate evidence of a causal relationship with SIJ dysfunction and management in endurance running, and their interpretation should be driven by specific guidelines (eg, individual- or population-based normative boundaries, thresholds, or trends; [Multimedia Appendix 3](#)).

Data Extraction and Collection

After the data search was complete, data were obtained and extracted from eligible studies in a custom form that was created in Microsoft Excel. The form included (1) author, title, journal, and publication year; (2) research design; (3) sample size; (4) participant characteristics (eg, age, gender); (5) intervention features (type, length, and frequency); (6) measures and settings (laboratory, field-based, the type of wearable technology used, and sensors); (7) analysis; (8) key findings relative to the pre, at, and post categories for clinical SIJ dysfunction management using wearables in a remote or clinical setting; and (9) research outcomes, the metrics used, and conclusive statements. Data were then synthesized into a table format in Microsoft Excel

and confirmed for data entry by the author. No automation tools were used in the process.

Results

Overview of Identified Articles

From the 4097 articles identified through the database search (Google Scholar, n=2263; Scopus, n=1624; and PubMed, n=210), and after removing duplicate items, 2245 publications were excluded based on title, abstract, and inappropriateness of topics (eg, knee arthroplasty in endurance runners). A further search was then performed in the databases with exclusion criteria (without the words) “knee” AND “lower back” AND “hip.” A search “with the words” was then refined to include “remote.” An additional 551 articles were removed due to “knee” appearing in the article while 2 papers were removed due to not being written in English. A further 4 were removed due to the topic being limited to physiological assessments only. A total of 1295 articles remained. Of these, 585 articles were discarded (most frequent reasons were not including wearables, not mentioning SIJ injury or SIJ dysfunction or running-related activities, and not describing the relationship between biomechanical quantities from wearables and the SIJ, or not defining the IrRL classification model relative to the SIJ and wearable usage in endurance runners). In addition, 665 records were removed due to technology not being classified as wearable, yielding a total of 45 studies to be considered for review.

A total of 151 participants were identified as being runners or endurance runners from the 45 papers analyzed. Descriptions of the included studies were either classified as a review of wearable sensors in the diagnosis and treatment of SIJ dysfunction, or wearable movement sensors for rehabilitation, or a combination of the above for SIJ gait analysis in an intelligent health care setting. Two papers [24,25] specifically mentioned wearable technology and the COVID-19 pandemic. However, only 1 of the review papers specifically mentioned measuring biomechanical loads and asymmetries in elite long-distance runners through inertial sensors [26]. One study [27] reported on SIJ pain relative to contralateral pelvic drop compared while the remaining research papers specifically mentioned iliac stress fractures in endurance runners linked to the SIJ, hip pain, or SIJ dysfunction. The remaining studies did not openly discuss the link between wearables and remote settings and SIJ dysfunction but mentioned such relationships as being possible or hypothetical. Thus, a total of 21 manuscripts remained, with overlapping reports on topics relative to SIJ dysfunction. No immediate forms of information bias (measurement bias) were detected in the final 21 studies.

Journals and Years

The 21 original manuscripts included in the review appeared in over 11 different journals, with 11 journals publishing nearly half of the total, and at least five relevant articles published in orthopedic, traumatology, or physical therapy journals. One paper was published in a rehabilitation journal while 3 papers were published in technology and engineering journals. The number of articles in the area under scrutiny appears to have increased over time, as 7 papers have been published since the

onset of the COVID-19 pandemic regarding telehealth (remote), sensors, and machine learning in endurance running injury management journals. The use of wearables in field-based and self-reliant monitoring seems to be increasing in popularity, as also demonstrated by the 7 review papers published between 2020 and November 2022.

More than half of the 21 studies scrutinized were literature reviews, 4 (19%) were case reports, and 1 was classified as fully

experimental; 5 (24%) attempted to develop a predictive model or a machine learning approach to identify risk factors for running-related SIJ dysfunction. One study was classified as being at the pre stage of SIJ dysfunction, while 6 (29%) were identified as being at the at stage of classification. Considerably fewer studies attempted to capture or classify actual SIJ injuries, and no study directly addressed injury recovery ([Table 4](#)).

Table 4. Validity and reliability and application (information extracted from each article included the classification of the study).

Author	Year	Location	IrRL ^a	Classification	Participants and gender, n	Age (years)	Metric(s)
Alcantara et al [28]	2021	Force-measuring treadmill (laboratory)	IrRL1: Research (exploring causal relationship)	Validity (at ^b)	37	Mean 20 (SD 2) years	Quantified accuracy of applying quantile regression forest and linear regression models to sacral-mounted accelerometer data to predict peak vertical ground reaction force, vertical impulse, and ground contact time across a range of running speeds.
Whitney et al [27]	2022	Treadmill (laboratory)	IrRL2: Development (building on established causal relationship)	Case-control (at)	81 runners (63 runners without SIJ ^c pain and 18 runners with SIJ pain)	Mean 27.3 (SD 12.9) years for runners without and 23.8 (SD 10.5) years for runners with SIJ pain	In midstance, runners with SIJ pain had greater contralateral pelvic drop compared with controls. For unilateral SIJ pain cases (n=15), greater contralateral pelvic drop was observed when loading the affected side compared with the unaffected side. Female runners with SIJ pain demonstrated greater contralateral pelvic drop during the mid-stance phase, along with less knee flexion, greater "tibial overstride," and greater ankle dorsiflexion at initial contact compared with controls.
Höfer and Siemsen [29]	2008	Treadmill (laboratory)	IrRL1: Research (exploring causal relationship)	Application (at) (proof of concept)	3 male participants	N/A ^d	The pressure between the sensor contact area and the lumbar region was measured with force sensitive resistor sensors.
Amorosa et al [30]	2014	N/A	IrRL1: Research (exploring causal relationship)	Review	1 female participant	24 years	Report on a second case of an isolated stress fracture of the iliac wing in a female marathon runner and the associated diagnosis of the female athlete triad.
Ueberschär et al [26]	2019	Treadmill (laboratory)	IrRL2: Development (building on established causal relationship)	Experimental (pre ^e)	45 healthy junior-elite long-distance runners	N/A	The mean peak tibial accelerations in junior-elite long-distance runners ranged between 14 (SD 3) and 16 (SD 3) g ($g \approx 9.81 \text{ m s}^{-2}$) for running speeds of 14–16 km h^{-1} . The corresponding mean peak sacral and scapular accelerations amounted to 4 (SD 1) to 5 (SD 1) g (32%, SD 8% of tibial load) and 4 (SD 1) g (mean 27%, SD 6%), respectively.
Liu et al [31]	2021	N/A	IrRL1: Research (exploring causal relationship)	Review	N/A	N/A	Daily monitoring of basic health data by wearable devices helps physicians in detecting the health problem. However, most current wearable sensors are not accurate enough for clinical evidence.
Banos et al [32]	2015	Laboratory	IrRL1: Research (exploring causal relationship)	Application (proof of concept/case report)	1 male participant	N/A	A novel mobile health system to support trunk endurance assessment. The system uses a wearable inertial sensor to track the patient's trunk posture, while portable electromyography sensors were used to seamlessly measure the electrical activity produced by the trunk muscle.

Author	Year	Location	IrRL ^a	Classification	Participants and gender, n	Age (years)	Metric(s)
Falowski et al [33]	2020	N/A	IrRL1: Research (exploring causal relationship)	Review	N/A	N/A	A review and algorithm for the diagnosis and treatment of sacroiliac joint pain.
Zadeh et al [34]	2021	Laboratory	IrRL2: Development (building on established causal relationship)	Application (at) (proof of concept)	55 (39 male and 16 female participants)	21.1 (SD 3.84) years for male and 20.1 (SD 1.18) years for female participants	Proof of concept that wearable technology has the potential to predict injury in sports.
Porciuncula et al [35]	2018	N/A	IrRL1: Research (exploring causal relationship)	Review	N/A	N/A	Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances.
Lorussi et al [36]	2018	Field based	IrRL3: Deployment	Application (at) (proof of concept)	N/A	N/A	A wearable system for remote monitoring of the treatments of musculoskeletal disorder.
Shen et al [24]	2021	N/A	IrRL1: Research (exploring causal relationship)	Review	N/A	N/A	Digital technology-based telemedicine for the COVID-19 pandemic.
Nascimento et al [37]	2020	N/A	IrRL1: Research (exploring causal relationship)	Review	N/A	N/A	Sensors and systems for physical rehabilitation and health monitoring.
Channa et al [25]	2021	N/A	IrRL1: Research (exploring causal relationship)	Review	N/A	N/A	The rise of wearable devices during the COVID-19 pandemic: a systematic review.
Rahlf et al [38]	2022	N/A	IrRL1: Research (exploring causal relationship)	Application (at) (proof of concept)	N/A	N/A	Proof of concept using runners who run at least 20 km. A prospective longitudinal cohort study using statistical analysis of the data was performed using machine learning methods.

^aIrRL: Injury-research Readiness Level.

^bAn at/post classification: if the scope was to identify and characterize SIJ injury factors, diagnosis, or underlying mechanisms; or track SIJ injury occurrences in endurance runners.

^cSIJ: sacroiliac joint.

^dN/A: not applicable.

^ePre: pre-SIJ dysfunction (ie, before the SIJ injury).

Experimental Setting

In the field-based study [36] that analyzed endurance runners at the SIJ dysfunction stage, the application was at the proof-of-concept stage only. None of the studies included in this review were deemed to be experimental or classified as an observational study design pertaining to the use of wearables in a self-monitoring or remote rehabilitation capacity. This was despite most studies being literature or systematic reviews that focused on wearables for self-monitoring, self-monitoring in a remote setting, or a combination of both.

Participant Characteristics

Overall, the studies included between 1 participant [30] and 81 participants [26], with the mean number of participants being 21 (SD 32). The mean age of participants was 22.2 (SD 3.7) years. None of the selected studies performed a comparison of SIJ dysfunction and related gait patterns across the selected age groups or compared SIJ dysfunction using a validity approach

in wearables. Many of the studies included both male and female participants; however, none of the selected studies examined differences between male and female participants in SIJ dysfunction using wearables. One study [26] focused on female runners with SIJ or sacral stress fractures, whereas another [29] included only male participants using pressure sensors in the lumbar region. Given the discrepancy in participant characteristics, a source of inequity, that is, gender bias, was prevalent in some studies analyzed.

Clarification of SIJ Pathomechanics

Overall, the SIJ appears to function as a stabilizer of the pelvis, absorbing ground reaction forces during gait and shear forces during movement [6]. The SIJ has also been described as a multidirectional force [39]. Activities that involve a 1-leg stance such as running would presumably increase the force in each SIJ, yet this was not specifically mentioned in the studies. Similarly, this would influence the vertical ground reaction force that occurs with each step. Another significant influence

is the center of mass, which is in slightly different positions for men and women. One study noted the importance of the center of mass, particularly in women, as it commonly passes in front of or through the SIJ [40]. Some of this can be explained due to sexual dimorphism being apparent in the pelvis, with the female sacrum being wider and with a more backward tilt. This would also account for the higher loads and stronger SIJs that are commonly seen in men [41]. This characteristic may also explain why men have more restricted mobility, as the average movement for men is approximately 40% less than that of women [42]. In this regard, the mechanism of SIJ dysfunction is primarily a result of a combination of axial loading and abrupt rotation [43]. DeRosa and Porterfield [44] delineated the primary influences as follows: the force of gravity, which acts downward through the spine, generating the flexion moment of the sacrum on the ilium, and the ground reaction force, which travels upward through the lower extremity from the heel strike, producing a posterior rotational moment (referred to as “torsional”) of the ilium on the sacrum; they termed these motions sacroiliac and iliosacral, respectively. Falowski et al [33] presented an algorithm for the diagnosis and treatment of SIJ pain. In this case, the authors believed that SIJ pain is an

underdiagnosed and undertreated element of LBP. Citing an emerging disconnect between the growing incidence of diagnosed SIJ pathology and the underwhelming efficacy of medical treatment, they created a diagnostic and treatment pathway to establish an algorithm for patients that can include conservative measures and interventional techniques once the diagnosis is identified.

Classification of Wearables

A total of 8 studies used wearables in some form; however, only 1 study [26] used a sensor (a triaxial accelerometer) to measure biomechanical loads in endurance runners, although this study did not specifically review SIJ dysfunction. In the 8 studies that mentioned wearables, accelerometers and gyroscopes featured; however, the authors did not provide enough information to establish the type, range, and technical specification of the devices. There was a large variation in the reported use of temperature sensors, pulse oximeters, BioHarness wearable technology, pressure sensors, correlated glycemic measurements, biosensitivity techniques, electrodes, environmental monitoring, smartphone accelerometers, and next-generation wearable movement sensors despite these studies not specifically mentioning SIJ in endurance runners (Table 5).

Table 5. Breakdown of various approaches used for wearables.

Approaches	Description
Referred to sensors’ validation within the cited article	<ul style="list-style-type: none"> Compared with gold standards (eg, stereophotogrammetry, force platforms, high-speed video, or photocells) [24,25,34-37]. Comparing classification results against human\validated software classification [24,25,35-38].
Pilot or proof studies	<ul style="list-style-type: none"> Biomechanical effect of a lumbar spine-relief orthosis for the treatment of sacroiliac pain [29].
Referred to ad hoc procedures for the performed measures	<ul style="list-style-type: none"> Describing procedures for sacroiliac joint monitoring or pain management measures using machine learning or similar approaches [45,46].

The reviewed studies that used proof-of-concept designs [34,38] included generic descriptions of wearables relating to self-monitoring use and remote rehabilitation monitoring despite inadequate information provided about SIJ for rehabilitation in endurance runners. Furthermore, while describing the technical features of the wearable is key to the accurate clarification of data quality and of the implication of the changes that a remote intervention may encourage, many studies did not report this information sufficiently. Notably, and as highlighted by recent systematic reviews on wearables and inertial sensors for sport performance evaluation [47], and on accelerometry of impact loading in runners [30], reporting the features of the wearable device used—as well as information on the attachment location and fixing methods—is essential.

Discussion

Principal Findings

This review examined 21 studies that evaluated the effects of wearable use in remote settings during SIJ dysfunction in endurance runners. A secondary purpose of this review was to evaluate the effectiveness of wearables in possible or probable

SIJ rehabilitation programs for endurance runners. Explicitly, this review reported on the (1) strength of evidence, time setting, and primary scope of studies relating to SIJ dysfunction in endurance runners; (2) characterization of SIJ dysfunction in terms of experimental conditions, setting (running field based or running laboratory based using treadmills), and the age of endurance runners tested; and (3) characteristics of the technologies and types of wearables and measures used relative to SIJ dysfunction in endurance runners. The author also defined and assessed (4) the IrRL relative to SIJ dysfunction. This review has demonstrated that the use of wearable technology for SIJ dysfunction monitoring in endurance running either from a laboratory or from a remote (telehealth) perspective is emerging, but further work is required to establish a standardized methodology and the validity or reliability of instrumentation.

This review provides a comprehensive overview of wearable technology used for an SIJ dysfunction in endurance runners as well as recommendations for future work.

Injury Type and Classification

The quality of the included studies varied, with one of the most challenging aspects of diagnosing and treating SIJ dysfunction

in the endurance running population being the inconsistent judgment and, in some instances, worrisome presentation of the injury. The main difficulty faced by authors appears to be related to diagnostic challenges given that the pathomechanics and diagnostic classification of SIJ dysfunction are inconsistent in the literature. This was mainly observed in studies that referred to SIJ dysfunction as either a potential source of LBP or symbolic of hip-related issues. Moreover, lumbopelvic rhythm (LPR) was used as a definitive term by some authors. This, then, makes any possible deployment of wearables for rehabilitation purposes challenging if the diagnosis is either missed or misdiagnosed. As specific characteristics of SIJ dysfunction in endurance runners are required for investigation, the number of eligible participants was limited given that acute injuries were investigated primarily in 1 study [30] and chronic SIJ dysfunction in another [27], both of which occurred in control settings. None of the studies monitored acute or chronic SIJ dysfunction using wearables in a remote setting.

There were additional variations among the reviewed studies. While 2 studies examined the usability of wearables through active engagement with endurance runners [27,38], many lacked consideration for the wearer's physical, psychological, and social preferences regarding the technology. Although 1 proof-of-concept study examined if wearable technology has the potential to predict injury in sports [34], many studies (42%) were found to be at the at stage of injury classification. However, it is important to consider the practicality of using wearables to classify SIJ dysfunction at the pre stage during running. Further research exploring the feasibility and necessity of using wearables is required, or whether this is feasible given the apparent difficulty in diagnosing SIJ dysfunction. Additional research will enhance our understanding of how wearables could be used at the onset of possible SIJ dysfunction to deliver the most pertinent data while enabling a clinical diagnosis.

A major issue in the approach to wearable instrument application is that more than half of the 21 studies analyzed were literature reviews, 4 (19%) were case reports, and 1 was classified as fully experimental relative to the classification of SIJ dysfunction. The results showed that although different wearables have been used for evaluating biomechanical parameters in the running gait analysis, as well as some relevant SIJ parameters pertaining to diagnostic or predictive stages of SIJ dysfunction, a paucity of research exists in the rehabilitation and remote monitoring of SIJ dysfunction. Indeed, the findings show that different descriptions related to possible or probable SIJ diagnosis exist in that injury classification is also referenced in relation to LBP and LPR. This, then, makes it difficult to draw firm conclusions regarding how wearables could be deployed remotely for rehabilitation purposes. Therefore, we are beginning to understand that the at stages of SIJ dysfunction require more than a concentration on the risk factors associated with injury occurrence.

Evidence also suggests that SIJ rehabilitation using wearable technology, in both controlled and remote settings, is highly nuanced (ie, varying across classification, injury stage, diagnosis, participant age, and gender). This complexity may extend to confusion in terminology and diagnosis between lower back injury and SIJ dysfunction, considering potential

differences in running gait mechanics when running in controlled (eg, laboratory) versus remote settings. For example, one study [48] noted that the most common complaints were pain in the lower back, buttocks, leg, groin, and hip. Although some studies acknowledged that pain originating from the lower back region is likely more common than most endurance runners realize, as a result of the difficulty in localizing symptoms and referred pain patterns, the results suggest that reference to running-related SIJ issues was infrequent. This is not necessarily surprising as LBP is among the most common human health problems and accounts for a significant amount of disability worldwide [49]. Interestingly, the SIJ has been estimated to contribute to pain in as much as 38% of cases of LBP [50]. Although topographical classifications such as "sacroiliac," "pelvis," and "spine" serve a crucial didactic purpose, they can impede understanding of normal and altered functional SIJ mechanisms. As different classifications exist, it remains somewhat unknown if greater SIJ dysfunction in endurance runners exists, thus making any reference to the possible role of wearables relative to injury classification and rehabilitation in remote monitoring challenging.

What is commonly stated among the papers reviewed is that the clinical examination of an endurance runner with SIJ dysfunction commonly begins with an evaluation of gait. The results suggest that this often commenced in a clinical setting with ongoing monitoring of the condition commonly requiring the patient to be in the same clinical and controlled setting. It is at this juncture that wearables could be used in a remote and personalized setting, whereby data are fed to the clinician to monitor and track gait-related patterns or irregularities. Notwithstanding the literature reviews discussed in this paper that highlight the obvious and practical gap in using this technology in an SIJ dysfunction setting, more research is needed to test the feasibility and validity of the different wearable devices currently available. This extends to the level of expertise needed to operate and interpret the data from the perspective of an operator, athlete (runner) and clinician. Additionally, the results point to LPR being frequently referenced in the literature alongside LBP and SIJ dysfunction. The literature suggests that LPR is the relationship between the lumbar spine, hip, and pelvis when the trunk is in flexion. The classification of LPR during torso forward bending and backward return has also been widely investigated and commonly related to lower back disorders [51]. This defines LPR and LBP without necessarily drawing on the biomechanical differences and classification of how these injuries are managed in endurance runners. Furthermore, the results show considerable differences in the methods used to measure, and approaches used to characterize, LPR. Overall, it appears as though the timing aspect of LPR has been examined to obtain insights into the neuromuscular control of torso motion. The lack of consensus in LPR, LBP, and SIJ dysfunction is further impacted by the fact that there are no "gold-standard" algorithms for the detection of running gait outcomes from wearable sensor setups, which likely explains the large variation of outcomes and definitions reported in the reviewed studies.

Treatment of SIJ Dysfunction

It appears that treatment and management of SIJ dysfunction are often nonsurgical and involve packages of care that can include analgesics, physiotherapy, corticosteroid injections, and radiofrequency ablation [52]. Non-face-to-face (remote) care models exist in which the athlete is physically separated from the physician (or other health care workers) and empowered by communication-based technologies, such as videoconferencing and the use of continuous patient monitoring (wearable or “surface sensor”) technologies that capture athlete metrics and deliver health data remotely to the physician. Although these technologies have existed for some time, widespread implementation has been constrained by laws, regulations, and policies. The use of wearables in movement science and sport is widespread [53]; however, relative to SIJ dysfunction detection using wearables in either a laboratory/clinical setting or a remote setting, it could be argued that their application is still in an “exploratory phase.” Therefore, the findings agree with Hughes et al [54] in that the technology and the associated methods still require further development and careful analysis.

The results concerning SIJ injury risk mitigation have been well addressed in the literature [55,56]. Notwithstanding injury mitigation factors, no exploratory research has been performed to systematically investigate the feasibility of wearables use as a rehabilitation tool in SIJ injury assessment or dysfunction in endurance runners. This includes how wearables could potentially be used to characterize the severity of SIJ dysfunction as well as exploring the use of acquired information to support either clinical preventive or rehabilitative interventions. The empirical and analytical study of SIJ motion dates back to the late nineteenth century. However, its widespread acceptance as a legitimate entity has only occurred recently [57,58]. This delay in acknowledgment may elucidate why SIJ dysfunction can often be mistaken for LBP and LPR issues. Moreover, nowadays, the topography of SIJ motion should be measured to establish the conceivable axes of motion. From the study of Wilder et al [59], translation must occur for any sagittal innominate rotation to be possible because of the irregular surfaces and taut ligament structure. Accordingly, clinical theories have been proposed regarding the details of these motions. Along this line, Lee et al [6] stated that nutation seems to occur bilaterally when moving from supine to standing and unilaterally with flexion of the hip joint. Moreover, this kind of information would be relevant to any treatment of SIJ dysfunction given that counternutation occurs bilaterally and sometimes near the end of trunk flexion and unilaterally during hip extension. Some authors (eg, [60]) suggested that individuals with SIJ dysfunction display symmetrical gait and a depressed synergy between muscles providing SIJ force closure. The disorder involves reduced coactivation of the gluteus maximus and contralateral activation of the latissimus dorsi, which together provide joint stability during running. The disorder would be exacerbated in endurance runners given their need for maximum activation of gluteus maximus and torso stability, both of which require consideration when treating SIJ dysfunction. Nevertheless, these results indicate that the information on SIJ dysfunction in endurance runners and the treatment options that exist using wearables are unrepresented.

Despite these limitations, it is pertinent to consider whether such treatment methodologies are clinically and practically feasible within a given wearables context.

Information Technology and Health Care

Outcomes obtained from this review posit that health services have experienced great changes, especially in remote monitoring [61]. Additional clinical studies (eg, [31]) have shown that wearables are widely used to monitor functional and daily activity inclusive of walking and running gait. The wearables used were commonly integrated with an IMU sensor and controlled with a smartphone app [62]. The increased use of wearable technologies, either in isolation or as part of integrated, preventative, or rehabilitative approaches, offers an opportunity to collect quantitative data “in the field,” less obtrusively, for extended periods, and with fewer spatial limitations than conventional motion-capture technologies (eg, [46]). In this regard, wearables are increasingly viewed as promising alternatives to expensive analytical instruments in health care when specificity and selectivity criteria are met. It could be that wearables are used to monitor for possible pain, therefore exploring the use of torso acceleration as a proxy with a triaxial accelerometer. As the goals of SIJ dysfunction treatment may include increasing suppleness, strengthening, and correcting any asymmetries, the opportunity remains to explore how wearables could be used as a viable treatment monitoring option. This, then, is an area for future research.

Wearables can help quantify spatiotemporal variables (eg, stride, step length, cadence) and physiology (eg, heart rate, recovery time) and are commonly used for human activity detection and quantified self-assessment. Until recently, or specifically since the emergence of the novel coronavirus, COVID-19 in January 2020, evidence for the effectiveness of remote usage and wearable monitoring, compared with traditional care models, has been scarce [63]. Along this line, the combination of telemedicine as an audiovisual communication platform and wearables that transmit field-based kinematic metrics provides numerous benefits to both health care providers and runners alike. Similarly, machine learning approaches have been widely used in gait biomechanics studies in the past decade [64-66]. However, among the papers included in this review, only 3 [31,32,36] focused on wearables for the sole purpose of remote monitoring of treatment of musculoskeletal disorders, clinical advances, and rehabilitation. This ambiguity further complicates the usage and uptake of wearables for SIJ dysfunction, which need to accommodate such conditions.

Although wearables can be used for home monitoring of activity and for the purposes of rehabilitation, little research has examined the potential of wearables when applied to acute or chronic SIJ dysfunction in endurance running. For example, when used remotely (ie, at home), the wearer (runner) could be required to complete standardized functional, rehabilitative assessments while data are continually recorded from the wearable device and relayed directly to the doctor or medic. Therefore, rather than comprising only standardized functional test data, as would be the case in a clinical setting, the runner’s ambulatory movement data set would contain data corresponding to all movements while wearing the sensor, including recovery

and running activity. Indeed, common day-to-day movements can be tracked using wearable devices equipped with an IMU sensor and controlled with a smartphone app [62]. Besides research into wearable use in stride, step, stance, and spatiotemporal variables relative to both performance and injury mitigation, a greater understanding of the processes and predictors of SIJ rehabilitation has the potential to inform and strengthen public health. In this regard, the findings agree with Regterschot et al [67] in that important challenges and barriers to the deployment of wearables in clinical care remain. Similarly, Lang et al [68] discussed the major barriers to the application of wearables in motor rehabilitation and proposed benchmarks for the implementation of wearables in clinical practice. These clinical barriers include the demanding clinical environments that are often present, as well as the lack of recognition by some health professionals of the valuable information that can be obtained from wearables. There are also technology-related barriers, including (1) wearables that are inaccurate for many athletic populations (ie, inconsistent data output or lack of validity), (2) wearables that are not user-friendly for clinicians or athletes, and (3) the lack of published data on the reliability and clinical validity of some wearables. This extends to the development and optimization of innovative wearable configurations and data analysis techniques (eg, machine learning–based algorithms that enable the detection of specific activities and movements in free-living conditions). While Regterschot et al [67] asserted the existence of reliable and valid wearables for clinical populations and free-living environments, medical technology professionals could be encouraged to assist allied health specialists in developing the knowledge and skills necessary to effectively use wearables for remote rehabilitation purposes. In concordance with Regterschot et al [67], barriers exist in deploying remote wearables for detecting specific activities and movements in free-living conditions. The results of this review suggest that clinical barriers extend to the busy medical environment and the lack of realization of the value of information that can be obtained using wearables. However, it appears as though technological barriers also exist, including (1) a perception that wearables are inaccurate for many patient populations, (2) wearables that are not user-friendly for clinicians or patients, and (3) a lack of published data regarding reliability and clinical validity of sensor systems. Relatedly, Lang et al [68] discussed the clinical barriers to the application of wearables in motor rehabilitation and proposed benchmarks for the implementation of sensors in clinical practice. Therefore, researchers are encouraged to investigate the usability, acceptance, feasibility, reliability, and clinical validity of wearable sensors in clinical populations to facilitate the application of wearable movement sensors in SIJ rehabilitation.

Limitations

Some caution should be exercised when considering these findings. It merits noting that this review was a single-author systematic review. The author performed manual searches of all databases stated in this review and then coded and analyzed all retrieved results. Despite this, being a single-author review ensured that the processes described were based on the author's judgment of eligible articles, albeit following the PRISMA

guidelines diligently. While systematicity was adhered to as best as possible, a single-author review does incur a possible likelihood of unintentional bias and methodological limitations when compared with group reviews. However, the processes described by the author are based on data accumulation with clear links between the knowledge and content of the subject as well as providing evidence for future research. Additionally, this review is not meant to be exhaustive and includes only a cursory evaluation of the issues. The clinical applications discussed are limited to SIJ dysfunction in an endurance running population. As a potential limitation, endurance running was classified as involving runners regularly completing over 5 km in training or competitive situations. Therefore, papers featuring experimental trials involving runners covering distances below this threshold were not included. This was motivated by the very high publication rate that made their inclusion infeasible. Nevertheless, this potential limitation did not alter the key points raised in the large number of papers included in this review and presented in the Discussion section. While an effective SIJ is fundamental in one's ability to run with biomechanical efficiency and effectiveness, this systematic review was not intended to review sensor-based methods solely for applied real-time gait analysis. As gait analysis can include sensors located on the shank and foot, which are most often used in combination with threshold or peak identification methods for gait detection for SIJ assessment, review papers on gait analysis were limited.

Recommendations

Despite these limitations, future studies should prioritize improving the quality of research aimed at reducing discrepancies in result interpretation, increasing reliability and validity, and promoting study generalizability. Given these findings, the review concurs with Block and Miller [69] that SIJ pain and dysfunction in endurance runners are likely highly underdiagnosed and undertreated. Additionally, clinicians should be mindful of a broader range of potential differential diagnoses regarding other sources of posterior hip and LBP in endurance runners.

Based on the findings of this review, wearables combined with smart devices could enable real-time data to be sent to health care professionals and clinicians, allowing for simultaneous tracking of endurance runners and monitoring the magnitude of SIJ dysfunction. This also challenges the engineering community to develop more intelligent, real-time, accurate information, making it user-friendly and offering athletes and clinicians actionable insights based on context-specific evaluation frameworks. As noted by Clermont et al [70], personalized and effective wearable technology should be rooted in a thorough understanding of the user's experience, attitudes, and opinions which, if not properly considered, can severely hamper the potential of applications.

The selected articles, particularly those from 2020 and the onset of the COVID-19 pandemic, undoubtedly reflect the widespread interest in the area and an increasing trend in popularity. The analysis resulted in some key conclusions, which were reported along with main reflection points that led to the formulation of

guidelines and good practices for future research and dissemination. These are as follows:

- Articles should explicitly state the rationale for choosing and analyzing specific biomechanical quantities relating to the SIJ and include a justification of what relationship may exist between the SIJ and the diagnosed dysfunction. When previous literature and reviews are cited to support the choice made, the strength of evidence of previous studies should be discussed, together with the context from which that evidence emerged.
- More effort should be spent to fully exploit the potential of wearable technologies to detect and manage SIJ dysfunction, particularly as part of an injury management plan (post). This would allow the unobtrusive monitoring and quantification of the effects of prescribed interventions (preventive or rehabilitative) more regularly.
- The continuous progress in wearables offers many opportunities to collect data on many athletes simultaneously, unobtrusively, for long periods, and in field-based situations. However, the great “power” that even consumer-level technologies (eg, smartphones,

watches, pods) currently offer does not come without problems, such as those associated with validity, user and clinician experience, and interpretation of data.

Conclusions

A current “state of play” in SIJ dysfunction among endurance runners for rehabilitation considerations using wearables in a remote setting was presented. This study took a systematic review approach to explore the existing literature on SIJ dysfunction in an endurance running population, using wearables as a rehabilitation tool. Viewed through the lens of wearable technology, the results from this review show that diagnosing, treating, and managing SIJ dysfunction in endurance runners vary considerably because of the inconsistent definition of the condition. To identify optimal rehabilitation considerations and effectively monitor this condition using remote wearables, further investigations are recommended to better clarify the condition. Moreover, greater utilization of wearables for measuring both biomechanics and pathomechanics is suggested to enhance the reliability and accuracy of remote wearable usage.

Acknowledgments

The author declares that no financial support was provided for this research. Acknowledgement is given to Dr Jim Lee who assisted with the preparation of this paper.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.

[[DOCX File , 32 KB - biomedeng_v9i1e46067_app1.docx](#)]

Multimedia Appendix 2

Google Scholar, PubMed, and Scopus search strategies.

[[DOCX File , 21 KB - biomedeng_v9i1e46067_app2.docx](#)]

Multimedia Appendix 3

The IrRL classification model, where different classes of research maturity (IrRL1-3; columns) are mapped against the following feature domains (rows): knowledge of causal relationships, experimental settings, testing technology, and normative guidelines. IrRL: Injury-research Readiness Level.

[[DOCX File , 13 KB - biomedeng_v9i1e46067_app3.docx](#)]

References

1. McGuine TA, Post EG, Hetzel SJ, Brooks MA, Trigsted S, Bell DR. A prospective study on the effect of sport specialization on lower extremity injury rates in high school athletes. *Am J Sports Med* 2017 Oct 23;45(12):2706-2712. [doi: [10.1177/0363546517710213](https://doi.org/10.1177/0363546517710213)] [Medline: [28735552](https://pubmed.ncbi.nlm.nih.gov/28735552/)]
2. Emery CA, Pasanen K. Current trends in sport injury prevention. *Best Pract Res Clin Rheumatol* 2019 Feb;33(1):3-15. [doi: [10.1016/j.berh.2019.02.009](https://doi.org/10.1016/j.berh.2019.02.009)] [Medline: [31431273](https://pubmed.ncbi.nlm.nih.gov/31431273/)]
3. Creighton D, Shrier I, Shultz R, Meeuwisse W, Matheson G. Return-to-play in sport: a decision-based model. *Clin J Sport Med* 2010 Sep;20(5):379-385. [doi: [10.1097/JSM.0b013e3181f3c0fe](https://doi.org/10.1097/JSM.0b013e3181f3c0fe)] [Medline: [20818198](https://pubmed.ncbi.nlm.nih.gov/20818198/)]
4. Dydyk AM, Forro SD, Hanna A. Sacroiliac joint injury. In: *StatPearls*. StatPearls Publishing: Treasure Island, FL; 2022.

5. Grayson A, Shawgo R, Johnson A, Flynn N, Li Y, Cima M, et al. A BioMEMS review: MEMS technology for physiologically integrated devices. *Proc IEEE* 2004 Jan;92(1):6-21. [doi: [10.1109/jproc.2003.820534](https://doi.org/10.1109/jproc.2003.820534)]
6. Lee SP, Ha G, Wright DE, Ma Y, Sen-Gupta E, Haubrich NR, et al. Highly flexible, wearable, and disposable cardiac biosensors for remote and ambulatory monitoring. *NPJ Digit Med* 2018;1:2. [doi: [10.1038/s41746-017-0009-x](https://doi.org/10.1038/s41746-017-0009-x)] [Medline: [31304288](https://pubmed.ncbi.nlm.nih.gov/31304288/)]
7. Taelman J, Adriaensen T, van der Horst C, Linz T, Spaepen A. Textile integrated contactless EMG sensing for stress analysis. *Annu Int Conf IEEE Eng Med Biol Soc* 2007;2007:3966-3969. [doi: [10.1109/IEMBS.2007.4353202](https://doi.org/10.1109/IEMBS.2007.4353202)] [Medline: [18002868](https://pubmed.ncbi.nlm.nih.gov/18002868/)]
8. Park JL, Fairweather MM, Donaldson DI. Making the case for mobile cognition: EEG and sports performance. *Neurosci Biobehav Rev* 2015 May;52:117-130 [FREE Full text] [doi: [10.1016/j.neubiorev.2015.02.014](https://doi.org/10.1016/j.neubiorev.2015.02.014)] [Medline: [25735956](https://pubmed.ncbi.nlm.nih.gov/25735956/)]
9. Hilty DM, Armstrong CM, Edwards-Stewart A, Gentry MT, Luxton DD, Krupinski EA. Sensor, wearable, and remote patient monitoring competencies for clinical care and training: scoping review. *J Technol Behav Sci* 2021 Jan 22;6(2):252-277 [FREE Full text] [doi: [10.1007/s41347-020-00190-3](https://doi.org/10.1007/s41347-020-00190-3)] [Medline: [33501372](https://pubmed.ncbi.nlm.nih.gov/33501372/)]
10. Tormene P, Bartolo M, De Nunzio AM, Fecchio F, Quaglini S, Tassorelli C, et al. Estimation of human trunk movements by wearable strain sensors and improvement of sensor's placement on intelligent biomedical clothes. *BioMed Eng OnLine* 2012 Dec 14;11(1):95. [doi: [10.1186/1475-925x-11-95](https://doi.org/10.1186/1475-925x-11-95)]
11. Seshadri DR, Drummond C, Craker J, Rowbottom JR, Voos JE. Wearable devices for sports: new integrated technologies allow coaches, physicians, and trainers to better understand the physical demands of athletes in real time. *IEEE Pulse* 2017 Jan;8(1):38-43. [doi: [10.1109/mpul.2016.2627240](https://doi.org/10.1109/mpul.2016.2627240)]
12. Capecci M, Ceravolo MG, Ferracuti F, Iarlori S, Monteriu A, Romeo L, et al. The KIMORE Dataset: KInematic Assessment of MOvement and Clinical Scores for Remote Monitoring of Physical REhabilitation. *IEEE Trans Neural Syst Rehabil Eng* 2019 Jul;27(7):1436-1448. [doi: [10.1109/tnsre.2019.2923060](https://doi.org/10.1109/tnsre.2019.2923060)]
13. McIntosh AS. Risk compensation, motivation, injuries, and biomechanics in competitive sport. *Br J Sports Med* 2005 Jan 01;39(1):2-3 [FREE Full text] [doi: [10.1136/bjism.2004.016188](https://doi.org/10.1136/bjism.2004.016188)] [Medline: [15618329](https://pubmed.ncbi.nlm.nih.gov/15618329/)]
14. Finch C. A new framework for research leading to sports injury prevention. *J Sci Med Sport* 2006 May;9(1-2):3-9; discussion 10. [doi: [10.1016/j.jsams.2006.02.009](https://doi.org/10.1016/j.jsams.2006.02.009)] [Medline: [16616614](https://pubmed.ncbi.nlm.nih.gov/16616614/)]
15. Bahr R, Alfredson H, Järvinen M, Järvinen T, Khan K, Kjær M, et al. *The IOC Manual of Sports Injuries*. Hoboken, NJ: John Wiley & Sons, Ltd; Jun 2012.
16. Jacobs JM, Cameron KL, Bojeskul JA. Lower extremity stress fractures in the military. *Clin Sports Med* 2014 Oct;33(4):591-613. [doi: [10.1016/j.csm.2014.06.002](https://doi.org/10.1016/j.csm.2014.06.002)] [Medline: [25280611](https://pubmed.ncbi.nlm.nih.gov/25280611/)]
17. Bennett H, Arnold J, Norton K, Davison K. Are we really "screening" movement? The role of assessing movement quality in exercise settings. *J Sport Health Sci* 2020 Dec;9(6):489-492 [FREE Full text] [doi: [10.1016/j.jshs.2020.08.002](https://doi.org/10.1016/j.jshs.2020.08.002)] [Medline: [32791205](https://pubmed.ncbi.nlm.nih.gov/32791205/)]
18. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med* 2021 Mar;18(3):e1003583 [FREE Full text] [doi: [10.1371/journal.pmed.1003583](https://doi.org/10.1371/journal.pmed.1003583)] [Medline: [33780438](https://pubmed.ncbi.nlm.nih.gov/33780438/)]
19. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice (3rd Edition)*. Saddle River, NJ: Pearson/Prentice Hall; 2015.
20. Preatoni E, Bergamini E, Fantozzi S, Giraud LI, Orejel Bustos AS, Vannozzi G, et al. The use of wearable sensors for preventing, assessing, and informing recovery from sport-related musculoskeletal injuries: a systematic scoping review. *Sensors (Basel)* 2022 Apr 22;22(9):3225 [FREE Full text] [doi: [10.3390/s22093225](https://doi.org/10.3390/s22093225)] [Medline: [35590914](https://pubmed.ncbi.nlm.nih.gov/35590914/)]
21. McKinney J, Velghe J, Fee J, Isserow S, Drezner JA. Defining athletes and exercisers. *Am J Cardiol* 2019 Feb 01;123(3):532-535. [doi: [10.1016/j.amjcard.2018.11.001](https://doi.org/10.1016/j.amjcard.2018.11.001)] [Medline: [30503799](https://pubmed.ncbi.nlm.nih.gov/30503799/)]
22. Sauser B, Verma D, Ramirez-Marquez J, Gove R. From TRL to SRL: the concept of systems readiness levels. 2006 Presented at: Conference on Systems Engineering Research; April 7-8, 2006; Los Angeles, CA p. 1-10 URL: https://www.researchgate.net/publication/228652562_From_TRL_to_SRL_The_concept_of_systems_readiness_levels/link/0c96051598e3c0b3b7000000/download?tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uInB1YmxpY2F0aW9uIn19
23. Henderson RM, Clark KB. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* 1990 Mar;35(1):9. [doi: [10.2307/2393549](https://doi.org/10.2307/2393549)]
24. Shen Y, Chen L, Yue W, Xu H. Digital technology-based telemedicine for the COVID-19 pandemic. *Front Med (Lausanne)* 2021 Jul 6;8:646506 [FREE Full text] [doi: [10.3389/fmed.2021.646506](https://doi.org/10.3389/fmed.2021.646506)] [Medline: [34295908](https://pubmed.ncbi.nlm.nih.gov/34295908/)]
25. Channa A, Popescu N, Skibinska J, Burget R. The rise of wearable devices during the COVID-19 pandemic: a systematic review. *Sensors (Basel)* 2021 Aug 28;21(17):5787 [FREE Full text] [doi: [10.3390/s21175787](https://doi.org/10.3390/s21175787)] [Medline: [34502679](https://pubmed.ncbi.nlm.nih.gov/34502679/)]
26. Ueberschär O, Fleckenstein D, Warschun F, Kränzler S, Walter N, Hoppe MW. Measuring biomechanical loads and asymmetries in junior elite long-distance runners through triaxial inertial sensors. *Sports Orthopaedics and Traumatology* 2019 Sep;35(3):296-308. [doi: [10.1016/j.orthtr.2019.06.001](https://doi.org/10.1016/j.orthtr.2019.06.001)]

27. Whitney KE, Sugimoto D, d'Hemecourt CA, d'Hemecourt DA, d'Hemecourt PA. Running gait biomechanics in female runners with sacroiliac joint pain. *J Phys Ther Sci* 2022 Apr;34(4):327-334 [FREE Full text] [doi: [10.1589/jpts.34.327](https://doi.org/10.1589/jpts.34.327)] [Medline: [35400840](https://pubmed.ncbi.nlm.nih.gov/35400840/)]
28. Alcantara R, Day E, Hahn M, Grabowski A. Sacral acceleration can predict whole-body kinetics and stride kinematics across running speeds. *PeerJ* 2021;9:e11199 [FREE Full text] [doi: [10.7717/peerj.11199](https://doi.org/10.7717/peerj.11199)] [Medline: [33954039](https://pubmed.ncbi.nlm.nih.gov/33954039/)]
29. Höfer S, Siemsen C. Proof of the biomechanical effect of a lumbar spine-relief orthosis for treatment of sacroiliac pain. *Z Orthop Unfall* 2008 Aug 14;146(4):439-443. [doi: [10.1055/s-2008-1038612](https://doi.org/10.1055/s-2008-1038612)] [Medline: [18704838](https://pubmed.ncbi.nlm.nih.gov/18704838/)]
30. Amorosa LF, Serota AC, Berman N, Lorich DG, Helfet DL. An isolated iliac wing stress fracture in a marathon runner. *Am J Orthop (Belle Mead NJ)* 2014 Feb;43(2):74-77. [Medline: [24551864](https://pubmed.ncbi.nlm.nih.gov/24551864/)]
31. Liu X, Zhao C, Zheng B, Guo Q, Duan X, Wulamu A, et al. Wearable devices for gait analysis in intelligent healthcare. *Front Comput Sci* 2021 May 13;3:661-676. [doi: [10.3389/fcomp.2021.661676](https://doi.org/10.3389/fcomp.2021.661676)]
32. Banos O, Moral-Munoz J, Diaz-Reyes I, Arroyo-Morales M, Damas M, Herrera-Viedma E, et al. mDurance: a novel mobile health system to support trunk endurance assessment. *Sensors (Basel)* 2015 Jun 05;15(6):13159-13183 [FREE Full text] [doi: [10.3390/s150613159](https://doi.org/10.3390/s150613159)] [Medline: [26057034](https://pubmed.ncbi.nlm.nih.gov/26057034/)]
33. Falowski S, Sayed D, Pope J, Patterson D, Fishman M, Gupta M, et al. A review and algorithm in the diagnosis and treatment of sacroiliac joint pain. *JPR* 2020 Dec;Volume 13:3337-3348. [doi: [10.2147/jpr.s279390](https://doi.org/10.2147/jpr.s279390)]
34. Zadeh A, Taylor D, Bertson M, Tillman T, Nosoudi N, Bruce S. Predicting sports injuries with wearable technology and data analysis. *Inf Syst Front* 2020 May 22;23(4):1023-1037 [FREE Full text] [doi: [10.1007/s10796-020-10018-3](https://doi.org/10.1007/s10796-020-10018-3)]
35. Porciuncula F, Roto AV, Kumar D, Davis I, Roy S, Walsh CJ, et al. Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances. *PM R* 2018 Sep 27;10(9 Suppl 2):S220-S232 [FREE Full text] [doi: [10.1016/j.pmrj.2018.06.013](https://doi.org/10.1016/j.pmrj.2018.06.013)] [Medline: [30269807](https://pubmed.ncbi.nlm.nih.gov/30269807/)]
36. Lorussi F, Lucchese I, Tognetti A, Tognetti N, Carbonaro N. A wearable system for remote monitoring of the treatments of musculoskeletal disorder. New York, NY: IEEE; 2018 Presented at: IEEE International Conference on Smart Computing (SMARTCOMP); June 18-20, 2018; Sicily, Italy p. 18-20. [doi: [10.1109/smartcomp.2018.00030](https://doi.org/10.1109/smartcomp.2018.00030)]
37. Nascimento L, Bonfati L, Freitas M, Mendes Junior JJA, Siqueira HV, Stevan SL. Sensors and systems for physical rehabilitation and health monitoring—a review. *Sensors (Basel)* 2020 Jul 22;20(15):4063 [FREE Full text] [doi: [10.3390/s20154063](https://doi.org/10.3390/s20154063)] [Medline: [32707749](https://pubmed.ncbi.nlm.nih.gov/32707749/)]
38. Rahlf AL, Hoenig T, Stürznickel J, Cremans K, Fohrmann D, Sanchez-Alvarado A, et al. A machine learning approach to identify risk factors for running-related injuries: study protocol for a prospective longitudinal cohort trial. *BMC Sports Sci Med Rehabil* 2022 Apr 26;14(1):75 [FREE Full text] [doi: [10.1186/s13102-022-00426-0](https://doi.org/10.1186/s13102-022-00426-0)] [Medline: [35473813](https://pubmed.ncbi.nlm.nih.gov/35473813/)]
39. Snijders CJ, Vleeming A, Stoeckart R. Transfer of lumbosacral load to iliac bones and legs Part 1: Biomechanics of self-bracing of the sacroiliac joints and its significance for treatment and exercise. *Clin Biomech (Bristol, Avon)* 1993 Nov;8(6):285-294. [doi: [10.1016/0268-0033\(93\)90002-Y](https://doi.org/10.1016/0268-0033(93)90002-Y)] [Medline: [23916048](https://pubmed.ncbi.nlm.nih.gov/23916048/)]
40. Bellamy N, Park W, Rooney PJ. What do we know about the sacroiliac joint? *Semin Arthritis Rheum* 1983 Feb;12(3):282-313. [doi: [10.1016/0049-0172\(83\)90011-2](https://doi.org/10.1016/0049-0172(83)90011-2)] [Medline: [6867741](https://pubmed.ncbi.nlm.nih.gov/6867741/)]
41. Vleeming A, Schuenke MD, Masi AT, Carreiro JE, Danneels L, Willard FH. The sacroiliac joint: an overview of its anatomy, function and potential clinical implications. *Journal of Anatomy* 2012 Sep 19;221(6):537-567 [FREE Full text] [doi: [10.1111/j.1469-7580.2012.01564.x](https://doi.org/10.1111/j.1469-7580.2012.01564.x)] [Medline: [22994881](https://pubmed.ncbi.nlm.nih.gov/22994881/)]
42. Stureson B, Uden A, Vleeming A. A radiostereometric analysis of the movements of the sacroiliac joints in the reciprocal straddle position. *Spine (Phila Pa 1976)* 2000 Jan 15;25(2):214-217. [doi: [10.1097/00007632-200001150-00012](https://doi.org/10.1097/00007632-200001150-00012)] [Medline: [10685486](https://pubmed.ncbi.nlm.nih.gov/10685486/)]
43. Dreyfuss P, Cole AJ, Pauza K. Sacroiliac joint injection techniques. *Phys Med Rehabil Clin North Am* 1995 Nov;6(4):785-813. [doi: [10.1016/s1047-9651\(18\)30434-0](https://doi.org/10.1016/s1047-9651(18)30434-0)]
44. DeRosa CP, Porterfield JA. A physical therapy model for the treatment of low back pain. *Phys Ther* 1992 Apr;72(4):261-9; discussion 270. [doi: [10.1093/ptj/72.4.261](https://doi.org/10.1093/ptj/72.4.261)] [Medline: [1533940](https://pubmed.ncbi.nlm.nih.gov/1533940/)]
45. Rahlf AL, Hoenig T, Stürznickel J, Cremans K, Fohrmann D, Sanchez-Alvarado A, et al. A machine learning approach to identify risk factors for running-related injuries: study protocol for a prospective longitudinal cohort trial. *BMC Sports Sci Med Rehabil* 2022 Apr 26;14(1):75 [FREE Full text] [doi: [10.1186/s13102-022-00426-0](https://doi.org/10.1186/s13102-022-00426-0)] [Medline: [35473813](https://pubmed.ncbi.nlm.nih.gov/35473813/)]
46. Mündermann L, Corazza S, Andriacchi TP. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *J Neuroeng Rehabil* 2006 Mar 15;3(1):6 [FREE Full text] [doi: [10.1186/1743-0003-3-6](https://doi.org/10.1186/1743-0003-3-6)] [Medline: [16539701](https://pubmed.ncbi.nlm.nih.gov/16539701/)]
47. Camomilla V, Bergamini E, Fantozzi S, Vannozzi G. Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors (Basel)* 2018 Mar 15;18(3):873 [FREE Full text] [doi: [10.3390/s18030873](https://doi.org/10.3390/s18030873)] [Medline: [29543747](https://pubmed.ncbi.nlm.nih.gov/29543747/)]
48. Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. *Spine (Phila Pa 1976)* 1995 Jan 01;20(1):31-37. [doi: [10.1097/00007632-199501000-00007](https://doi.org/10.1097/00007632-199501000-00007)] [Medline: [7709277](https://pubmed.ncbi.nlm.nih.gov/7709277/)]
49. March L, Smith EU, Hoy DG, Cross MJ, Sanchez-Riera L, Blyth F, et al. Burden of disability due to musculoskeletal (MSK) disorders. *Best Pract Res Clin Rheumatol* 2014 Jun;28(3):353-366. [doi: [10.1016/j.berh.2014.08.002](https://doi.org/10.1016/j.berh.2014.08.002)] [Medline: [25481420](https://pubmed.ncbi.nlm.nih.gov/25481420/)]

50. Yoshihara H. Sacroiliac joint pain after lumbar/lumbosacral fusion: current knowledge. *Eur Spine J* 2012 Sep 13;21(9):1788-1796 [[FREE Full text](#)] [doi: [10.1007/s00586-012-2350-8](https://doi.org/10.1007/s00586-012-2350-8)] [Medline: [22581257](#)]
51. Vazirian M, Van Dillen L, Bazrgari B. Lumbopelvic rhythm during trunk motion in the sagittal plane: a review of the kinematic measurement methods and characterization approaches. *Phys Ther Rehabil* 2016;3(1):5. [doi: [10.7243/2055-2386-3-5](https://doi.org/10.7243/2055-2386-3-5)] [Medline: [29034099](#)]
52. Dale M, Evans J, Carter K, O'Connell S, Morgan H, Carolan-Rees G. iFuse implant system for treating chronic sacroiliac joint pain: a NICE medical technology guidance. *Appl Health Econ Health Policy* 2020 Jun 27;18(3):363-373. [doi: [10.1007/s40258-019-00539-7](https://doi.org/10.1007/s40258-019-00539-7)] [Medline: [31879828](#)]
53. Adesida Y, Papi E, McGregor AH. Exploring the role of wearable technology in sport kinematics and kinetics: a systematic review. *Sensors (Basel)* 2019 Apr 02;19(7):1597 [[FREE Full text](#)] [doi: [10.3390/s19071597](https://doi.org/10.3390/s19071597)] [Medline: [30987014](#)]
54. Hughes GT, Camomilla V, Vanwanseele B, Harrison AJ, Fong DT, Bradshaw EJ. Novel technology in sports biomechanics: some words of caution. *Sports Biomech* 2024 Apr 26;23(4):393-401. [doi: [10.1080/14763141.2020.1869453](https://doi.org/10.1080/14763141.2020.1869453)] [Medline: [33896368](#)]
55. Van Hooren B, Goudsmit J, Restrepo J, Vos S. Real-time feedback by wearables in running: current approaches, challenges and suggestions for improvements. *J Sports Sci* 2020 Jan 03;38(2):214-230. [doi: [10.1080/02640414.2019.1690960](https://doi.org/10.1080/02640414.2019.1690960)] [Medline: [31795815](#)]
56. Sheerin KR, Reid D, Besier TF. The measurement of tibial acceleration in runners—a review of the factors that can affect tibial acceleration during running and evidence-based guidelines for its use. *Gait Posture* 2019 Jan;67:12-24. [doi: [10.1016/j.gaitpost.2018.09.017](https://doi.org/10.1016/j.gaitpost.2018.09.017)] [Medline: [30248663](#)]
57. Alderink GJ. The sacroiliac joint: review of anatomy, mechanics, and function. *J Orthop Sports Phys Ther* 1991;13(2):71-84. [doi: [10.2519/jospt.1991.13.2.71](https://doi.org/10.2519/jospt.1991.13.2.71)] [Medline: [18796854](#)]
58. Kissling R, Brunner C, Jacob HAC. Mobility of the sacroiliac joint in vitro. *Z Orthop* 1990;128(3):282-288.
59. Wilder DG, Woodworth BB, Frymoyer JW, Pope MH. Vibration and the human spine. *Spine (Phila Pa 1976)* 1982;7(3):243-254. [doi: [10.1097/00007632-198205000-00008](https://doi.org/10.1097/00007632-198205000-00008)] [Medline: [6214030](#)]
60. Feeney DF, Capobianco RA, Montgomery JR, Morreale J, Grabowski AM, Enoka RM. Individuals with sacroiliac joint dysfunction display asymmetrical gait and a depressed synergy between muscles providing sacroiliac joint force closure when walking. *J Electromyogr Kinesiol* 2018 Dec;43:95-103. [doi: [10.1016/j.jelekin.2018.09.009](https://doi.org/10.1016/j.jelekin.2018.09.009)] [Medline: [30267967](#)]
61. Xiaolong LM, Lehang GM, Liping SM, PhD. WYM, Huixiong XMP. Teleultrasound for the COVID-19 pandemic: a statement from China. *Advanced Ultrasound in Diagnosis and Therapy* 2020;4(2):50. [doi: [10.37015/audt.2020.200036](https://doi.org/10.37015/audt.2020.200036)]
62. Al-Azwani I, Aziz HA. Integration of wearable technologies into patient's electronic medical records. *Qual Prim Care* 2016 Mar 24;24(4):155 [[FREE Full text](#)]
63. Yoong NKM, Perring J, Mobbs RJ. Commercial postural devices: a review. *Sensors (Basel)* 2019 Nov 23;19(23):5128 [[FREE Full text](#)] [doi: [10.3390/s19235128](https://doi.org/10.3390/s19235128)] [Medline: [31771130](#)]
64. Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. *J Biomech* 2018 Nov 16;81:1-11 [[FREE Full text](#)] [doi: [10.1016/j.jbiomech.2018.09.009](https://doi.org/10.1016/j.jbiomech.2018.09.009)] [Medline: [30279002](#)]
65. Xiang L, Gu Y, Mei Q, Wang A, Shim V, Fernandez J. Automatic classification of barefoot and shod populations based on the foot metrics and plantar pressure patterns. *Front Bioeng Biotechnol* 2022;10:843204 [[FREE Full text](#)] [doi: [10.3389/fbioe.2022.843204](https://doi.org/10.3389/fbioe.2022.843204)] [Medline: [35402419](#)]
66. Vijayan V, Connolly J, Condell J, McKelvey N, Gardiner P. Review of wearable devices and data collection considerations for connected health. *Sensors (Basel)* 2021 Aug 19;21(16):5589 [[FREE Full text](#)] [doi: [10.3390/s21165589](https://doi.org/10.3390/s21165589)] [Medline: [34451032](#)]
67. Regterschot G, Ribbers G, Bussmann J. Wearable movement sensors for rehabilitation: from technology to clinical practice. *Sensors (Basel)* 2021 Jul 12;21(14):4744 [[FREE Full text](#)] [doi: [10.3390/s21144744](https://doi.org/10.3390/s21144744)] [Medline: [34300484](#)]
68. Lang CE, Barth J, Holleran CL, Konrad JD, Bland MD. Implementation of wearable sensing technology for movement: pushing forward into the routine physical rehabilitation care field. *Sensors (Basel)* 2020 Oct 10;20(20):5744 [[FREE Full text](#)] [doi: [10.3390/s20205744](https://doi.org/10.3390/s20205744)] [Medline: [33050368](#)]
69. Block J, Miller L. Minimally invasive arthrodesis for chronic sacroiliac joint dysfunction using the SIMmetry SI Joint Fusion system. *MDER* 2014 May:125. [doi: [10.2147/mder.s63575](https://doi.org/10.2147/mder.s63575)]
70. Clermont CA, Duffett-Leger L, Hettinga BA, Ferber R. Runners' perspectives on 'smart' wearable technology and its use for preventing injury. *International Journal of Human-Computer Interaction* 2019 Mar 29;36(1):31-40. [doi: [10.1080/10447318.2019.1597575](https://doi.org/10.1080/10447318.2019.1597575)]

Abbreviations

- IMU:** inertial measurement unit
- IrRL:** Injury-research Readiness Level
- LBP:** lower back pain
- LPR:** lumbopelvic rhythm

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

SIJ: sacroiliac joint

Edited by A Mavragani; submitted 28.01.23; peer-reviewed by A Angelucci, Z Lin; comments to author 27.07.23; revised version received 02.10.23; accepted 30.10.23; published 20.05.24.

Please cite as:

Evans S

Sacroiliac Joint Dysfunction in Endurance Runners Using Wearable Technology as a Clinical Monitoring Tool: Systematic Review
JMIR Biomed Eng 2024;9:e46067

URL: <https://biomedeng.jmir.org/2024/1/e46067>

doi: [10.2196/46067](https://doi.org/10.2196/46067)

PMID: [38875697](https://pubmed.ncbi.nlm.nih.gov/38875697/)

©Stuart Evans. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 20.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Investigation of Deepfake Voice Detection Using Speech Pause Patterns: Algorithm Development and Validation

Nikhil Valsan Kulangareth¹, PhD; Jaycee Kaufman¹, MSc; Jessica Oreskovic¹, MASc; Yan Fossat¹, MSc

Klick Labs, Toronto, ON, Canada

Corresponding Author:

Yan Fossat, MSc

Klick Labs

175 Bloor St E #300

3rd floor

Toronto, ON, M4W3R8

Canada

Phone: 1 6472068717

Email: yfossat@klick.com

Abstract

Background: The digital era has witnessed an escalating dependence on digital platforms for news and information, coupled with the advent of “deepfake” technology. Deepfakes, leveraging deep learning models on extensive data sets of voice recordings and images, pose substantial threats to media authenticity, potentially leading to unethical misuse such as impersonation and the dissemination of false information.

Objective: To counteract this challenge, this study aims to introduce the concept of innate biological processes to discern between authentic human voices and cloned voices. We propose that the presence or absence of certain perceptual features, such as pauses in speech, can effectively distinguish between cloned and authentic audio.

Methods: A total of 49 adult participants representing diverse ethnic backgrounds and accents were recruited. Each participant contributed voice samples for the training of up to 3 distinct voice cloning text-to-speech models and 3 control paragraphs. Subsequently, the cloning models generated synthetic versions of the control paragraphs, resulting in a data set consisting of up to 9 cloned audio samples and 3 control samples per participant. We analyzed the speech pauses caused by biological actions such as respiration, swallowing, and cognitive processes. Five audio features corresponding to speech pause profiles were calculated. Differences between authentic and cloned audio for these features were assessed, and 5 classical machine learning algorithms were implemented using these features to create a prediction model. The generalization capability of the optimal model was evaluated through testing on unseen data, incorporating a model-naive generator, a model-naive paragraph, and model-naive participants.

Results: Cloned audio exhibited significantly increased time between pauses ($P < .001$), decreased variation in speech segment length ($P = .003$), increased overall proportion of time speaking ($P = .04$), and decreased rates of micro- and macropauses in speech (both $P = .01$). Five machine learning models were implemented using these features, with the AdaBoost model demonstrating the highest performance, achieving a 5-fold cross-validation balanced accuracy of 0.81 (SD 0.05). Other models included support vector machine (balanced accuracy 0.79, SD 0.03), random forest (balanced accuracy 0.78, SD 0.04), logistic regression, and decision tree (balanced accuracies 0.76, SD 0.10 and 0.72, SD 0.06). When evaluating the optimal AdaBoost model, it achieved an overall test accuracy of 0.79 when predicting unseen data.

Conclusions: The incorporation of perceptual, biological features into machine learning models demonstrates promising results in distinguishing between authentic human voices and cloned audio.

(JMIR Biomed Eng 2024;9:e56245) doi:[10.2196/56245](https://doi.org/10.2196/56245)

KEYWORDS

voice; vocal biomarkers; deepfakes; artificial intelligence; vocal; sound; sounds; speech; audio; deepfake; cloning; text to speech; cloned; deep learning; machine learning; model-naive

Introduction

An increasing number of individuals rely on digital platforms as their primary sources of news and information [1]. People often trust what they consume on the internet without doing any research on the source. There is a technological advancement significantly influencing the production of digital media known as “deepfake.” Deepfake constitutes a synthetic reproduction of media content, both auditory and visual, carefully crafted to closely represent the physical attributes and vocal characteristics of a specific individual. Its use spans many domains, notably in entertainment, where it can be used for the digital replication of actors for special effects or the creation of intricately detailed characters in video games [2].

Deepfakes are generated through the aggregation of substantial data sets, including voice recordings, images, and video segments [3]. This research specifically targets the detection of audio deepfakes, relying solely on voice data for both deepfake development and detection method testing. The voice data sets serve as the foundation for training deep learning models, predominantly deep neural networks, with the primary objective of encoding unique and distinguishable attributes and characteristics found in human voices, like speech patterns and intonation [3]. Following successful model training, it gains the capability to produce replicated voice data by processing input audio or text [3]. While initially trained with substantial data sets, deepfake generation models posttraining can produce new voice clones with minimal audio input, synthesizing voice data to replicate the target voice’s distinctive traits based on learned patterns during the training phase.

This technology is valuable in many domains including voice assistants, voice dubbing for multimedia, professional voiceovers, and the narration of audiobooks [4]. Deepfake content can be generated rapidly once a model is trained, thereby significantly improving efficiency across many industries. Unfortunately, the irresponsible and unethical misuse of deepfakes is prevalent, encompassing impersonation, the dissemination of false information, and violation of privacy [5,6]. Due to the dynamic and rapidly evolving nature of this technology, remaining updated with the ongoing advancements in deepfake detection is challenging [7].

Individuals need a reliable tool to verify that the information they are consuming is authentic. Several outdated deepfake detection machine learning methods have high levels of accuracy, achieving up to 100% accuracy on a data set [8]. However, these accurate predictions are restricted to the level of advancement of the deepfakes that the detection models are trained with [9]. For example, the previously mentioned tool that achieved 100% accuracy was trained and tested on a data set of deepfakes generated in 2019, which are of much lower quality than the level of deepfakes available in 2023 [8]. Furthermore, recent work has shown that out-of-domain voice clone detectors (ie, voice detectors applied outside of the data set in which they were applied) had extremely low performance, obtaining an area under the receiver operator curve (AUC) of 25% [10]. A more robust detection method might involve

searching for the absence of biological features in the cloned voice, rather than the presence of digital features [11].

Activities such as respiration, swallowing, and cognitive processes can influence speech production and the pattern of pauses in authentic speech. Although voice cloning processes may closely mimic human speech production, machines have no requirements for speech breaks and instead rely on training data to indicate where these pauses occur. This may result in subtle but detectable differences in the way pauses are present in authentic versus cloned audio. Indeed, when humans were asked to distinguish between audio deepfakes and authentic voices, one of the primary justifications for a fake audio classification was unnatural pauses in the recordings [10]. Furthermore, when these features were integrated into a classification regime, a moderate accuracy (approximately 85%) was achieved when analyzing deepfakes by perceptual features such as the amplitude of speech and pauses within a recording [12]. However, that study only assessed the use of a single voice cloning software (ElevenLabs) and a small number of cloned voices (9 built-in text-to-speech (TTS) voices and voices cloned from 2 celebrities). Furthermore, the training, validation, and testing sets were not split by participants, so it is assumed that recordings from the same participant are present in both the training and testing data sets.

We posit that the absence of regular human vocal biomarkers, characterized by the pause pattern in a speech segment, will be effective in differentiating cloned audio from authentic audio. For a more comprehensive understanding of model performance on out-of-domain data, we test the proposed methodology in the following ways:

1. On real and cloned audio recordings the model was not exposed to during training, including built-in TTS obtained from the cloning models
2. On a paragraph the model was not exposed to during training
3. On a new cloning software the model was not exposed to during training

Methods

Recruitment

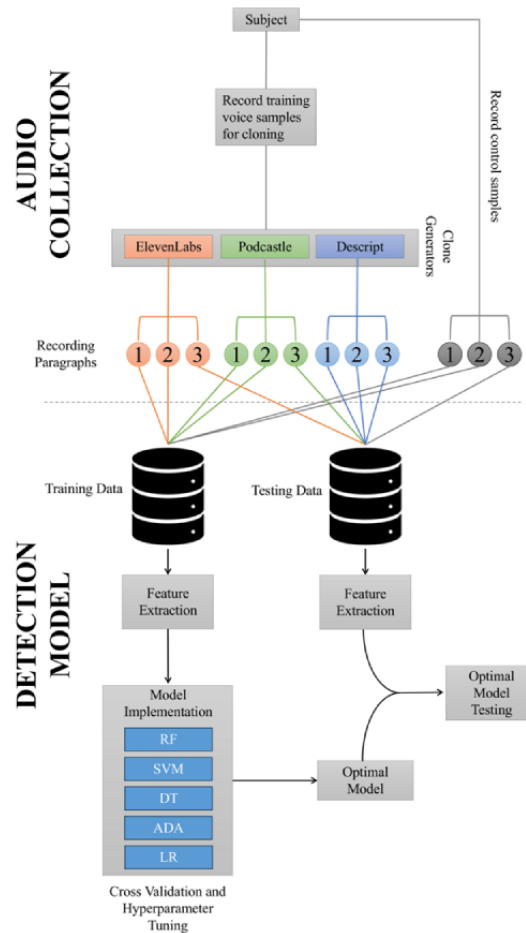
A total of 49 adult participants (20 male) were recruited for this study between June and August 2023 in Toronto, Canada. The participant pool exhibited diversity in terms of ethnicity and had various types and strengths of accents. Exclusion criteria for recruitment included: (1) any person not living in Canada, (2) any person below the age of 18 years, and (3) any speech pathology or condition impeding the production of standard speech, such as stuttering, vocal cord pathology, tracheostomy, or the common cold. No restrictions on gender, ethnicity, accents, or other demographic data were implemented in the recruitment procedure.

The summarized protocol, as illustrated in Figure 1, involves participants recording the required voice samples for the training of 3 distinct deepfake models and a control version of 3 test paragraphs. Subsequently, each deepfake model generates each test paragraph, resulting in a total of 9 deepfake audio samples,

in addition to the 3 control samples for each participant. It is worth noting that some participants were unable to complete the necessary training voice recordings for 1 or 2 of the deepfake

generators due to time constraints, resulting in varying numbers of recordings and deepfakes among participants.

Figure 1. General study protocol overview comprising the audio collection section and detection model development for a participant used in model training. Note that for participants not used in model training (“Model-Naïve Participants”), all data are used for model testing. ADA: AdaBoost; DT: decision tree; LR: logistic regression; RF: random forest; SVM: support vector machine.



Ethics Approval

The research protocol received approval from the Canadian SHIELD Ethics Review Board (REB Tracking Number 2023-06-003).

Audio Samples

In this study, we generated deepfakes using 3 publicly available and user-friendly web-based models: ElevenLabs [13], Podcastle [14], and Describe [15]. Each of these models required different training data. ElevenLabs had the least specific training requirements and was provided approximately 10 minutes of voice recordings, Describe required 10 minutes of speech samples, and Podcastle required participants to read 70 short phrases.

Recordings took place in a quiet room with participants seated in front of a MacBook Pro with 2.8 GHz Quad-Core Intel Core i7. They were instructed to articulate their speech clearly at a standard speaking volume, using the laptop’s built-in microphone to record. The laptop screen displayed the text that participants were required to read for the collection of voice sample data, including the 3 test paragraphs used in the development of the classification model.

All audio samples were saved in the Waveform Audio Format. The respective voice sample data were input for each deepfake generation model for the training process. Upon completion of the model training, a TTS technique was used to generate deepfake versions of the 3 test paragraphs for each model.

Each voice cloning platform also provides pregenerated TTS voices. We generated each of the 3 paragraphs using all available pregenerated TTS to be used in model testing.

Feature Generation

The aim of the analysis was to characterize cloned voices using amplitude-agnostic perceptual voice features, primarily characterized by the pause patterns within a speech segment. Speech segments were identified using a voice activity detector (VAD Solero) in Python [16]. The time between speech segments was calculated and classified as a micropause if the time between segments was greater than or equal to 0.1 seconds and less than 0.5 seconds. It was classified as a macropause if the time between segments was greater than or equal to 0.5 seconds (Figure 2). The recording was trimmed so that the recording began at the beginning of the first speech segment and concluded at the end of the final speech segment. Overall, five features were obtained to denote the pause pattern:

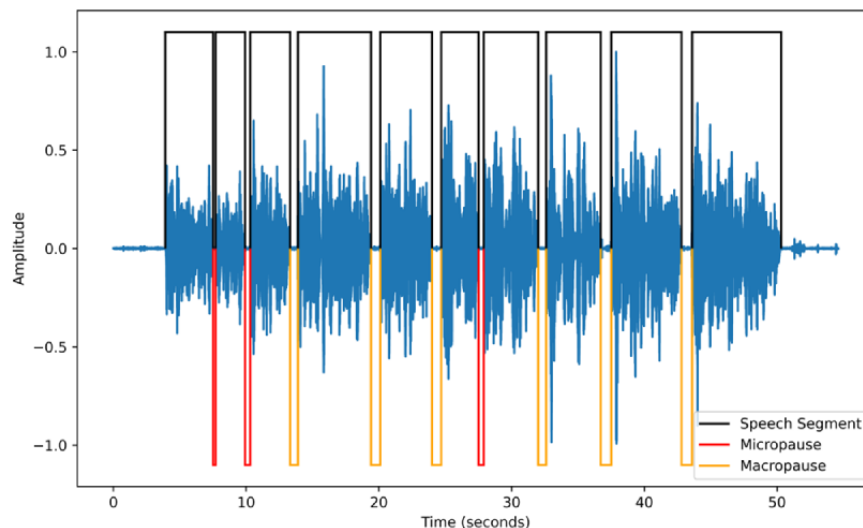
1. SpeechAV: The average speech segment length.
2. SpeechSD: The SD of the speech segment lengths.
3. SpeechProp: The proportion of time speaking, calculated by the sum of all the speech segment lengths divided by the length of the entire recording.
4. MiRate: The rate of micropauses, calculated by dividing the number of micropauses by the length of the trimmed recording (in minutes).
5. MaRate: The rate of macropauses, calculated by dividing the number of macropauses by the length of the trimmed recording (in minutes).

Previous work published by Barrington et al [12] evaluated perceptual features to compare audio deepfakes and authentic voices. In this work, 4 summary metrics to characterize the pauses were generated: the average length of a pause, the SD of the pauses, the pause ratio, and the total number of pauses. We slightly modified and expanded these features to align with

our hypothesis. Rather than the average length and SD of the pauses, we used the average length and SD of the speech segments. We hypothesized that cloned audio would have longer periods between pauses, as they would have no requirements for biological processes such as breathing or swallowing. Furthermore, instead of reporting the number of pauses, which is dependent on the text spoken and the length of the recording, we exclusively reported pause rates. To account for the differences in pause lengths, we calculated the rates of both micropauses and macropauses.

Contrary to the work published by Barrington et al [12], we chose not to include amplitude features. The amplitude of a voice recording can be influenced by the type of microphone used in recording and the distance of the participant to the microphone. Due to this variation, and the desire to evaluate pause metrics exclusively, we chose to remove amplitude-associated features from our feature set.

Figure 2. Sample speech and pause illustration. Black segments indicate speech segments, red segments illustrate micro pauses (pauses < 0.5 seconds and ≥ 0.1 seconds), and yellow segments indicate macro pauses (pauses ≥ 0.5 seconds).



Audio Feature Information

Audio features were compared between authentic and cloned audio. All analysis was conducted in Python. Statistical analysis was conducted using the scipy Python package [17]. *P* values were calculated using the Mann–Whitney U test. Statistical significance is defined as $P < .05$.

Detection Model Generation

An experiment was conducted to assess 5 models to determine the most suitable machine learning tool for this application: random forest (RF), decision tree (DT), logistic regression (LR), support vector machine (SVM), and AdaBoost (ADA) models. Neural networks, although useful in previous deepfake detection methods, perform best with large amounts of training data and tend to overfit with smaller data sets. We aimed to show speech pause patterns could be used to create a robust model even with a small amount of training data, so neural networks were not included in the current analysis.

A 5-fold stratified group cross-validation was used during model training and hyperparameter tuning to find the optimal model. Paragraphs 1 and 2 in [Multimedia Appendix 1](#), and ElevenLabs

and Podcastle generators were used in model training. A total of 30 participants were used in cross-validation (approximately 60% of participants). All recordings corresponding to a participant were kept in the same group, such that if a participant was in one of the folds, all the authentic and cloned recordings obtained from that participant were in the same fold. The total number of recordings used in cross-validation model training is displayed in [Table 1](#).

All analysis was conducted in Python. Models were trained using the scikit-learn Python package [18]. Hyperparameters were tuned using the GridSearch algorithm in scikit-learn, using the parameters denoted in [Multimedia Appendix 2](#). Accuracy is defined as

$$\frac{TP + TN}{TP + FP + FN + TN}$$

Model performance was assessed by the average balanced accuracy of all folds for a model, defined as

$$\frac{BA_1 + BA_2 + \dots + BA_n}{n}$$

where k is the fold number, sensitivity is the accuracy of the model in predicting audio deepfakes, and specificity is the accuracy of the model in predicting authentic audio.

Table 1. Number of recordings collected and generated.

	Training data set (P1 ^a /P2 ^b), n	Testing data set (P1/P2/P3 ^c), n	Total data (P1/P2/P3), n
All recordings	127 (63/64)	257 (63/58/136)	384 (126/122/136)
ElevenLabs			
Pretrained recordings	— ^d	19 (7/5/7)	19 (7/5/7)
Cloned recordings	45 (22/23)	28 (4/0/24)	73 (26/23/24)
Total recordings	45 (22/23)	47 (11/5/31)	92 (33/28/31)
Podcastle			
Pretrained recordings	—	53 (18/18/17)	53 (18/18/17)
Cloned recordings	27 (13/14)	30 (6/4/20)	57 (19/18/20)
Total recordings	27 (13/14)	83 (24/22/37)	110 (37/36/37)
Descript			
Pretrained recordings	—	6 (2/2/2)	6 (2/2/2)
Cloned recordings	—	46 (13/16/17)	46 (13/16/17)
Total recordings	—	52 (15/17/18)	52 (15/17/18)
Authentic			
Total recordings	55 (28/27)	75 (13/13/49)	130 (41/40/49)

^aP1: paragraph 1.

^bP2: paragraph 2.

^cP3: paragraph 3.

^dNot applicable.

Optimal Model Testing

The optimal model from the detection model generation was tested on unseen data. For testing, there were three subgroups of data:

1. Audio recordings from individuals the model was not exposed to during training. This subgroup consists of:
 - Participant audio recordings that were not used in model training (“Model-Naïve Participants”). Note that for a participant to be “Model-Naïve”, neither authentic nor cloned audio obtained from that participant was used in model training.
 - Built-in, pretrained TTS obtained from the cloning models (“Pre-Generated TTS”)
2. A paragraph the model was not exposed to during training (“Model-Naïve Paragraph”; P3, [Multimedia Appendix 1](#)).

3. A new cloning software the model was not exposed to during training (“Model-Naïve Generator”). This was the Descript generator.

The model was tested in such a way that each testing datapoint was Model-Naïve in at least 1 of the 3 above subgroups. Data classes used in model training are denoted as “Model-Trained”.

Results

Audio Feature Information

The 5 audio features corresponding to the speech pause profiles were calculated from the training data and are displayed in [Table 2](#). Overall, cloned audio was significantly associated with increased time between pauses ($P < .001$), decreased variation in the length of speech segments ($P = .003$), increased overall proportion of time speaking ($P = .04$), and a decreased rate of micro- and macropauses in speech (both $P = .01$).

Table 2. Participant and recording data for model features for training data.

Feature	Authentic audio, mean (SD)	Cloned audio, mean (SD)	<i>P</i> values ^a
SpeechAV	2.93 (1.76)	3.49 (1.23)	<.001
SpeechSD	1.51 (1.83)	1.22 (0.89)	.003
SpeechProp	0.87 (0.04)	0.89 (0.04)	.04
MiRate	11.72 (4.34)	9.47 (4.25)	.01
MaRate	7.04 (3.39)	5.78 (2.74)	.01

^a*P* value calculated using Mann-Whitney U test. Statistical significance defined as $P < .05$.

Detection Model Generation

Five classical machine learning algorithms were implemented to create the prediction model, using the 5 features presented in Table 2. A total of 127 recordings were used to train each model and 257 recordings were used to test each model (see Table 1). The optimal performance was obtained by an ADA model, achieving a 5-fold cross-validation balanced accuracy of 0.81 (SD 0.05). The subsequent models were SVM (balanced accuracy 0.79, SD 0.03) and RF (balanced accuracy 0.78, SD 0.04), followed by LR and DT (balanced accuracies 0.76, SD 0.10 and 0.72, SD 0.06). Unsurprisingly, the models that are

traditionally less prone to overfitting (ADA and SVM) were the models that had the best performance, whereas the model that was more likely to overfit (DT) had the poorest performance. Furthermore, ADA and other boosted models can experience the curse of dimensionality when data have many features. By using a small feature set (5 features), we avoided this problem, and ADA achieved a high cross-validated accuracy. Receiver operator curves of all models are shown in Figure 3, and additional model metrics are presented in Table 3. Tuned model hyperparameters are presented in Multimedia Appendix 2.

Figure 3. Average receiver operator curves with variability of all models. The results presented are calculated using the optimal parameter set for each model after Grid Search cross-validation. ADA: AdaBoost; AUC: area under the receiver operator curve; DT: decision tree; LR: logistic regression; RF: random forest; ROC: receiver operator curves; SVM: support vector machine.

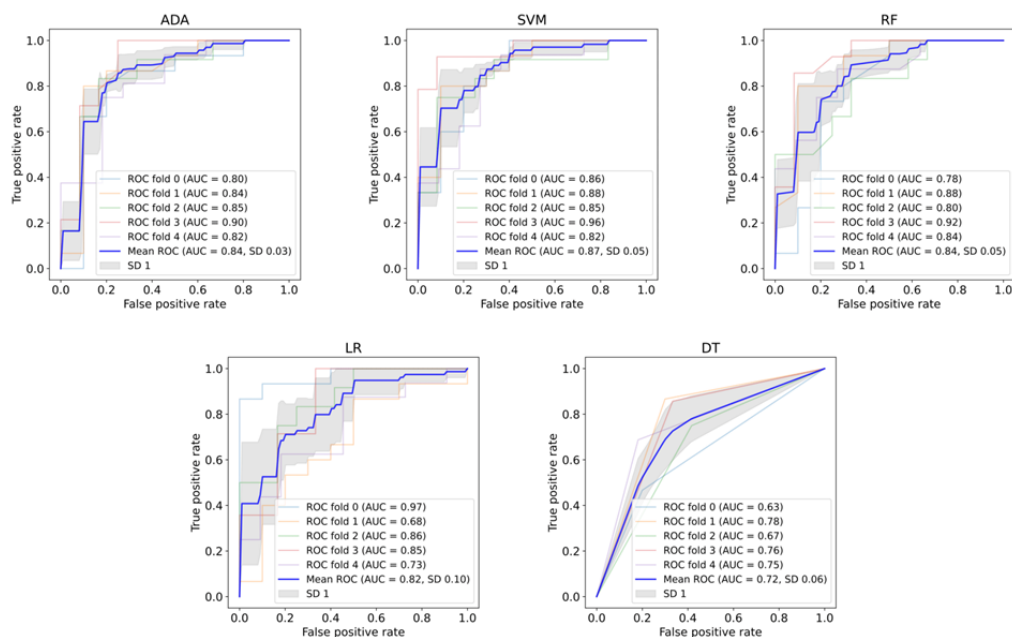


Table 3. Model prediction results for all models.

Model ^a	Balanced accuracy, mean (SD)	Authentic voice accuracy, mean (SD)	Cloned voice accuracy, mean (SD)	Precision, mean (SD)	f1-score, mean (SD)
AdaBoost ^b	0.81 (0.05)	0.75 (0.09)	0.87 (0.08)	0.82 (0.07)	0.84 (0.04)
Support vector machine	0.79 (0.03)	0.73 (0.06)	0.85 (0.05)	0.80 (0.03)	0.82 (0.02)
Random Forest	0.78 (0.04)	0.73 (0.08)	0.83 (0.07)	0.80 (0.07)	0.81 (0.05)
Logistic Regression	0.76 (0.10)	0.70 (0.16)	0.83 (0.09)	0.79 (0.11)	0.81 (0.08)
Decision Tree	0.72 (0.06)	0.71 (0.08)	0.73 (0.15)	0.77 (0.07)	0.73 (0.09)

^aResults presented are calculated using the optimal parameter set for each model after Grid Search cross-validation.

^bOptimal model.

Optimal Model Testing

The optimal ADA model was tested on trained and naïve generators and participants with the paragraphs used in model training (Table 4), and a Model-Naïve paragraph (Table 5). The optimal overall testing performance was obtained when the model was tested on pretrained paragraphs for naïve participants (0.89 overall accuracy). The poorest authentic classification accuracy was obtained when trained participants spoke a new paragraph (accuracy 0.70), potentially indicating the model was overfit to the paragraphs used in training by trained participants. The highest authentic classification accuracy was obtained by model-naïve participants speaking model-trained paragraphs with an accuracy of 0.96. Conversely, the detection of cloned

and pregenerated voices typically performed better on Model-Naïve paragraphs (most accuracies >0.70). The exception to this was the Model-Naïve Generator which had an overall accuracy of 0.67. However, the number of datapoints for this category was extremely small (N=3) so this accuracy may not be the best representation of the Model-Naïve Generator performance. Pregenerated voices with the trained paragraphs had the lowest performance of all the model testing (overall 0.67 accuracy), but classification performance was much higher in the model-naïve paragraph (overall accuracy 0.89). When the results of all confusion matrices in Tables 4 and 5 are compiled, the overall accuracy of all testing data was 0.79 with an AUC of 0.88.

Table 4. Confusion matrices of model test results for model-trained paragraphs (P1 and P2).

	Predicted authentic	Predicted fake	Accuracy
Model-trained participants			
Authentic	— ^a	—	—
Model-trained generator	—	—	—
Model-naïve generator	5	17	0.773
Overall	—	—	0.773
Model-naïve participants			
Authentic	25	1	0.962
Model-trained generator	3	10	0.769
Model-naïve generator	1	7	0.875
Overall	—	—	0.894
Pregenerated TTS^b			
Authentic	—	—	—
Model-trained generator	17	31	0.646
Model-naïve generator	0	4	1.00
Overall	—	—	0.673

^aNot applicable.

^bTTS: text-to-speech.

Table 5. Confusion matrices of model test results for the Model-Naïve paragraph (P3).

	Predicted authentic	Predicted fake	Accuracy
Model-trained participants			
Authentic	19	8	0.704
Model-trained generator	7	29	0.806
Model-naïve generator	0	14	1.00
Overall	— ^a	—	0.805
Model-naïve participants			
Authentic	16	6	0.727
Model-trained generator	1	7	0.875
Model-naïve generator	1	2	0.667
Overall	—	—	0.758
Pregenerated TTS^b			
Authentic	—	—	—
Model-trained generator	3	21	0.875
Model-naïve generator	0	2	1.00
Overall	—	—	0.885

^aNot applicable.

^bTTS: text-to-speech.

Discussion

Principal Findings

This paper outlines the development of an audio deepfake detection model that capitalizes on the distinctive biological vocal characteristics to distinguish between genuine human speech and machine-generated audio. Voice clone samples were created for each participant using 3 publicly available platforms: Descript, ElevenLabs, and Podcastle. To compare these cloned samples with the participants' authentic voice recordings, a variety of perceptual features were calculated to characterize the pause pattern in a recording. The hypothesis was that the speech and pause pattern would be distinguishable between authentic voice recordings and voice clones, as a machine-generated audio sample would not be under the same biological requirements as a human. Machines have no requirements for breathing or swallowing, and their processing time is magnitudes shorter than humans. Even if machines falsely replicate the pauses in speech, their lack of necessity for these processes may create subtle distinctions in the overall pause patterns. Our results support this finding, and 5 perceptual pause features were used to create a detection model for cloned audio.

To generate the voice classification model, 5 machine learning algorithms were used. An ADA model emerged as the most capable of classification, achieving an accuracy of 0.81 (SD 0.05) in 5-fold cross-validation and similar accuracy (0.79) across all testing experiments. The accuracy is in line with previous pause rate detection methods [12], although the testing methodology presented here allows for more comprehensive conclusions about the extendibility of the model results and possible implications for future work. Overall, Model-Naïve

participants, a variety of generators, and Model-Naïve paragraphs were used to test the feasibility of the approach.

In the 5-fold cross-validation model optimization, we achieved an accuracy of 0.75 (SD 0.09) for authentic audio and 0.87 (SD 0.08) for cloned audio. Authentic accuracy may have been lower due to the inherent variation in real human speech, as demonstrated by the higher SDs of the pause metrics in [Table 1](#) compared with cloned audio. This could result in decreased performance, as authentic audio may be more likely to overlap with cloned audio features and thus be harder to classify. Furthermore, we did not prioritize authentic speech accuracy in cross-validation, instead optimizing based on balanced class accuracy. Future models could prioritize authentic audio accuracy in model training and hyperparameter tuning if higher authentic accuracy is preferred.

It is important to note that the text the model was tested on had a distinct effect on the performance of the model. In authentic audio samples, the model performed better on known text for both Model-Trained and Model-Naïve participants. Conversely, in Model-Naïve clones, performance improved when the model was tested on a new paragraph. This effect was evident in both pregenerated TTS and Model-Naïve Participant clones for the Model-Trained generators. This may indicate a tendency for the model to slightly overfit to the paragraphs on which it was trained. When exposed to new participants, its performance declines. That being said, the model accuracy for authentic audio from Model-Naïve participants was 0.73. This is within half an SD of the cross-validated authentic audio accuracy (0.75, SD 0.09), further supporting the use of speech pause metrics for robust model prediction.

Incorporating features associated with real, biological processes (such as breathing, thinking, and swallowing) into a deepfake prediction algorithm is likely to enhance its reliability and longevity in the face of ongoing advancements in deepfake technologies. Instead of solely relying on a model trained on the current state of deepfake generation, which may struggle to maintain accuracy as technology evolves, the inclusion of biological features offers valuable insights that enable the model to adapt and effectively detect inauthentic voices. This approach enhances the model's resilience against evolving deepfake techniques.

Comparison to Prior Work

High-performance current models are typically trained on spectral or deep-learned audio features obtained from the current state of deepfake generation. This permits for an extremely high accuracy in voice clones in a similar domain to the training data but new advancements and subtle changes in these obscure features could soon make these prediction models obsolete. Indeed, when a high-accuracy prediction model was tested on new, out-of-domain voice clones in a recent study, the prediction accuracy was abysmal (AUC is approximately 25%) [10]. We aimed to evaluate the use of perceptual features in current and future model implementations by testing model performance on a completely new generator. Overall, our model performance on a new generator was a success, and the average accuracy of classification of the new generator was 0.87. This generator provided no audio files for model training, and as such, we can conclude that this technique may be extended to out-of-domain cloning processes.

Limitations

This research identified certain limitations in the audio quality variation, linguistic diversity, and deepfake generators used in our study. First, since we created a new cloned audio data set, we only had a small amount of data to train and test the prediction model, and the exclusively English-focused experiments did not account for the potential impact of diverse accents or languages on our results. Small data sets may lead to model overfitting, which we attempted to mitigate using a comprehensive model testing methodology. Further exploration in this domain with a larger and more diverse data set encompassing various accents and languages is warranted, as it has the potential to strengthen the robustness of our conclusions and provide a more comprehensive understanding of model performance across linguistic variations.

Acknowledgments

The authors would like to thank Klick Inc for their support in this research. The authors also thank Anirudh Thommandram for his consultation and insight on the study methodology. The authors used the generative AI tool ChatGPT by OpenAI [20] for general-purpose grammatical editing. All generated text was further reviewed and revised by the study group. No results or conclusions were impacted. The study was internally funded by Klick Inc.

Data Availability

The data sets generated or analyzed during this study are not publicly available due terms in the ethics approval and informed consent. The code used for analysis is available from the corresponding author on reasonable request.

Second, although the pause rate biomarker enhanced prediction accuracy, it introduced the time requirement of sufficiently long audio samples to accurately calculate pause rate data. An older data set that has been widely used for testing and training previous detection tools consisted of samples shorter than 5 seconds, rendering them incompatible with our model [19]. We prioritize the analysis of longer samples due to their higher potential for misuse in the context of misinformation or impersonation scams. Therefore, our detection tool was optimized for modern voice cloning generators and prioritized longer audio outputs over compatibility with previous deepfake data sets.

Third, another limitation concerns the variation of deepfake generation methods. Our study featured 3 distinct tools to introduce variability in deepfake audio samples. Nevertheless, numerous other models exist and possess subtle distinctions that were not covered in our investigation. While we anticipate that the incorporation of vocal biomarkers will enable accurate predictions regardless of the generation method, we did not test deepfakes produced by alternative tools. This decision stemmed from the recognition that there are numerous methods with slight variations in cloned audio samples, compelling us to focus on some of the most prominent and accessible tools.

Future Directions

In this study, we aim to show that speech pause metrics may contribute to robust deepfake detection models, and that trained models using these features perform well on out-of-domain data such as new audio deepfake generators or audio samples from new individuals. Further research should perform an ablation study to compare spectral features and pause pattern features, specifically focusing on testing on unknown data.

Conclusions

In conclusion, the integration of vocal biomarkers into machine learning models shows promise in distinguishing between authentic voice recordings and cloned samples. Given the escalating prevalence of unethical deepfake applications involving impersonation, fraud, and the dissemination of misinformation, establishing a reliable method for verifying source authenticity is crucial. Biological processes and vocal biomarkers offer a potential avenue for enhancing detection methodologies, suggesting a possible means to mitigate the risk of detection tools being rapidly outpaced by advancing deepfake generation technologies.

Authors' Contributions

NVK was responsible for conceptualization, methodology, software, investigation, data curation, and writing of the original draft and its review and editing. JK contributed to methodology, software, validation, formal analysis, investigation, and writing of the original draft and its review and editing. Additionally, JK was involved in visualization. JO participated in the methodology, investigation, and writing of the original draft and its review and editing. JO also contributed to visualization. YF played a role in the conceptualization, methodology, investigation, and writing of the original draft and its review and editing. YF also provided supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Speech paragraphs.

[[DOCX File, 14 KB - biomedeng_v9i1e56245_app1.docx](#)]

Multimedia Appendix 2

Hyperparameter tuning.

[[DOCX File, 14 KB - biomedeng_v9i1e56245_app2.docx](#)]

References

1. Chen Y, Conroy NK, Rubin VL. News in an online world: the need for an “automatic crap detector”. *Proc Assoc Info Sci Tech* 2016;52(1):1-4 [FREE Full text] [doi: [10.1002/pr2.2015.145052010081](#)]
2. Murphy G, Ching D, Twomey J, Linehan C. Face/Off: changing the face of movies with deepfakes. *PLoS One* 2023;18(7):e0287503 [FREE Full text] [doi: [10.1371/journal.pone.0287503](#)] [Medline: [37410765](#)]
3. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. 2014 Presented at: Neural Information Processing Systems; December 8-11, 2014; Montreal, Canada URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Reviews.html
4. Chadha A, Kumar V, Kashyap S, Gupta M. Deepfake: an overview. In: Rodrigues JJPC, Ganzha M, Singh PK, Tanwar S, Wierchoń ST, editors. *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020*. Singapore: Springer; 2021:557-566.
5. Borges L, Martins B, Calado P. Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *J Data Inf Qual* 2019;11(3):1-26 [FREE Full text] [doi: [10.1145/3287763](#)]
6. Vaccari C, Chadwick A. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Soc Media Soc* 2020 Feb 19;6(1):205630512090340 [FREE Full text] [doi: [10.1177/2056305120903408](#)]
7. Engler A. Fighting deepfakes when detection fails. Brookings. 2019. URL: <https://www.brookings.edu/articles/fighting-deep-fakes-when-detection-fails/> [accessed 2024-02-21]
8. Malik H, Chandalvala R. Fighting AI with AI: fake speech detection using deep learning. In: Audio Engineering Society. 2019 Presented at: 2019 AES International Conference on Audio Forensics; June 18-20, 2019; Porto, Portugal URL: <https://www.aes.org/e-lib/browse.cfm?elib=20479> [doi: [10.17743/aesconf.2019.978-1-942220-28-2](#)]
9. Mcuba M, Singh A, Ikuesan RA, Venter H. The effect of deep learning methods on deepfake audio detection for digital investigation. *Procedia Comput Sci* 2023;219:211-219 [FREE Full text] [doi: [10.1016/j.procs.2023.01.283](#)]
10. Mai KT, Bray S, Davies T, Griffin LD. Warning: humans cannot reliably detect speech deepfakes. *PLoS One* 2023;18(8):e0285333 [FREE Full text] [doi: [10.1371/journal.pone.0285333](#)] [Medline: [37531336](#)]
11. Patil K, Kale S, Dhokey J, Gulhane A. Deepfake detection using biological features: a survey. *ArXiv Preprint* posted online January 14, 2023 [FREE Full text]
12. Barrington S, Barua R, Koorma G, Farid H. Single and multi-speaker cloned voice detection: from perceptual to learned features. : IEEE; 2023 Presented at: 2023 IEEE International Workshop on Information Forensics and Security (WIFS); December 4-7, 2023; Nürnberg, Germany p. 1-6. [doi: [10.1109/wifs58808.2023.10374911](#)]
13. Generative voice AI. ElevenLabs. 2023. URL: <https://elevenlabs.io/> [accessed 2024-01-21]
14. Podcasting made easy. Podcastle. 2023. URL: <https://podcastle.ai/> [accessed 2024-02-21]
15. Descript. 2023. URL: <https://www.descript.com/> [accessed 2024-02-21]
16. Silero vad: pre-trained enterprise-grade voice activity detector (vad), number detector and language classifier. GitHub. 2021. URL: <https://github.com/snakers4/silero-vad> [accessed 2024-02-21]
17. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17(3):261-272 [FREE Full text] [doi: [10.1038/s41592-019-0686-2](#)] [Medline: [32015543](#)]

18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-2830 [[FREE Full text](#)]
19. Liu X, Wang X, Sahidullah M, Patino J, Delgado H, Kinnunen T, et al. Asvspoof 2021: towards spoofed and deepfake speech detection in the wild. *IEEE/ACM Trans Audio Speech Lang Process* 2023;31:2507-2522 [[FREE Full text](#)] [doi: [10.1109/taslp.2023.3285283](https://doi.org/10.1109/taslp.2023.3285283)]
20. ChatGPT 3.5. OpenAI. 2024. URL: <https://chat.openai.com/chat> [accessed 2024-02-21]

Abbreviations

ADA: AdaBoost
AUC: area under the receiver operator curve
DT: decision tree
LR: logistic regression
RF: random forest
SVM: support vector machine
TTS: text-to-speech
VAD: voice activity detector

Edited by T Leung; submitted 16.01.24; peer-reviewed by R Iyer, K Mai; comments to author 25.01.24; revised version received 31.01.24; accepted 17.02.24; published 21.03.24.

Please cite as:

Kulangareth NV, Kaufman J, Oreskovic J, Fossat Y

Investigation of Deepfake Voice Detection Using Speech Pause Patterns: Algorithm Development and Validation

JMIR Biomed Eng 2024;9:e56245

URL: <https://biomedeng.jmir.org/2024/1/e56245>

doi: [10.2196/56245](https://doi.org/10.2196/56245)

PMID: [38875685](https://pubmed.ncbi.nlm.nih.gov/38875685/)

©Nikhil Valsan Kulangareth, Jaycee Kaufman, Jessica Oreskovic, Yan Fossat. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 21.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Validation of a Novel Noninvasive Technology to Estimate Blood Oxygen Saturation Using Green Light: Observational Study

Sanjay Gokhale¹, MD; Vinoop Daggubati², MD; Georgios Alexandrakis¹, PhD

¹Department of Biomedical Engineering, The University of Texas at Arlington, Arlington, TX, United States

²Shani Biotechnologies LLC, Austin, TX, United States

Corresponding Author:

Sanjay Gokhale, MD

Department of Biomedical Engineering

The University of Texas at Arlington

500 UTA Blvd

Arlington, TX, 76010

United States

Phone: 1 8172645227

Email: rajhanssanjay@gmail.com

Abstract

Background: Pulse oximeters work within the red-infrared wavelengths. Therefore, these oximeters produce erratic results in dark-skinned subjects and in subjects with cold extremities. Pulse oximetry is routinely performed in patients with fever; however, an elevation in body temperature decreases the affinity of hemoglobin for oxygen, causing a drop in oxygen saturation or oxyhemoglobin concentrations.

Objective: We aimed to determine whether our new investigational device, the Shani device or SH1 (US Patent 11191460), detects a drop in oxygen saturation or a decrease in oxyhemoglobin concentrations.

Methods: An observational study (phase 1) was performed in two separate groups to validate measurements of hemoglobin and oxygen concentrations, including 39 participants recruited among current university students and staff aged 20-40 years. All volunteers completed baseline readings using the SH1 device and the commercially available Food and Drug Administration–approved pulse oximeter Masimo. SH1 uses two light-emitting diodes in which the emitted wavelengths match with absorption peaks of oxyhemoglobin (hemoglobin combined with oxygen) and deoxyhemoglobin (hemoglobin without oxygen or reduced hemoglobin). Total hemoglobin was calculated as the sum of oxyhemoglobin and deoxyhemoglobin. Subsequently, 16 subjects completed the “heat jacket study” and the others completed the “blood donation study.” Masimo was consistently used on the finger for comparison. The melanin level was accounted for using the von Luschan skin color scale (VLS) and a specifically designed algorithm. We here focus on the results of the heat jacket study, in which the subject wore a double-layered heated jacket and pair of trousers including a network of polythene tubules along with an inlet and outlet. Warm water was circulated to increase the body temperature by 0.5-0.8 °C above the baseline body temperature. We expected a slight drop in oxyhemoglobin concentrations in the heating phase at the tissue level.

Results: The mean age of the participants was 24.1 (SD 0.8) years. The skin tone varied from 12 to 36 on the VLS, representing a uniform distribution with one-third of the participants having fair skin, brown skin, and dark skin, respectively. Using a specific algorithm and software, the reflection ratio for oxyhemoglobin was displayed on the screen of the device along with direct hemoglobin values. The SH1 device picked up more minor changes in oxyhemoglobin levels after a change in body temperature compared to the pulse oximeter, with a maximum drop in oxyhemoglobin concentration detected of 6.5% and 2.54%, respectively.

Conclusions: Our new investigational device SH1 measures oxygen saturation at the tissue level by reflectance spectroscopy using green wavelengths. This device fared well regardless of skin color. This device can thus eliminate racial disparity in these key biomarker assessments. Moreover, since the light is shone on the wrist, SH1 can be readily miniaturized into a wearable device.

(*JMIR Biomed Eng* 2024;9:e46974) doi:[10.2196/46974](https://doi.org/10.2196/46974)

KEYWORDS

reflectance spectroscopy; tissue oxygen measurements; oxygen saturation; pulse oximeter; oxyhemoglobin concentration; oxygen level; racial disparity

Introduction

Pulse oximetry is routinely performed in all patients with elevated body temperature. However, high blood temperature decreases the affinity of oxygen for hemoglobin (Hb) [1], and an elevation in temperature by approximately 1 °C decreases arterial oxygen saturation (sO₂) by only 0.5% [2]. Therefore, the decrease is very minimal and is not clinically significant in subjects with a baseline sO₂ level within the normal range [3].

Blood sO₂ measured by pulse oximetry is currently used to monitor tissue hypoxia. This method uses red and infrared wavelengths in the light spectrum and there is no correction for the level of skin melanin. Consequently, the estimated values are often inaccurate in people with darker skin tones due to the overlapping absorption spectra of melanin [4,5]. Administrative and health authorities have also recognized this issue; however, a solution has not yet been put forward [6].

To overcome this limitation, we have developed a novel technology to estimate Hb and tissue oxygenation. The scientific basis, details, and underlying technology of the device are published elsewhere [7]. In brief, the Shani device (SH1) measures reflectance of light from the skin by a pair of light-emitting diodes (LEDs), displayed as the reflectance ratio from LED1 (E1) and LED2 (E2) and as the sum of the ratios (E; E1+E2). The method of measurement is described in our previous report [7]. The light is shone on the wrist and only the reflected light is picked up by the sensors. This analog signal is then converted to a digital format by the processor, which can be analyzed, displayed, and stored in digital form (US Patent 11191460). In this device and associated technology, the reflectance ratio varies inversely with the concentration of Hb or Hb combined with oxygen (OxyHb) [8].

With an increase in body temperature, a slight drop in OxyHb is expected. Here, we focus on the results of the phase 1 study (heat jacket study) to validate the device in healthy human volunteers. The body temperature of the participants was increased by circulating warm water in a double-layered heat jacket; therefore, a slight drop in the OxyHb concentration at

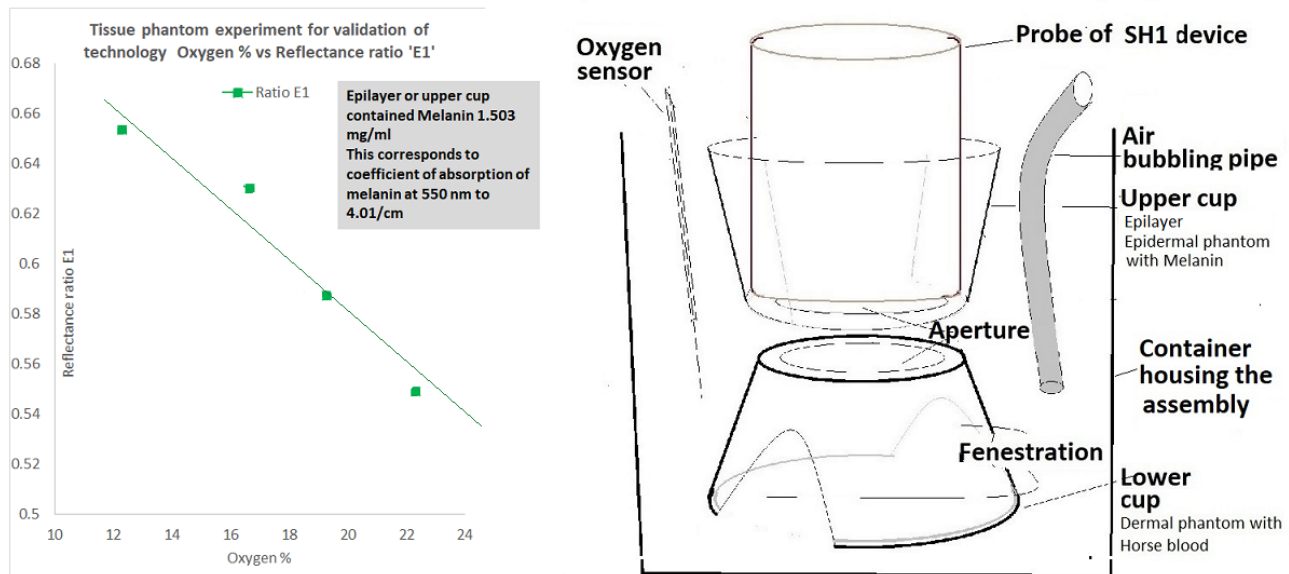
the tissue level was expected during the heating phase. The aim of this study was to determine whether our new device can detect the drop in sO₂ or a decrease in tissue OxyHb concentrations.

Methods

Investigational Device

The SH1 device was validated in tissue phantom experiments. Here, we are presenting the results of the phase 1 heat jacket study in healthy human volunteers. Details of preclinical studies and the results of these experiments have been recently published elsewhere [9]. In the preclinical experiments, we used synthetic melanin as an epilayer mimicking melanin and we used horse blood in the lower layer corresponding to the dermis of the skin. As a part of preclinical studies, we performed an absorption scan of synthetic melanin (Sigma-Aldrich, Instrument-Infinite 200 and BME089922 software system). At a melanin concentration of 1.5 mg/ml, the absorption coefficient of melanin (ie, the absorption of light per unit length) was determined to be 4.01/ml. Even with this high melanin concentration in the epilayer, our device could detect changes in OxyHb levels in the lower dermal layer. Dark-skinned subjects showed an absorption coefficient of melanin at 550 nm, corresponding to 2.5/ml. Therefore, in this preclinical study, we tested melanin levels that are higher (darker) than those measured for dark-skinned subjects. The relationship of melanin concentration with the absorption coefficient of melanin at a wavelength of 550 nm was determined in the previous study [9]. This new investigational device (SH1) measures Hb by shining light from LED1 and LED2 sequentially to obtain measures of both OxyHb and reduced Hb (ie, DeoxyHb). This is termed the “hemoglobin mode” of the operation. The reflectance ratio for LED1 is referred to as E1 and that for LED2 is referred to as E2. The reflectance ratio E1 corresponds to OxyHb, which is inversely related to the oxygen content (Figure 1). The device could detect changes in oxygen concentration in the blood, even in the presence of high melanin in the epilayer, thereby mimicking the detection of hypoxia in dark-skinned subjects.

Figure 1. Results (left) and setup (right) of the tissue phantom experiment [9]. The graph shows the relationship between the reflectance ratio of light-emitting diode 1 (E1) with the measured oxygen saturation level (%). The epilayer or upper cup contained 1.503 mg/ml melanin, corresponding to a coefficient of melanin absorption (at 550 nm) of 4.01/cm. This figure was adapted from Gokhale et al [9] which is published under Creative Commons Attribution 4.0 International License [10].



Eligibility and Recruitment

As per the institutional policy, only current students and members of the staff of University of Texas at Arlington were recruited for this validation study. External candidates, including past students, were not allowed to participate. The age restriction for participation was 20-40 years. For recruitment, flyers were sent by email and posters were displayed in designated locations, including the lobby, near elevators, the cafeteria, and common rooms.

The phase 1 study was performed with two separate cohorts independently for validation of measurements of Hb and oxygen concentrations. There were 39 participants in the two studies; there was only one staff member and the remaining participants were students. Both studies included baseline measurements at visit 1. All 39 volunteers completed baseline readings using our new investigational SH1 device. Subsequently, 16 participants completed the heat jacket study and the others were included in the blood donation study. The commercially available pulse oximeter Masimo was used on the finger for comparison. Skin melanin was accounted for using the von Luschan skin color scale (VLS) and a specifically designed algorithm. Skin tone measurements were performed by two observers independently and the mean value was noted and rounded to the nearest integer. We here focus only on the results of the heat jacket study.

Ethical Considerations

The heat jacket study was approved by the institutional review board (IRB) at The University of Texas at Arlington (STU-2021-0150; approval date February 23, 2021). Written informed consent was obtained from each participant. Copies of consent forms are maintained by IRB authorities. The privacy and confidentiality of the participants were respected and data are stored in a coded, anonymous format. Each participant received monetary compensation as per the stipulated rules and

regulations laid out by the IRB of The University of Texas at Arlington.

Study Design

The heat jacket study included 16 volunteers and was performed under the IRB-approved protocol. At visit 1, baseline measurements of Hb and oxygen concentrations were taken using our novel SH1 device and the commercial pulse oximeter Masimo. At visit 2, after obtaining appropriate consent, the baseline demographic information was obtained. The participant was then asked to wear a double-layered heat jacket and a pair of trousers, which comprise a network of polythene tubules and an inlet and outlet. Subsequently, the participant swallowed a telemetry pill with some water; this is a small pill-shaped electronic object that is used to sense temperature. After a few minutes, the pill reaches the stomach, measures internal body temperature, and emits a signal. A sensor attached to the jacket receives these signals, which are then relayed to a monitor via a cable. Baseline readings were taken with the participant lying down. Warm water was circulated in the jacket and trousers to increase the body temperature by 0.5-0.8 °C above the baseline body temperature as measured with the telemetry pill. This increase in body temperature simulates clinically relevant fever settings. The heating phase lasted for 40 minutes, followed by cooling for the next 20 minutes. Cooling was achieved by circulating cold water through the jacket. Baseline and serial readings were taken with the new investigational SH1 device at 10-minute intervals. The measurements taken with the Masimo Pronto pulse oximeter were used for comparison. A total of 7 sets of observations were obtained for each participant over a period of 60 minutes. Readings in the heating phase and cooling phase were rescaled for each participant with baseline measurements taken as 100%. The percentage drop in sO₂ or the difference between maximum and minimum readings by the SH1 and Masimo devices was plotted for each participant.

Results

The mean age of the study population was 24.1 (SD 0.8) years. The skin tone varied from 12 to 36 on the VLS with a uniform distribution: one-third of the participants had fair skin, brown skin, and dark skin, respectively. Using a specific algorithm accounting for melanin, as determined from the VLS, and associated software, the reflection ratio for OxyHb is displayed on the screen along with direct Hb values [11]. We had baseline readings for all 39 participants (Figure 2) with our SH1 device and the commercial pulse oximeter Masimo. As seen in Figure 2, baseline oxygen concentrations of all 39 participants as measured by SH1 device readings fell within a range similar to those measured by the pulse oximeter.

In the heat jacket study, we expected a slight drop in OxyHb concentrations in the heating phase at the tissue level. Our hypothesis was confirmed after analyzing the data for the reflection ratios E1, E2, and E (E1+E2), followed by computation of OxyHb and total Hb values. Figure 3 shows the distribution of the skin tone (according to the VLS) of the participants compared to the percentage drop in oxygen concentrations in the 16 subjects participating in the heat jacket study. Our device could pick up more minor changes in OxyHb levels after a change in body temperature than possible with the pulse oximeter. The maximum drop in OxyHb concentrations picked up by our device was 6.5% compared to a drop of only 2.54% sensed by the pulse oximeter. The average change in OxyHb measured by our device was 2.98%, whereas that of the pulse oximeter was 1.33%, with a median of 3% and 1%, respectively.

Figure 2. Skin tone versus oxygen saturation measured by the Shani device and Masimo pulse oximeter at baseline in all 39 participants. VLS: von Luschan skin coloration scale.

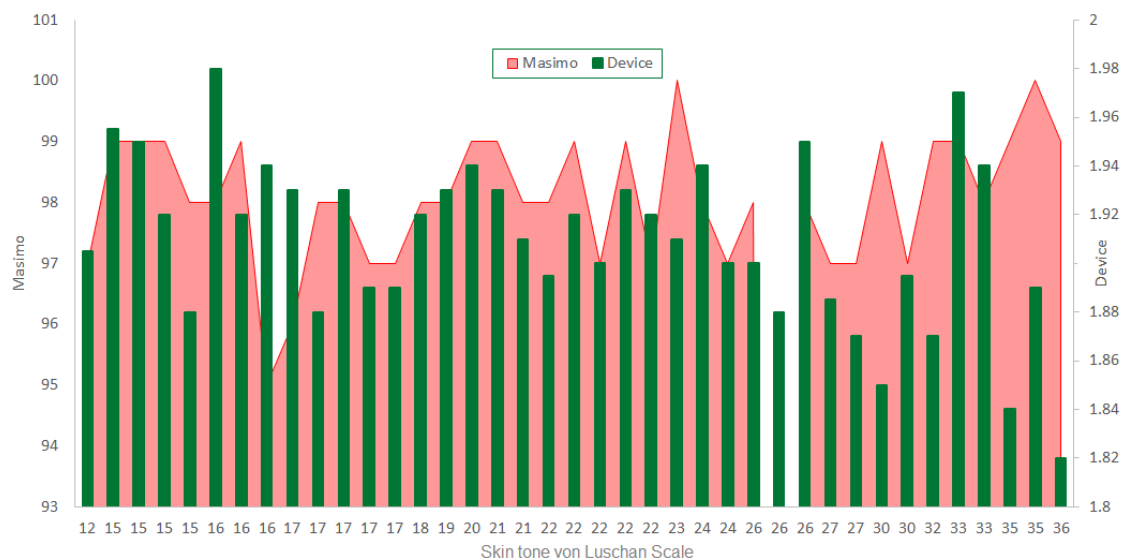
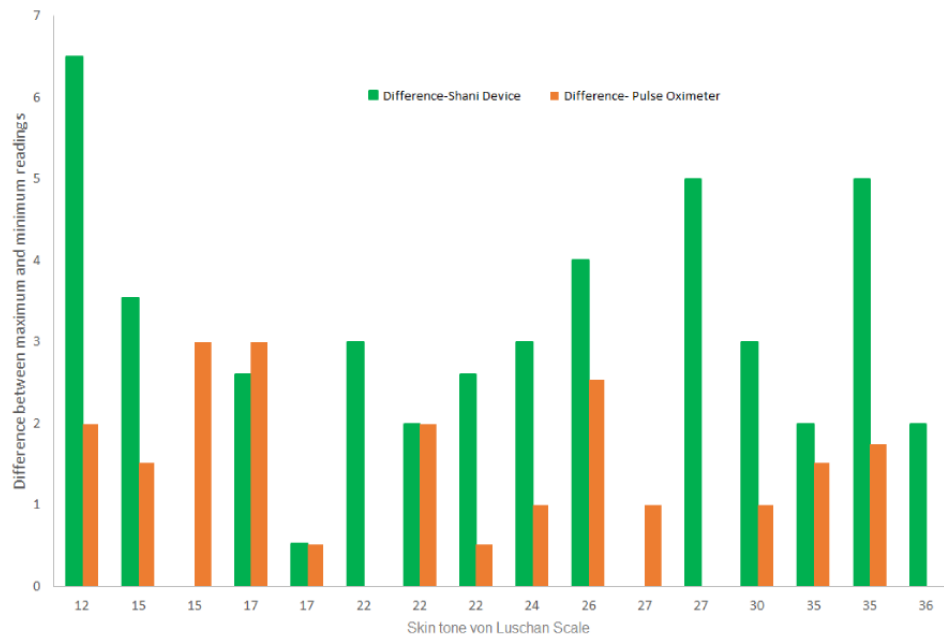


Figure 3. Skin tone versus the percentage drop in oxygen concentrations for the 16 participants in the heat jacket study, as measured by the new device and the commercial pulse oximeter Masimo. VLS: von Luschan skin color scale.



Discussion

Principal Findings

The baseline readings taken by the SH1 device (Figure 2) while the participants were lying in supine position showed some interesting features, including different readings for subjects with the same skin tone (eg, subjects with skin tones 15, 16, 22, and 26 on the VLS) and similar readings for subjects with different skin tones (eg, subjects with skin tones 15, 16, and 26 on the VLS). Therefore, the baseline measurement of tissue oxygen levels (ie, OxyHb levels) as determined by the SH1 device is independent of skin tone and likely depends on tissue metabolism and the basal metabolic rate.

In this heat jacket study, all participants showed a drop in OxyHb levels during the heating phase, followed by a rise in the cooling phase. In some cases, the rise was slightly higher than baseline levels.

We compared the difference in maximum and minimum readings obtained by our SH1 device and the commercial pulse

oximeter Masimo. Figure 3 shows the distribution of the difference across different skin tones. This drop in oxygen saturation was magnified by the SH1 device compared to the readings obtained with the pulse oximeter. The difference between maximum and minimum readings by the SH1 device was 6.5%, whereas that for Masimo was only 2.54%.

As stated earlier, the decrease in arterial oxygen levels with fever is very small, even in patients in the intensive care unit, and this drop in sO_2 is not clinically significant in patients with baseline sO_2 within the normal range. Nonetheless, this minor change is picked up well by our device and is not detected by the pulse oximeter. The commercial pulse oximeter Masimo Pronto measures arterial sO_2 by photoplethysmography and red-infrared wavelengths [12], whereas the novel SH1 device measures sO_2 at the tissue level by reflectance spectroscopy and using two green wavelengths. These are two important fundamental differences between these technologies. Table 1 compares the properties of the SH1 device and pulse oximeter.

Table 1. Comparison of the properties of the new investigational device (Shani device) and a conventional pulse oximeter.

Properties	Pulse oximeter	Shani device
Radiation used	Red and infrared wavelengths	Green visible light
Method	Transmittance of light	Reflectance of light
Any solution for melanin interference?	None so far	Yes; accounted for using a special algorithm
Accuracy of O ₂ % in dark-skinned subjects?	Doubtful; erratic results are obtained with darker skin tones	High accuracy in all subjects, regardless of skin color
Diagnostic ability (eg, hypotension, shock)	Poor accuracy	High accuracy
Continuous monitoring possible	Yes	Yes
Site of testing	Fingertip	Back of wrist
Can be transitioned into a wearable?	Unclear	Certainly
Data storage and transfer	Yes	Yes
Battery operated	Yes	Yes

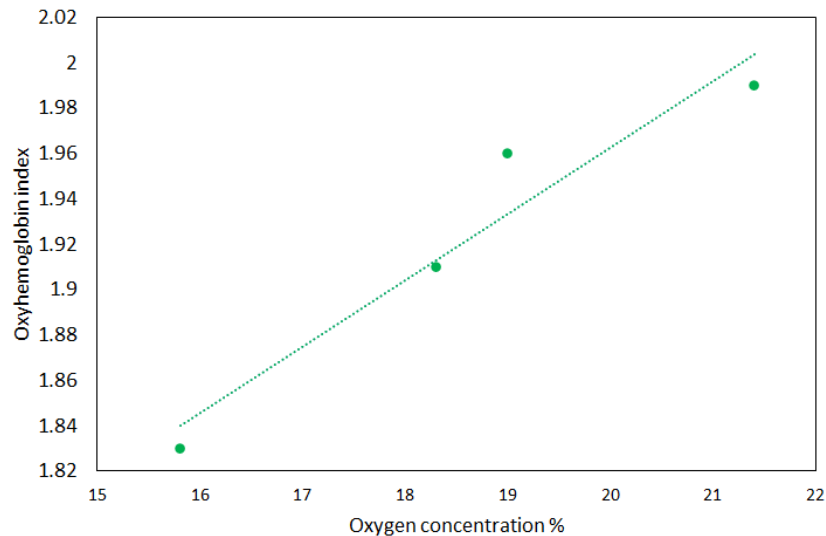
Working of the Device

The emitted wavelengths of the two LEDs matched with the absorption peaks of OxyHb and DeoxyHb (ie, reduced Hb) between 520 nm and 580 nm. As mentioned above, in our new device, the input from LED1 and LED2 of the probe is received as an analog signal. This signal is then converted to digital format by the processor, which can be analyzed, displayed, and stored in digital form. In tissue phantom experiments, we bubbled air and measured the increase in oxygen concentrations in horse blood using an oxygen sensor [9]. The reflection ratio E1 received from LED1 correlates inversely with OxyHb, E2 from LED2 correlates with DeoxyHb, and E (sum of E1 and E2) correlates with total Hb. After analyzing data from the tissue phantom experiments, we derived another new parameter termed the “OxyHb index” as a measure of OxyHb concentration. In our earlier tissue phantom experiments [9], we found that this OxyHb index varies directly with the oxygen concentration in the blood (Figure 4). This means that low oxygen concentrations are reflected as a low OxyHb index and vice versa.

Tissue oxygenation parameters include the concentrations of Hb and OxyHb in the tissue [13]. Tissue sO₂ monitoring is a relatively new technology, and a drop in tissue sO₂ is an early warning sign of peripheral hypoperfusion and the onset of tissue hypoxia [14]. All oximeters currently available using red and infrared wavelengths to target arterial blood flow, which measures sO₂ in the conducting vascular/arterial system. Our device targets the capillary-venous network and measures tissue

oxygenation. Sepsis and shock result in disturbances in microcirculatory perfusion and a change in tissue oxygen utilization that may not be reflected in arterial sO₂ levels. According to many authorities, tissue oxygenation is a better marker of the underlying pathological processes as well as responsiveness to some treatments [15,16]. Tissue oxygen levels are more important, because the arterial O₂ content and an adequate bulk transport of oxygen by the cardiovascular system may not guarantee delivery of oxygen to the critical tissues of the body [17]. Additionally, tissue Hb sO₂ has been determined to be a better predictor of the multiorgan failure outcome [18]. Near-infrared spectroscopy also has well-known limitations [19]. To overcome these limitations, our new device uses green light (520-570 nm) for the estimation of Hb and OxyHb, while accounting for the impact of skin color on the measurements. A summary presentation displays relevant information about this technology in [Multimedia Appendices 1 and 2](#). The video demonstrates the working of this new device.

Using green wavelengths for measurements of Hb and oxygen concentrations with a special algorithm to account for melanin is a novel concept and our efforts have already been appreciated by experts in the field [20,21]. In this study, we have used the VLS for the measurements of skin tone. This is an interval scale with measurements such as 18, 23, and 36. However, we are developing a technology to quantify melanin in the skin more precisely in a noninvasive manner, providing measurements on a continuous scale (eg, 15.25, 17.50, and 30.75). This method is patented and under development.

Figure 4. Oxygen concentrations (%) versus the oxyhemoglobin index.

Strengths and Limitations

The main limitation of this study is the small sample size. Although our device fared better than the pulse oximeter, larger studies are needed in patients with diverse skin tones and a variable degree of tissue hypoxia for further clinical development of the device.

The main strength of this study is that the SH1 device offers an early warning system. A drop in tissue sO_2 is an early warning sign of peripheral hypoperfusion and the onset of tissue hypoxia. Our device can sense this before traditional oximeters can raise the alarm, thereby demonstrating the potential for saving lives.

Our device can also be used in diverse settings, from home monitoring to intensive care, making critical data readily available. Light is shone on the wrist of the individual and is then measured. Therefore, in the future, this device can be miniaturized for wearable technology. The device can be operated in either hemoglobin or oximetry mode. In hemoglobin mode, the device can be used for noninvasive measurements of Hb, whereas in oximetry mode, the device can be used for the continuous monitoring of sO_2 .

Racial equity is another important advantage of our device. Unlike most pulse oximeters on the market, our device works for all skin tones, eliminating bias and improving care for people of color.

The specific technology underlying the design of our device offers specific advantages. First, the use of green light (520-570 nm) can obtain information from the microcirculation, revealing tissue oxygen levels invisible to red and infrared wavelengths of light. Second, our unique algorithm accounts for variations

in melanin, ensuring accurate readings. These advantages can consequently lead to improved outcomes, as early detection and tailored therapies lead to better health for all, especially marginalized groups. Moreover, the ability to obtain a faster diagnosis and intervention can save both resources and lives. Finally, the device offers global reach as it is affordable and adaptable, thereby demonstrating potential to improve health care in resource-limited settings.

Conclusion

The changes in sO_2 at the tissue level in normoxemic subjects are very minor or minimal. Current pulse oximeters, limited to red and infrared wavelengths, only capture the “big picture” of arterial blood flow by measuring oxygen saturation in the conducting arterial system while missing critical changes in the microcirculation, where sepsis and shock wreak havoc, before the arterial oxygen dips.

Our device is a game-changer for measuring tissue oxygenation by shining green light and offering a more sensitive marker of these hidden dangers. Our device was validated to accurately measure tissue oxygen levels and could pick up very minor changes after a change in body temperature in the heat jacket study, demonstrating improved performance compared to the commercial pulse oximeter. Tissue oxygenation parameters include the concentrations of Hb and OxyHb in the tissue. Our device worked better and appeared to be more sensitive than the pulse oximeter even for subjects with light skin or skin tone (eg, VLS 12-18). Since the sample size of this study was small, additional studies with large sample size, a diverse population, and varied degree of hypoxia are required for further confirmation.

Acknowledgments

We thank Dr Michael Nelson, Associate Professor, Department of Kinesiology, University of Texas at Arlington for allowing the use of the Kinesiology Laboratory for the clinical study. The project was funded by Shani Biotechnologies (to GA).

Data Availability

Since this is a patented technology, the data cannot be widely shared. However, the data are available with the institutional review board of the University of Texas at Arlington and can be accessed through Shani Biotechnologies LLC after completing nondisclosure agreement or confidential disclosure agreement formalities.

Authors' Contributions

SGG and GA contributed to the conceptualization, formal analysis, and methodology. VD and GA acquired the funding. GA contributed to project administration and supervision. All authors wrote the original draft and reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary Presentation

[[PPTX File, 5624 KB - biomedeng_v9i1e46974_app1.pptx](#)]

Multimedia Appendix 2

Shani Biotech Demo(10mb)

[[MP4 File \(MP4 Video\), 9402 KB - biomedeng_v9i1e46974_app2.mp4](#)]

References

1. Goldberg S, Heitner S, Mimouni F, Joseph L, Bromiker R, Picard E. The influence of reducing fever on blood oxygen saturation in children. *Eur J Pediatr* 2018 Jan;177(1):95-99. [doi: [10.1007/s00431-017-3037-2](https://doi.org/10.1007/s00431-017-3037-2)] [Medline: [29101451](#)]
2. Kiekkas P, Brokalaki H, Manolis E, Askotiri P, Karga M, Baltopoulos GI. Fever and standard monitoring parameters of ICU patients: a descriptive study. *Intensive Crit Care Nurs* 2007 Oct;23(5):281-288. [doi: [10.1016/j.iccn.2007.04.001](https://doi.org/10.1016/j.iccn.2007.04.001)] [Medline: [17531490](#)]
3. Lahav DZ, Picard E, Mimouni F, Joseph L, Goldberg S. The effect of fever on blood oxygen saturation in children. *Harefuah* 2015 Mar;154(3):162-165, 213, 212. [Medline: [25962244](#)]
4. Fawzy A, Wu TD, Wang K, Robinson ML, Farha J, Bradke A, et al. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. *JAMA Intern Med* 2022 Jul 01;182(7):730-738 [FREE Full text] [doi: [10.1001/jamainternmed.2022.1906](https://doi.org/10.1001/jamainternmed.2022.1906)] [Medline: [35639368](#)]
5. Zonios G, Bykowski J, Kollias N. Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo using diffuse reflectance spectroscopy. *J Invest Dermatol* 2001 Dec;117(6):1452-1457 [FREE Full text] [doi: [10.1046/j.0022-202x.2001.01577.x](https://doi.org/10.1046/j.0022-202x.2001.01577.x)] [Medline: [11886508](#)]
6. Pulse oximeter accuracy and limitations: FDA Safety Communication. US Food and Drug Administration. URL: <https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication> [accessed 2024-03-11]
7. Gokhale SG, Daggubati V, Alexandrakis G. Innovative technology to eliminate the racial bias in non-invasive, point-of-care (POC) haemoglobin and pulse oximetry measurements. *BMJ Innov* 2022 Sep 26;9(2):73-77. [doi: [10.1136/bmjinnov-2022-001018](https://doi.org/10.1136/bmjinnov-2022-001018)]
8. Jacques SL. Quick analysis of optical spectra to quantify epidermal melanin and papillary dermal blood content of skin. *J Biophotonics* 2015 Apr;8(4):309-316. [doi: [10.1002/jbio.201400103](https://doi.org/10.1002/jbio.201400103)] [Medline: [25491716](#)]
9. Gokhale SG, Daggubati VS, Alexandrakis G. Developing a novel device based on a new technology for non-invasive measurement of blood biomarkers irrespective of skin color. *Ger Med Sci* 2023;21:Doc09 [FREE Full text] [doi: [10.3205/000323](https://doi.org/10.3205/000323)] [Medline: [37426887](#)]
10. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: <https://creativecommons.org/licenses/by/4.0/> [accessed 2024-03-22]
11. Swiatoniowski AK, Quillen EE, Shriver MD, Jablonski NG. Technical note: comparing von Luschan skin color tiles and modern spectrophotometry for measuring human skin pigmentation. *Am J Phys Anthropol* 2013 Jun;151(2):325-330. [doi: [10.1002/ajpa.22274](https://doi.org/10.1002/ajpa.22274)] [Medline: [23633083](#)]
12. Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. *Med Devices* 2014;7:231-239 [FREE Full text] [doi: [10.2147/MDER.S47319](https://doi.org/10.2147/MDER.S47319)] [Medline: [25031547](#)]
13. Teng Y, Ding H, Huang L, Li Y, Shan Q, Ye D, et al. Non-invasive measurement and validation of tissue oxygen saturation covered with overlying tissues. *Prog Nat Sci* 2008 Sep;18(9):1083-1088. [doi: [10.1016/j.pnsc.2008.01.035](https://doi.org/10.1016/j.pnsc.2008.01.035)]
14. Epstein CD, Hagenbeck KT. Bedside assessment of tissue oxygen saturation monitoring in critically ill adults: an integrative review of the literature. *Crit Care Res Pract* 2014;2014:709683. [doi: [10.1155/2014/709683](https://doi.org/10.1155/2014/709683)] [Medline: [24900919](#)]

15. Jones N, Terblanche M. Tissue saturation measurement--exciting prospects, but standardisation and reference data still needed. *Crit Care* 2010;14(3):169 [FREE Full text] [doi: [10.1186/cc8970](https://doi.org/10.1186/cc8970)] [Medline: [20619003](https://pubmed.ncbi.nlm.nih.gov/20619003/)]
16. Swartz HM, Flood AB, Schaner PE, Halpern H, Williams BB, Pogue BW, et al. How best to interpret measures of levels of oxygen in tissues to make them effective clinical tools for care of patients with cancer and other oxygen-dependent pathologies. *Physiol Rep* 2020 Aug;8(15):e14541 [FREE Full text] [doi: [10.14814/phy2.14541](https://doi.org/10.14814/phy2.14541)] [Medline: [32786045](https://pubmed.ncbi.nlm.nih.gov/32786045/)]
17. Dantzker DR. Monitoring tissue oxygenation. The search for the grail. *Chest* 1997 Jan;111(1):12-14. [doi: [10.1378/chest.111.1.12](https://doi.org/10.1378/chest.111.1.12)] [Medline: [8995985](https://pubmed.ncbi.nlm.nih.gov/8995985/)]
18. Santora RJ, Moore FA. Monitoring trauma and intensive care unit resuscitation with tissue hemoglobin oxygen saturation. *Crit Care* 2009;13(Suppl 5):S10 [FREE Full text] [doi: [10.1186/cc8008](https://doi.org/10.1186/cc8008)] [Medline: [19951382](https://pubmed.ncbi.nlm.nih.gov/19951382/)]
19. Scheeren TWL, Schober P, Schwarte LA. Monitoring tissue oxygenation by near infrared spectroscopy (NIRS): background and current applications. *J Clin Monit Comput* 2012 Aug;26(4):279-287 [FREE Full text] [doi: [10.1007/s10877-012-9348-y](https://doi.org/10.1007/s10877-012-9348-y)] [Medline: [22467064](https://pubmed.ncbi.nlm.nih.gov/22467064/)]
20. Yang B, Moss J. Evolution of the Pulse Ox. *Chest* 2023 Jul;164(1):24-26 [FREE Full text] [doi: [10.1016/j.chest.2022.12.042](https://doi.org/10.1016/j.chest.2022.12.042)] [Medline: [37423695](https://pubmed.ncbi.nlm.nih.gov/37423695/)]
21. Updating pulse oximeters. National Heart, Lung and Blood Institute. 2023 Jul. URL: <https://www.nhlbi.nih.gov/news/2023/updating-pulse-oximeters> [accessed 2024-03-11]

Abbreviations

DeoxyHb: deoxyhemoglobin
Hb: hemoglobin
IRB: institutional review board
LED: light-emitting diode
OxyHb: oxyhemoglobin
SH1: Shani device
sO₂: oxygen saturation
VLS: von Luschan skin color scale

Edited by T Leung; submitted 03.03.23; peer-reviewed by J Spigulis, MS Arefin; comments to author 11.07.23; revised version received 21.07.23; accepted 29.02.24; published 27.03.24.

Please cite as:

Gokhale S, Daggubati V, Alexandrakis G

Validation of a Novel Noninvasive Technology to Estimate Blood Oxygen Saturation Using Green Light: Observational Study

JMIR Biomed Eng 2024;9:e46974

URL: <https://biomedeng.jmir.org/2024/1/e46974>

doi: [10.2196/46974](https://doi.org/10.2196/46974)

PMID: [38875701](https://pubmed.ncbi.nlm.nih.gov/38875701/)

©Sanjay Gokhale, Vinoop Daggubati, Georgios Alexandrakis. Originally published in *JMIR Biomedical Engineering* (<http://biomsedeng.jmir.org>), 27.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Biomedical Engineering*, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Preliminary Assessment of an Ambulatory Device Dedicated to Upper Airway Muscle Training in Patients With Sleep Apnea: Proof-of-Concept Study

Patrice Roberge¹, PhD; Jean Ruel¹, PhD; André Bégin-Drolet¹, PhD; Jean Lemay¹, PhD; Simon Gakwaya²; Jean-François Masse², DMD; Frédéric Sériès², MD

¹Mechanical Engineering Department, Université Laval, Quebec City, QC, Canada

²Centre de recherche, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Quebec City, QC, Canada

Corresponding Author:

Jean Ruel, PhD

Mechanical Engineering Department

Université Laval

1065 avenue de la Médecine

Quebec City, QC, G1V 0A6

Canada

Phone: 1 418 656 2131 ext 412245

Email: Jean.Ruel@gmc.ulaval.ca

Abstract

Background: Obstructive sleep apnea/hypopnea syndrome (OSAHS) is a prevalent condition affecting a substantial portion of the global population, with its prevalence increasing over the past 2 decades. OSAHS is characterized by recurrent upper airway (UA) closure during sleep, leading to significant impacts on quality of life and heightened cardiovascular and metabolic morbidity. Despite continuous positive airway pressure (CPAP) being the gold standard treatment, patient adherence remains suboptimal due to various factors, such as discomfort, side effects, and treatment unacceptability.

Objective: Considering the challenges associated with CPAP adherence, an alternative approach targeting the UA muscles through myofunctional therapy was explored. This noninvasive intervention involves exercises of the lips, tongue, or both to improve oropharyngeal functions and mitigate the severity of OSAHS. With the goal of developing a portable device for home-based myofunctional therapy with continuous monitoring of exercise performance and adherence, the primary outcome of this study was the degree of completion and adherence to a 4-week training session.

Methods: This proof-of-concept study focused on a portable device that was designed to facilitate tongue and lip myofunctional therapy and enable precise monitoring of exercise performance and adherence. A clinical study was conducted to assess the effectiveness of this program in improving sleep-disordered breathing. Participants were instructed to perform tongue protrusion, lip pressure, and controlled breathing as part of various tasks 6 times a week for 4 weeks, with each session lasting approximately 35 minutes.

Results: Ten participants were enrolled in the study (n=8 male; mean age 48, SD 22 years; mean BMI 29.3, SD 3.5 kg/m²; mean apnea-hypopnea index [AHI] 20.7, SD 17.8/hour). Among the 8 participants who completed the 4-week program, the overall compliance rate was 91% (175/192 sessions). For the tongue exercise, the success rate increased from 66% (211/320 exercises; SD 18%) on the first day to 85% (272/320 exercises; SD 17%) on the last day ($P=.05$). AHI did not change significantly after completion of training but a noteworthy correlation between successful lip exercise improvement and AHI reduction in the supine position was observed ($R_s=-0.76$; $P=.03$). These findings demonstrate the potential of the device for accurately monitoring participants' performance in lip and tongue pressure exercises during myofunctional therapy. The diversity of the training program (it mixed exercises mixed training games), its ability to provide direct feedback for each exercise to the participants, and the easy measurement of treatment adherence are major strengths of our training program.

Conclusions: The study's portable device for home-based myofunctional therapy shows promise as a noninvasive alternative for reducing the severity of OSAHS, with a notable correlation between successful lip exercise improvement and AHI reduction, warranting further development and investigation.

KEYWORDS

obstructive sleep apnea/hypopnea syndrome; OSAHS; myofunctional therapy; myotherapy; oral; orofacial; myology; musculature; labial; buccal; lingual; speech therapy; physiotherapy; physical therapy; oropharyngeal exercises; oropharyngeal; pharyngeal; pharynx; hypopnea; lip; home-based; portable device; devices; ambulatory; portable; monitoring; apnea; mouth; lips; tongue; facial; exercise; exercises; myofunctional; continuous monitoring; sleep-disordered breathing; sleep; breathing; tongue exercise; lip exercise; mHealth; muscle; muscles; muscular; airway; sleep apnea

Introduction

Obstructive sleep apnea/hypopnea syndrome (OSAHS) is a common condition that affects a large portion of the world population [1]. It is estimated that mild to severe OSAHS affects 24% of men and 9% of women in North America [2], with an increase in prevalence over the last 2 decades [3]. OSAHS originates from repetitive closure of the upper airway (UA). The negative impacts of OSAHS include a deterioration of quality of life [4] and an increase in cardiovascular and metabolic morbidity [5-10]. Currently, the gold standard for treatment of this condition is continuous positive airway pressure (CPAP) [5] machines, which provide constant pressure to the sleeping patient via an oral or nasal mask. While this method has proven to be effective in reducing the adverse effects of OSAHS, it has been reported that from 46% to 83% of patients do not adhere to the treatment [11]. The causes of this low adherence rate may include treatment unacceptability, general discomfort, side effects (mask leaks, pressure intolerance, skin irritation, mouth dryness), bed partner intolerance, or a combination of these causes [12].

Alternatively, therapies targeting the UA muscles have been developed to decrease the disease severity [13-15]. Myofunctional therapy is a noninvasive approach in which patients are tasked with exercises of the lips, tongue, or both to target oropharyngeal functions [16]. It has been observed that myofunctional therapy may decrease the apnea-hypopnea index (AHI) by 50% in adults and by 62% in children [16]. For myofunctional therapy to be effective, the patient must perform the exercises daily. However, monitoring the quality and frequency of the exercises is pivotal to supporting implantation of such treatment and may be challenging outside of the laboratory setting. Therefore, there is a need for developing a home reeducation setup where patients can perform daily exercises with continuous monitoring of program adherence and exercise performance.

We developed a portable device that allows completion of tongue and lip myofunctional therapy while providing precise performance monitoring of performance and adherence to exercise. The aims of this clinical study were to evaluate task performance and treatment adherence to a 4-week training session and its efficacy in improving sleep-disordered breathing.

Methods

Study Design

Ten patients with untreated mild or moderate OSAHS who were referred to our sleep clinic volunteered to participate in this

study. These patients were men and postmenopausal women aged ≤ 65 years who had a BMI ≤ 30 kg/m² and regular sleep habits free of sleep debt (caused by, eg, insomnia or sleep deprivation). Their initial OSAHS diagnosis and severity were established by conventional sleep studies (level 1 or 3) performed at our local sleep clinic. Consecutive patients fulfilling the entry criteria were offered enrollment, and recruitment was completed within 6 months. A polysomnographic study (level 2, Embla Titanium; Natus) and the Epworth Sleepiness Scale (ESS) were completed just before and after a 4-week training program. A registered sleep technician who was blind to the protocol performed polysomnography scoring according to standard American Academy of Sleep Medicine criteria [17]. The participants were asked to perform the full training (35 minutes) 6 days a week for 4 weeks. The first session was completed in our research laboratory and the remainder were done at the participant's home. A follow-up was completed by phone on the first and third home training days during the first week and once a week thereafter.

Ethical Considerations

The Ethics Review Board of Institut Universitaire de Cardiologie et de Pneumologie de Québec approved the protocol (2020-3246), which conformed to the guidelines set forth by the Declaration of Helsinki, and written informed consent was obtained from all participants.

Module Overview

A briefcase-sized module that can precisely monitor lip and tongue pressure with a custom-made mouthpiece is presented in this paper. The module presented in [Figure 1](#) is composed of 3 main parts: the mouthpiece, the interface electronics, and the user interface. The mouthpiece design is based on a 3D scan of the patient's teeth and has 2 separate embedded internal cavities to record lip and tongue pressure. The mouthpiece is made out of silicone cast in 3D-printed sugar molds. The pressure developed by contraction of the lips or tongue is read from inside the mouthpiece with transducers and transferred to a computing unit. The module includes a touch screen with an intuitive interface for user interaction. The software combines precise pressure measurements with user calibration and engaging games to maximize therapy adherence. The training presented in the software is based on the tongue-protrusion task program presented by Svensson et al [18], in which the participants are asked to exert a force with their tongue on a force transducer and maintain a certain level of force for 1.5 seconds before releasing. The module is packaged in a customized briefcase for a robust and easy-to-transport solution.

Figure 1. Pictures of the module packaged in a customized briefcase with a touch screen, computing unit, and hardware.



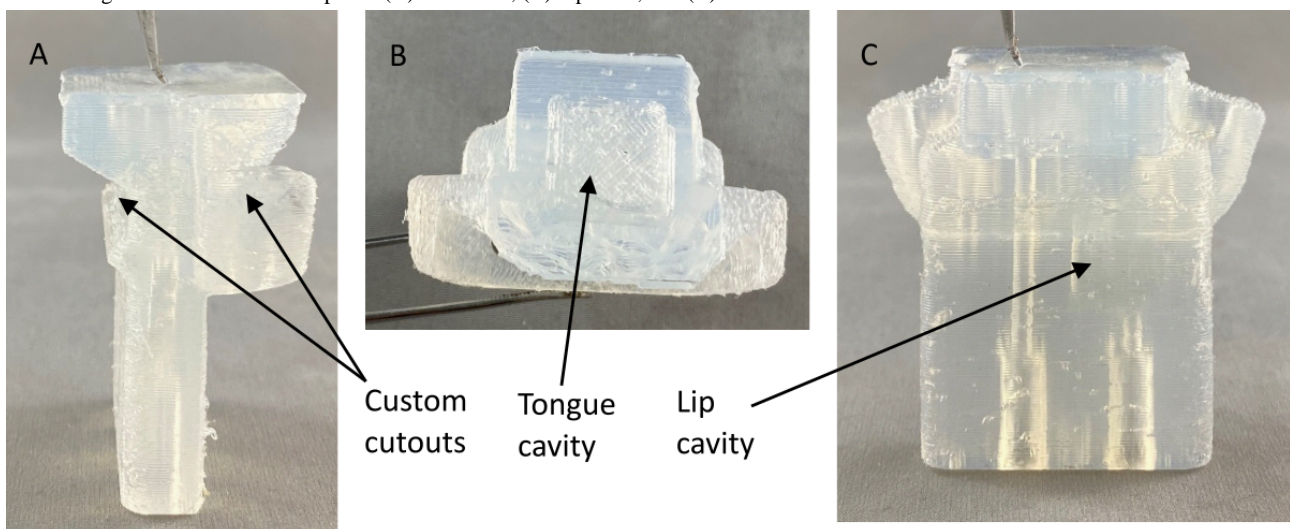
Mouthpiece Design

To maximize adherence to the treatment, a custom mouthpiece was made for each participant. A 3D scan was made of a participant's teeth by a dentist, and the dental print was digitally removed from a mouthpiece template using Meshmixer (version 3.5.474; Autodesk). Having a custom cutout allowed the mouthpiece to clamp naturally onto the patient's teeth and gums, as presented in Figure 2A. The mouthpiece design includes 2 distinct internal cavities acting as pressure chambers, one for the lips and one for the tongue, as presented in Figures 2B and 2C. The lip cavity is located in the front part of the mouthpiece where the lips naturally rest. The cavity has a thin bottom and top wall where the pressure from the lips is applied. The second

cavity is located in the back portion of the mouthpiece (behind the incisors) and has a thin back wall where the tip of the tongue is positioned during tongue exercises. Both cavities have tunnels connecting them to the front of the mouthpiece, where connectors can be installed to 2 distinct pressure transducers. The changes in cavity volume produced by thin wall deformation from lip or tongue movements increase the respective inner pressure.

From the digital model of the mouthpiece, a mold was created with sugar using a custom 3D printer and molding technique [19-21]. Silicon was poured into the mold, and air bubbles were removed in a vacuum chamber. Once the silicon solution solidified, the sugar was dissolved in water to free the mouthpiece.

Figure 2. Images of the custom mouthpiece. (A) Side view; (B) top view; and (C) front view.

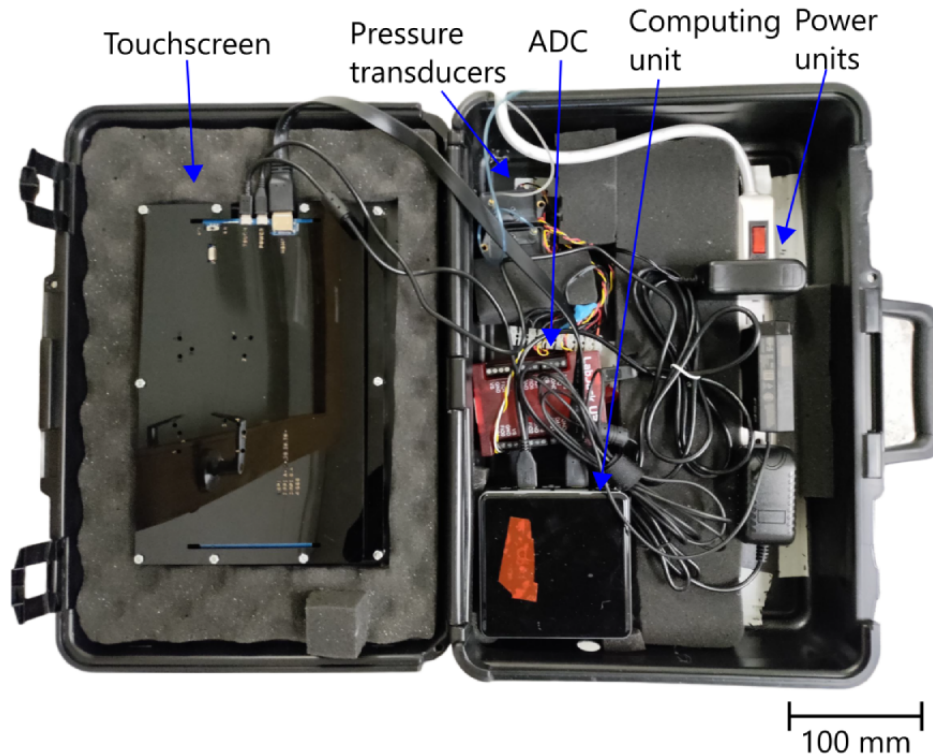


Hardware Description

The absolute pressure in the lip and tongue cavities is linked to 2 pressure transducers (Omega PX142-002D5V) by EVA tubing and adaptors. The output voltage of the transducers is then adapted to maximize the operating range of the analog-to-digital

converter (Labjack U3-LV). A computing unit (NUC7i7BNH) is used to read the values from the 2 converters. The user interacts with the software on the computing unit with a touch screen (Waveshare; this unit uses a 10.1-inch HDMI-connected LCD). All the described hardware, as well as power units, are packaged inside a briefcase, as shown in Figure 3.

Figure 3. Picture of the different hardware components packaged inside the briefcase. ADC: analog-to-digital converter.

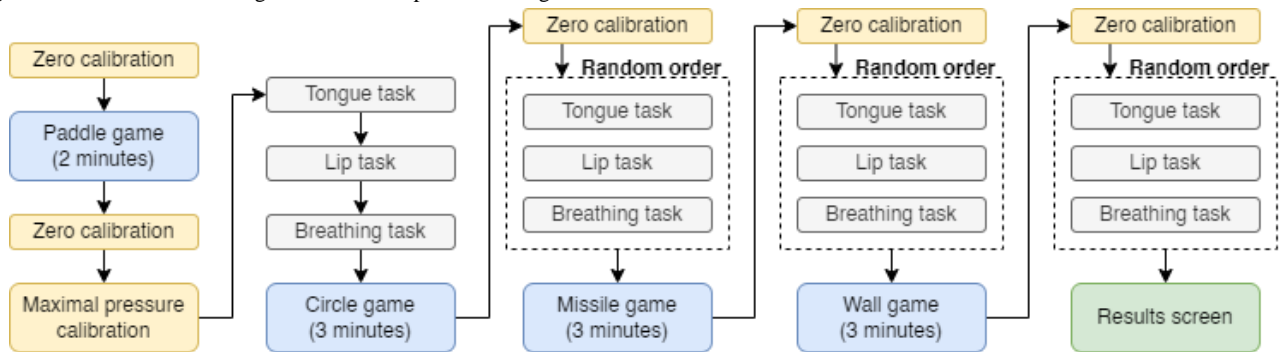


Software Description

The software comprises 3 main parts: initialization, training, and games. The participant must go through a sequence

alternating between these 3 parts to complete the session. **Figure 4** presents a flow chart describing the process of a training session.

Figure 4. Flow chart describing the different steps of a training session.



Zero Calibration

Once the mouthpiece has been installed, an initialization step must be performed. The participant is asked to release pressure from the tongue and lips to measure the baseline values with the display (presented in [Multimedia Appendix 1](#), Figure S1. This step is very important since the pressure in the cavities is also influenced by temperature variation. When the participant first puts the mouthpiece on, it will gradually heat up until a steady state is reached after approximately 1 minute. The first zero calibration is followed by a 2-minute game that does not require precision in the pressure measurements. This game allows the mouthpiece temperature to reach a steady-state level. A second zero calibration takes place at the end of the game. The zero calibration step is automatically repeated before each training session to ensure that no new temperature offset affects the pressure measurements.

Maximal Pressure Calibration

After the first zero calibration and the warm-up game, the participant is asked to set a baseline for both tongue protrusion and lip contraction maximal pressure. The participant starts on a page with a vertical bar corresponding to the actual developed tongue pressure, as presented in [Multimedia Appendix 1](#), Figure S2. The participant is asked to develop maximal tongue pressure before releasing and then press OK in 3 consecutive attempts. If the absolute deviation divided by the mean value of the maximal value is greater than 7.5%, the process is repeated; otherwise, the mean value is used as the baseline maximal tongue pressure. Subsequently, the same steps are completed for the lips.

Training

The training section consists of 3 tasks targeting, respectively, the tongue, the lips, and control of breathing. The tongue and

lip pressure tasks are based on the work of Svensson et al [18]. The display includes a white background, a green box, and an orange square, as presented in [Multimedia Appendix 1](#), Figure S3. The orange square displays the pressure level applied to the corresponding cavity. The participant is tasked with placing the orange square into the green box. The cycle starts with the green box at the bottom for the first 8 seconds, and the participant is asked to release the pressure from the cavity. The green box then rises to a level of pressure between 3% and 5% of the maximal pressure for 3 seconds. The first 1.5 seconds allow the user to react and adjust muscle contraction to fit within the targeted pressure range. A success score corresponding to the percentage of time the participant successfully applied pressure in the given range is computed during the last 1.5 seconds. The task repeats for 10 cycles, after which a cumulative score is computed.

For the breathing task, the participant is asked to continuously apply a small lip pressure of 4% of the maximal value. Visual feedback on the lip pressure is given to the user via a vertical bar. The participant is tasked with following a breathing pattern indicated on the screen. This task includes 7 breathing cycles, with 8 seconds for inhaling and 8 seconds for exhaling. The success of this task is defined by the fraction of the total time when the user has applied sufficient pressure. The user progression during this exercise is displayed with a horizontal progress bar, as presented in [Multimedia Appendix 1](#), Figure S4.

Games

The games are intended as relaxing activities between the formal metered training exercises described above. The first game is a paddle and ball game where a ball bounces off the top and bottom walls as well as paddles on the side. The user can move the paddles up in proportion to the pressure applied to the lip (blue paddle) and tongue (green paddle) cavities, respectively. The objective is to prevent the ball from hitting the edges behind the paddles. Every time the ball bounces off the paddle, a counter is incremented and the ball speeds up. The counter resets once the ball hits the edge behind a paddle. The user can track the high score of the current session on the screen, as presented in [Multimedia Appendix 1](#), Figure S5.

The second game is called the circle game. The user controls the position of a point; lip pressure controls the horizontal position and tongue pressure controls the vertical position. A yellow circle appears on the screen, and the user must combine lip and tongue pressure to place the point inside the circle. The circle turns green once the point is within its radius and must remain green for 3 seconds to succeed. Subsequently, the circle shrinks and appears at a new position. The user has 30 seconds to place the point steadily in the circle before the score resets; 7 seconds are added to the timer after each success. The current session's high score is also displayed on-screen. The circle game display is presented in [Multimedia Appendix 1](#), Figure S6.

In the missile game, the user must apply a small degree of tongue pressure and then release it to send a colored missile. The missile travels upward to a circle with 6 equal sections of different colors. Changing lip pressure allows the user to rotate the circle so that when the missile touches the circle, it collides

with the section with the matching color. For each success, the missile velocity increases. The missile game display is presented in [Multimedia Appendix 1](#), Figure S7.

The fourth game is the wall game, in which a ball moves horizontally toward a wall with a hole. The user must move the hole to let the ball through. The hole moves up incrementally when the user presses then releases lip pressure and moves down in the same way with tongue pressure. After each success, a new ball appears at a new height with a greater speed. The current and high scores are displayed on-screen, as presented in [Multimedia Appendix 1](#), Figure S8.

Results Screen

Once the user goes through the full training session (ie, 4 breathing tasks, 4 lip tasks, 4 tongue tasks, and 4 games), the results of each task are presented as a bar graph and are archived in a file.

Outcomes

The primary outcome of this study was the degree of completion of the exercises and adherence to treatment. The secondary outcomes were changes in sleep and breathing variables following training.

Compliance was defined as the number of completed sessions during the 4 weeks divided by 24 (6 sessions per week for 4 weeks). Tongue, lip, and breathing exercise success rates were defined by the percentage of time the participant successfully applied pressure in the given range. Results are presented with a 96% CI (1.96 SD). A mixed model was defined using a random intercept for the analysis of the changes in anthropometric data, the ESS score, sleep data, and the exercise success rate. One factor was associated to the before-and-after-intervention comparison. As the data were correlated, the normality assumption was verified with the Shapiro-Wilk test using residuals from the statistical model and transformed with the Cholesky metric. The graphical representation of marginal linear predictors with studentized residuals suggests the homogeneity of variances. Statistical significance was defined as a 2-tailed $P < .05$. Associations between AHI and adherence, as well as success rates, were assessed with Spearman correlations. Analyses were performed using SAS (version 9.4; SAS Institute).

Results

The characteristics of our study population are displayed in [Table 1](#). This study included 2 female participants. Breathing disturbances were mostly of moderate severity, except for 1 participant who had severe sleep apnea documented during the pretraining home sleep recording. A total of 2 of the 10 recruited participants did not complete the 4-week training program due to a lack of motivation. The remainder of the participants successfully completed at least 75% (18/24) of all sessions, as presented in [Table 2](#), with an average compliance of 91% (175/192).

For the tongue exercise, the success rate increased from 66% (211/320; SD 18%) on the first day to 85% (272/320; SD 17%) on the last day ($P = .05$). For the lip exercise, it increased from

78% (248/320; SD 18%) on the first day to 87% (278/320; SD 16%) on the last day ($P=.25$), as presented in Table 3. It is important to note that for both exercises the success score decreased in participant 2 while it improved for all other

participants, except for the lip exercise for participant 7. The success rate of the breathing exercise increased from 86% (275/320; SD 24%) the first day to 96% (307/320; SD 10%) on the last day ($P=.24$).

Table 1. Anthropometric, symptom, and sleep characteristics before and after 4 weeks.

	Pretraining value, mean (96% CI)	Posttraining value, mean (96% CI)	<i>P</i> value
Age (y)	48 (26 to 70)	N/A ^a	N/A
BMI (kg/m ²)	29.3 (25.8 to 32.8)	29.3 (26.0 to 32.6)	.73
ESS ^b score	11.3 (–0.1 to 22.7)	9.5 (–1.7 to 20.7)	.23
AHI ^c (events/h)	20.7 (2.9 to 38.5)	17.7 (–3.5 to 38.9)	.10
AHI supine (events/h)	33.4 (–3.4 to 70.2)	26.7 (–1.6 to 54.0)	.37
AHI RMI ^d (events/h)	25.5 (–6.1 to 57.1)	23.3 (–15.1 to 61.7)	.49
ODI ^e (events/h)	20.3 (–6.0 to 46.6)	19.2 (–9.0 to 47.4)	.66
Time supine (%)	43.7 (–19.0 to 106.4)	43.6 (–19.1 to 106.3)	.99
TST ^f (minutes)	420 (347 to 493)	395 (268 to 522)	.17
TST <90% SaO ₂ ^g (%)	1.6 (–0.9 to 4.1)	1.7 (2.0 to 5.4)	.75

^aN/A: not applicable.

^bESS: Epworth Sleepiness Scale.

^cAHI: apnea-hypopnea index.

^dRMI: respiratory mechanic instability.

^eODI: oxygen desaturation index.

^fTST: total sleep time.

^gSaO₂: oxygen saturation of arterial blood.

Table 2. Number of training sessions completed during the 4-week training program.

Participant	Week 1	Week 2	Week 3	Week 4
1 (total=24)	6	6	6	6
2 (total=20)	5	4	5	6
3 (total=24)	6	6	6	6
4 (total=22)	6	5	5	6
5 (total=22)	6	6	6	4
6 (total=23)	6	6	6	5
7 (total=18)	4	3	6	5
8 (total=22)	6	4	6	6

Table 3. Success rate at baseline and after 4 weeks of training.

Participant	Tongue success rate, %		Lip success rate, %	
	Baseline	Posttraining	Baseline	Posttraining
1	34	82	38	80
2	73	47	78	51
3	61	99	90	95
4	53	79	69	89
5	62	90	92	97
6	73	95	80	97
7	92	95	91	90
8	79	92	88	95

During the study period, there was no significant change in BMI, ESS score, AHI, and other polysomnography-derived parameters, as displayed in Table 1. Table 4 presents the variation in AHI before and after the 4-week training for each

participant. It illustrates that the index of each participant improved, with the exception of participants 2 and 3.

A significant correlation was found between the decrease in AHI in the supine position and the change in success rate for the lip exercise ($R_s = -0.76$; $P = .03$).

Table 4. Effects of 1 month of training on obstructive sleep apnea/hypopnea syndrome severity.

Participant	Pretraining AHI ^a , events/h	Posttraining AHI, events/h
1	13.6	11.6
2	15.4	19.6
3	41.0	40.8
4	17.4	6.3
5	25.0	21.3
6	20.5	14.9
7	12.6	8.3
8	20.2	19.0

^aAHI: apnea-hypopnea index.

Discussion

Principal Findings

The results of this study illustrate the feasibility of performing a training task focusing on the recruitment of different UA muscles while collecting major information such as adherence to the training program and objective measurements of task completion and success.

Comparison to Prior Work

Myofunctional therapy is a relatively new treatment for sleep-disordered breathing and is based on a combination of regular exercises aiming at enhancing muscle recruitment from various oral and oropharyngeal structures [13]. Although its effect on AHI and sleep apnea-related symptoms has been shown in an increasing number of studies [16], the main challenge to its success remains the objective assessment of program adherence and exercise performance [22]. Of the 10 participants initially recruited, 8 successfully completed more than 75% (18/24) of the total number of sessions over a period of 4 weeks. Two of the initially recruited participants left the

study less than 3 days after the start of the training. In comparison, rates of participant adherence reported in recent studies using myofunctional therapy with a mobile app were 75%, 90%, and 65% (15 minutes per session, 5 times/week for 3 months) [23-25]. In addition, Kim et al [26] found that in a myofunctional therapy support program with the help of exercise diaries, the reported adherence was 82%. In studies from Kim et al [26] and O'Connor et al [23-25], one could note the strong encouragement given to the patients through easy access to a health professional, encouraging text messages, or the use of a mobile app. In this study, a higher adherence rate was observed compared to the aforementioned studies. One potential explanation for this improvement is that the exercises were designed to induce motivation with games and visual feedback. Recently, a mobile app has been developed for this purpose that only requires a smartphone [24]. It first teaches the patient how to perform the exercise and then provides timely feedback on their performance. The results are saved over time, and the app promotes assiduity. It was observed that after 3 months, 75% of the patients completed the training at least 5 days a week [24] and the AHI of patients who adhered to the treatment decreased by 53.4% [23]. However, the quality tracking of the

exercises is limited by the functionality of the smartphone and requires covering the screen with cling film or hypoallergenic plastic wrap every session, since the tongue touches the screen [24]. The proposed custom mouthpiece introduced in this paper enables more accurate measurement of lip and tongue pressure, ensuring enhanced exercise quality.

Strengths

The multidimensional nature (exercises mixed with various training games) of the training program, its ability to immediately provide performance results for each exercise to the participants, and the measurement of treatment adherence are important strengths of our training program. As myofunctional therapy is based on an integrative approach, it is not possible to define which of the exercises may contribute most significantly to treatment success [13]. In recent studies, a combination of 9 exercises has been found to be sufficient to significantly decrease AHI [23,24]. Previous studies have focused on a single exercise and have found variable success [14,27]. Here, we used 3 exercises as a compromise between recruiting more of the muscles involved in the pathogenesis of OSAHS and keeping the workload at an acceptable level for participant motivation. Our results seem to show that a greater number of participants should have been recruited in order to see a higher impact on AHI reduction. Based on the results of our previous study [14], the training program duration was set at 4 weeks. This may have affected the results, as clear benefits in previously published myofunctional therapy studies were observed after 3 months of training. This difference may have helped to obtain a higher adherence rate while limiting the AHI reduction.

Limitations

There were 3 limitations of this study. The first was the success rate. This 4-week program did not significantly influence OSAHS severity to the degree that we previously found with an intensive in-lab tongue-protrusion training session that lasted 1 hour in 1 week [14]. However, it should be emphasized that although AHI did not decrease significantly, a correlation was found between the increase in success rate for the lip exercise and AHI decrease in the supine position. Therefore, one possible explanation for the modest decrease in AHI could be that the success rate of the present exercises was much higher than our previous in-lab trial (for both success rate during the first session and the rate of increase during the training period). Previously, it started from an average of 28% up to 65% for the last session. Similarly, an initial success rate of 25% was observed by Svensson et al [18] in a similar 1-week tongue-training program that was devised to increase corticomotor excitability. The high success rate in this study was likely mainly due to the adjustments that were made to the experimental set-up in order to make it ambulatory. However, in our success rate calculation, we did not take into account the results of the game sessions, which also involved a learning process and accounted for about half the duration of a training session. These games were designed mainly to boost the patients' motivation to continue

the program. Future exercise settings could be individually adjusted to adapt exercise targets to participants' baseline UA performance, with the goal of improving the success rate over time.

The second limitation was the selection of participants. Since it is not known to what extent anatomical UA abnormalities contribute to training program efficacy, no such selection criteria were used for our study population. It could be interesting to complete further studies focusing on patients with limited anatomical abnormalities according to practical clinical scores (ie, Mallampati and velopharyngeal scores). Apart from sleep apnea severity and degree of obesity, there were no selection criteria for participant selection. However, there may have been an indirect selection criterion due to the need for the participants to complete an additional preliminary visit at a dentist's office to perform the 3D tooth scan. This may have interfered with individual willingness to enter into the trial. Identifying participants who will remain engaged with the training program is a crucial issue for such a treatment strategy. Having a training device available for demonstration in the setting of a sleep clinic could definitely help to identify participants who are likely to follow the requirements of a training program.

The third limitation was the sample size. Yet another explanation for the lack of an AHI decrease is that our sample size was affected by the dropout of 2 participants and by the increase in AHI observed after the intervention in 1 participant. This participant was the only one with a decreasing overall success rate. This particular patient did have difficulty remaining focused throughout the training month, probably due to excessive sleepiness. As OSAHS is a multifactorial disorder [28] in which anatomical and nonanatomical factors can interact to modulate the severity of the disease, patient selection may play an important role in the success of OSAHS muscle training. Therefore, for this particular participant, treatments targeting traits other than low muscle tone or function would have been effective for decreasing AHI and related symptoms.

Conclusions

This study was an attempt to develop a prototype aimed at completing a simplified oral/oropharyngeal exercise program in an entertaining way in the comfort of a patient's own home. The program gives the patient visual feedback, as well as the ability to monitor improvement. Patients were instructed to perform tongue protrusion exercises, lip pressure exercises, and controlled breathing in various playful tasks 6 times a week for 4 weeks. Session duration was about 35 minutes. While the AHI reduction was not significant, we found that the success rate for improvement in the lip exercise was correlated with AHI reduction in the supine position ($R_s = -0.76$; $P = .03$). These results are a first steps toward the tuning of an ambulatory myofunctional therapy module able to accurately monitor the performance of participants in lip and tongue pressure exercises. This noninvasive approach may decrease the severity of OSAHS and represent an alternative to more invasive solutions, such as CPAP devices.

Acknowledgments

The authors would like to gratefully acknowledge the Fonds Alphonse-L'Espérance and the Ministère de l'Économie, de l'Innovation et de l'Énergie du Gouvernement du Québec for the financial support they provided for this research.

Data Availability

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

Authors' Contributions

PR contributed to writing (original draft, review, and editing), software, and visualization. JR contributed to conceptualization, project administration, funding acquisition, supervision, and writing (review and editing). AB-D contributed to supervision and writing (review and editing). JL contributed to supervision and writing (review and editing). SG contributed to data curation, formal analysis, methodology, and writing (review and editing). JFM contributed to resources and writing (review and editing). FS contributed to conceptualization, project administration, funding acquisition, supervision, and writing (review and editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the software.

[\[DOCX File , 439 KB - biomedeng_v9i1e51901_app1.docx \]](#)

References

1. McKeown P, O'Connor-Reina C, Plaza G. Breathing re-education and phenotypes of sleep apnea: a review. *J Clin Med* 2021 Jan 26;10(3):471 [FREE Full text] [doi: [10.3390/jcm10030471](https://doi.org/10.3390/jcm10030471)] [Medline: [33530621](https://pubmed.ncbi.nlm.nih.gov/33530621/)]
2. Young T, Palta M, Dempsey J, Peppard PE, Nieto FJ, Hla KM. Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study. *WMJ* 2009 Aug;108(5):246-249 [FREE Full text] [Medline: [19743755](https://pubmed.ncbi.nlm.nih.gov/19743755/)]
3. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol* 2013 May 01;177(9):1006-1014 [FREE Full text] [doi: [10.1093/aje/kws342](https://doi.org/10.1093/aje/kws342)] [Medline: [23589584](https://pubmed.ncbi.nlm.nih.gov/23589584/)]
4. Flemons WW, Reimer MA. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. *Am J Respir Crit Care Med* 1998 Aug;158(2):494-503. [doi: [10.1164/ajrccm.158.2.9712036](https://doi.org/10.1164/ajrccm.158.2.9712036)] [Medline: [9700127](https://pubmed.ncbi.nlm.nih.gov/9700127/)]
5. Courtney R. Breathing retraining in sleep apnoea: a review of approaches and potential mechanisms. *Sleep Breath* 2020 Dec;24(4):1315-1325. [doi: [10.1007/s11325-020-02013-4](https://doi.org/10.1007/s11325-020-02013-4)] [Medline: [31940122](https://pubmed.ncbi.nlm.nih.gov/31940122/)]
6. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med* 2005 Nov 10;353(19):2034-2041. [doi: [10.1056/nejmoa043104](https://doi.org/10.1056/nejmoa043104)]
7. Terán-Santos J, Jiménez-Gómez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. Cooperative Group Burgos-Santander. *N Engl J Med* 1999 Mar 18;340(11):847-851. [doi: [10.1056/NEJM199903183401104](https://doi.org/10.1056/NEJM199903183401104)] [Medline: [10080847](https://pubmed.ncbi.nlm.nih.gov/10080847/)]
8. Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000 May 11;342(19):1378-1384. [doi: [10.1056/NEJM200005113421901](https://doi.org/10.1056/NEJM200005113421901)] [Medline: [10805822](https://pubmed.ncbi.nlm.nih.gov/10805822/)]
9. He J, Kryger MH, Zorick FJ, Conway W, Roth T. Mortality and apnea index in obstructive sleep apnea. Experience in 385 male patients. *Chest* 1988 Jul;94(1):9-14. [doi: [10.1378/chest.94.1.9](https://doi.org/10.1378/chest.94.1.9)] [Medline: [3289839](https://pubmed.ncbi.nlm.nih.gov/3289839/)]
10. Baguet J, Hammer L, Lévy P, Pierre H, Launois S, Mallion J, et al. The severity of oxygen desaturation is predictive of carotid wall thickening and plaque occurrence. *Chest* 2005 Nov;128(5):3407-3412. [doi: [10.1378/chest.128.5.3407](https://doi.org/10.1378/chest.128.5.3407)] [Medline: [16304292](https://pubmed.ncbi.nlm.nih.gov/16304292/)]
11. Weaver TE, Grunstein RR. Adherence to continuous positive airway pressure therapy: the challenge to effective treatment. *Proc Am Thorac Soc* 2008 Feb 15;5(2):173-178 [FREE Full text] [doi: [10.1513/pats.200708-119MG](https://doi.org/10.1513/pats.200708-119MG)] [Medline: [18250209](https://pubmed.ncbi.nlm.nih.gov/18250209/)]
12. Zozula R, Rosen R. Compliance with continuous positive airway pressure therapy: assessing and improving treatment outcomes. *Curr Opin Pulm Med* 2001 Nov;7(6):391-398. [doi: [10.1097/00063198-200111000-00005](https://doi.org/10.1097/00063198-200111000-00005)] [Medline: [11706314](https://pubmed.ncbi.nlm.nih.gov/11706314/)]
13. Guimarães KC, Drager LF, Genta PR, Marcondes BF, Lorenzi-Filho G. Effects of oropharyngeal exercises on patients with moderate obstructive sleep apnea syndrome. *Am J Respir Crit Care Med* 2009 May 15;179(10):962-966. [doi: [10.1164/rccm.200806-981OC](https://doi.org/10.1164/rccm.200806-981OC)] [Medline: [19234106](https://pubmed.ncbi.nlm.nih.gov/19234106/)]
14. Rousseau E, Silva C, Gakwaya S, Sériès F. Effects of one-week tongue task training on sleep apnea severity: a pilot study. *Can Respir J* 2015;22(3):176-178 [FREE Full text] [doi: [10.1155/2015/583549](https://doi.org/10.1155/2015/583549)] [Medline: [25874736](https://pubmed.ncbi.nlm.nih.gov/25874736/)]
15. Melo-Silva C, Gakwaya S, Rousseau E, Borel JC, Sériès F. Effects of one week tongue-task training on sleep apnea severity, force and endurance of the tongue protrusion in obstructive sleep apnea patients: a pilot study. *Am J Respir Crit Care Med* 2014;189:A5622 [FREE Full text] [doi: [10.1155/2015/583549](https://doi.org/10.1155/2015/583549)] [Medline: [25874736](https://pubmed.ncbi.nlm.nih.gov/25874736/)]

16. Camacho M, Certal V, Abdullatif J, Zaghi S, Ruoff CM, Capasso R, et al. Myofunctional therapy to treat obstructive sleep apnea: a systematic review and meta-analysis. *Sleep* 2015 May 01;38(5):669-675 [FREE Full text] [doi: [10.5665/sleep.4652](https://doi.org/10.5665/sleep.4652)] [Medline: [25348130](https://pubmed.ncbi.nlm.nih.gov/25348130/)]
17. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, American Academy of Sleep Medicine. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2012 Oct 15;8(5):597-619 [FREE Full text] [doi: [10.5664/jcsm.2172](https://doi.org/10.5664/jcsm.2172)] [Medline: [23066376](https://pubmed.ncbi.nlm.nih.gov/23066376/)]
18. Svensson P, Romaniello A, Arendt-Nielsen L, Sessle BJ. Plasticity in corticomotor control of the human tongue musculature induced by tongue-task training. *Exp Brain Res* 2003 Sep;152(1):42-51. [doi: [10.1007/s00221-003-1517-2](https://doi.org/10.1007/s00221-003-1517-2)] [Medline: [12830348](https://pubmed.ncbi.nlm.nih.gov/12830348/)]
19. Bégin-Drolet A, Dussault M, Fernandez SA, Larose-Dutil J, Leask RL, Hoesli CA, et al. Design of a 3D printer head for additive manufacturing of sugar glass for tissue engineering applications. *Addit Manuf* 2017 May;15:29-39. [doi: [10.1016/j.addma.2017.03.006](https://doi.org/10.1016/j.addma.2017.03.006)]
20. Gauvin-Rossignol G, Legros P, Ruel J, Fortin M, Bégin-Drolet A. Sugar glass fugitive ink loaded with calcium chloride for the rapid casting of alginate scaffold designs. *Heliyon* 2018 Jul;4(7):e00680 [FREE Full text] [doi: [10.1016/j.heliyon.2018.e00680](https://doi.org/10.1016/j.heliyon.2018.e00680)] [Medline: [29998199](https://pubmed.ncbi.nlm.nih.gov/29998199/)]
21. Moeun BN, Fernandez SA, Collin S, Gauvin-Rossignol G, Lescot T, Fortin M, et al. Improving the 3D printability of sugar glass to engineer sacrificial vascular templates. *3D Print Addit Manuf* 2023 Oct 01;10(5):869-886 [FREE Full text] [doi: [10.1089/3dp.2021.0147](https://doi.org/10.1089/3dp.2021.0147)] [Medline: [37886415](https://pubmed.ncbi.nlm.nih.gov/37886415/)]
22. Koka V, De Vito A, Roisman G, Petitjean M, Filograna Pignatelli GR, Padovani D, et al. Orofacial myofunctional therapy in obstructive sleep apnea syndrome: a pathophysiological perspective. *Medicina (Kaunas)* 2021 Apr 01;57(4):323 [FREE Full text] [doi: [10.3390/medicina57040323](https://doi.org/10.3390/medicina57040323)] [Medline: [33915707](https://pubmed.ncbi.nlm.nih.gov/33915707/)]
23. O'Connor-Reina C, Ignacio Garcia JM, Rodriguez Ruiz E, Morillo Dominguez MDC, Ignacio Barrios V, Baptista Jardin P, et al. Myofunctional therapy app for severe apnea-hypopnea sleep obstructive syndrome: pilot randomized controlled trial. *JMIR Mhealth Uhealth* 2020 Nov 09;8(11):e23123 [FREE Full text] [doi: [10.2196/23123](https://doi.org/10.2196/23123)] [Medline: [33093013](https://pubmed.ncbi.nlm.nih.gov/33093013/)]
24. O'Connor Reina C, Plaza G, Ignacio-Garcia JM, Baptista Jardin P, Garcia-Iriarte MT, Casado-Morente JC, et al. New mHealth application software based on myofunctional therapy applied to sleep-disordered breathing in non-compliant subjects. *Sleep Sci Pract* 2020 Feb 05;4(1):3. [doi: [10.1186/s41606-019-0040-8](https://doi.org/10.1186/s41606-019-0040-8)]
25. O'Connor-Reina C, Ignacio Garcia JM, Rodriguez Alcala L, Rodríguez Ruiz E, Garcia Iriarte MT, Casado Morente JC, et al. Improving adherence to myofunctional therapy in the treatment of sleep-disordered breathing. *J Clin Med* 2021 Dec 09;10(24):5772 [FREE Full text] [doi: [10.3390/jcm10245772](https://doi.org/10.3390/jcm10245772)] [Medline: [34945068](https://pubmed.ncbi.nlm.nih.gov/34945068/)]
26. Kim J, Oh EG, Choi M, Choi SJ, Joo EY, Lee H, et al. Development and evaluation of myofunctional therapy support program (MTSP) based on self-efficacy theory for patients with obstructive sleep apnea. *Sleep Breath* 2020 Sep;24(3):1051-1058. [doi: [10.1007/s11325-019-01957-6](https://doi.org/10.1007/s11325-019-01957-6)] [Medline: [31811542](https://pubmed.ncbi.nlm.nih.gov/31811542/)]
27. Poncin W, Correvon N, Tam J, Borel J, Berger M, Liistro G, et al. The effect of tongue elevation muscle training in patients with obstructive sleep apnea: a randomised controlled trial. *J Oral Rehabil* 2022 Nov;49(11):1049-1059 [FREE Full text] [doi: [10.1111/joor.13369](https://doi.org/10.1111/joor.13369)] [Medline: [36081312](https://pubmed.ncbi.nlm.nih.gov/36081312/)]
28. Eckert DJ. Phenotypic approaches to obstructive sleep apnoea - New pathways for targeted therapy. *Sleep Med Rev* 2018 Feb;37:45-59. [doi: [10.1016/j.smrv.2016.12.003](https://doi.org/10.1016/j.smrv.2016.12.003)] [Medline: [28110857](https://pubmed.ncbi.nlm.nih.gov/28110857/)]

Abbreviations

- AHI:** apnea-hypopnea index
CPAP: continuous positive airway pressure
ESS: Epworth Sleepiness Scale
OSAHS: obstructive sleep apnea/hypopnea syndrome
UA: upper airway

Edited by T Leung; submitted 16.08.23; peer-reviewed by C O'Connor-Reina, T Penzel; comments to author 02.10.23; revised version received 20.01.24; accepted 31.01.24; published 15.04.24.

Please cite as:

Roberge P, Ruel J, Bégin-Drolet A, Lemay J, Gakwaya S, Masse JF, Sériès F

Preliminary Assessment of an Ambulatory Device Dedicated to Upper Airway Muscle Training in Patients With Sleep Apnea: Proof-of-Concept Study

JMIR Biomed Eng 2024;9:e51901

URL: <https://biomedeng.jmir.org/2024/1/e51901>

doi: [10.2196/51901](https://doi.org/10.2196/51901)

PMID: [38875673](https://pubmed.ncbi.nlm.nih.gov/38875673/)

©Patrice Roberge, Jean Ruel, André Bégin-Drolet, Jean Lemay, Simon Gakwaya, Jean-François Masse, Frédéric Sériès. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 15.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Impact of Audio Data Compression on Feature Extraction for Vocal Biomarker Detection: Validation Study

Jessica Oreskovic¹, MAS; Jaycee Kaufman¹, MS; Yan Fossat¹, MSc

Klick Labs, Toronto, ON, Canada

Corresponding Author:

Yan Fossat, MSc

Klick Labs

175 Bloor St E #300

3rd floor

Toronto, ON, M4W3R8

Canada

Phone: 1 6472068717

Email: yfossat@klick.com

Abstract

Background: Vocal biomarkers, derived from acoustic analysis of vocal characteristics, offer noninvasive avenues for medical screening, diagnostics, and monitoring. Previous research demonstrated the feasibility of predicting type 2 diabetes mellitus through acoustic analysis of smartphone-recorded speech. Building upon this work, this study explores the impact of audio data compression on acoustic vocal biomarker development, which is critical for broader applicability in health care.

Objective: The objective of this research is to analyze how common audio compression algorithms (MP3, M4A, and WMA) applied by 3 different conversion tools at 2 bitrates affect features crucial for vocal biomarker detection.

Methods: The impact of audio data compression on acoustic vocal biomarker development was investigated using uncompressed voice samples converted into MP3, M4A, and WMA formats at 2 bitrates (320 and 128 kbps) with MediaHuman (MH) Audio Converter, WonderShare (WS) UniConverter, and Fast Forward Moving Picture Experts Group (FFmpeg). The data set comprised recordings from 505 participants, totaling 17,298 audio files, collected using a smartphone. Participants recorded a fixed English sentence up to 6 times daily for up to 14 days. Feature extraction, including pitch, jitter, intensity, and Mel-frequency cepstral coefficients (MFCCs), was conducted using Python and Parselmouth. The Wilcoxon signed rank test and the Bonferroni correction for multiple comparisons were used for statistical analysis.

Results: In this study, 36,970 audio files were initially recorded from 505 participants, with 17,298 recordings meeting the fixed sentence criteria after screening. Differences between the audio conversion software, MH, WS, and FFmpeg, were notable, impacting compression outcomes such as constant or variable bitrates. Analysis encompassed diverse data compression formats and a wide array of voice features and MFCCs. Wilcoxon signed rank tests yielded *P* values, with those below the Bonferroni-corrected significance level indicating significant alterations due to compression. The results indicated feature-specific impacts of compression across formats and bitrates. MH-converted files exhibited greater resilience compared to WS-converted files. Bitrate also influenced feature stability, with 38 cases affected uniquely by a single bitrate. Notably, voice features showed greater stability than MFCCs across conversion methods.

Conclusions: Compression effects were found to be feature specific, with MH and FFmpeg showing greater resilience. Some features were consistently affected, emphasizing the importance of understanding feature resilience for diagnostic applications. Considering the implementation of vocal biomarkers in health care, finding features that remain consistent through compression for data storage or transmission purposes is valuable. Focused on specific features and formats, future research could broaden the scope to include diverse features, real-time compression algorithms, and various recording methods. This study enhances our understanding of audio compression's influence on voice features and MFCCs, providing insights for developing applications across fields. The research underscores the significance of feature stability in working with compressed audio data, laying a foundation for informed voice data use in evolving technological landscapes.

(*JMIR Biomed Eng* 2024;9:e56246) doi:[10.2196/56246](https://doi.org/10.2196/56246)

KEYWORDS

vocal biomarker; biomarker; biomarkers; sound; sounds; audio; compression; voice; acoustic; acoustics; audio compression; feature extraction; Python; speech; detect; detection; algorithm; algorithms

Introduction

Background

Vocal biomarkers are emerging as a promising accessible and noninvasive avenue for medical screening, diagnostics, and monitoring [1]. These biomarkers are unique characteristics or acoustic patterns of an individual's voice that can hold valuable information about their physical and mental well-being [2]. Human voice production requires the coordination of multiple biological systems; perturbations in these systems induced by various conditions or diseases can result in alterations in the characteristics of the human voice [3]. Potential applications of vocal biomarkers are diverse, including the identification of neurological disorders, cardiovascular diseases, respiratory conditions, and mental health disorders, among others [2,4-6].

In our previous work, "Acoustic Analysis and Prediction of Type 2 Diabetes Mellitus Using Smartphone-Recorded Voice Segments" [7], smartphone-recorded speech was used to predict type 2 diabetes mellitus through a comprehensive acoustic analysis [7]. The study demonstrated the feasibility of using acoustic features from smartphone-recorded voice data to predict the presence of this disorder, highlighting the valuable diagnostic potential of vocal biomarkers in the context of a specific health condition [7]. Building upon this prior research, we aim to assess the impact of audio compression on acoustic vocal biomarker development, which is crucial for the broader applicability of this emerging field.

The development of acoustic vocal biomarkers relies on the analysis of voice data, and this process is multifaceted. One critical aspect of this analysis is feature extraction, which involves identifying and quantifying relevant acoustic features from the voice data [8]. These features may encompass a wide range of parameters such as pitch, spectral properties, prosodic patterns, and various other characteristics that carry meaningful information about the speaker's health status [2,8]. Accurate and robust feature extraction is pivotal for the successful identification and interpretation of vocal biomarkers.

Voice data are often captured, transmitted, and stored in various digital formats that may include compression, a common practice used to reduce the size of audio files, making them more manageable and efficient for storage and transmission [9]. It is necessary to consider the potential impact of audio data compression on the overall process of vocal biomarker development as the process can have significant effects on the audio [10]. Compression algorithms are widely applied to raw, high-quality audio (typically waveform audio file format) and can be classified as lossy or lossless [11]. Lossy compression algorithms reduce file size to as low as 10% of the original size by removing mostly inaudible audio data, while lossless preserves all the original audio data and only compresses to approximately 50% [12]. Some of the most common lossy formats include MP3, M4A, and WMA [12]. These formats offer different trade-offs between file size and audio quality,

and each may introduce specific artifacts and alterations to the original acoustic data.

Previous research on how data compression impacts voice signals has found that different microphones and MP3 compression bitrates on sustained vowel sounds can significantly affect feature values [10]. Research has found that various digital platforms and their audio codecs affect the voice in a way that challenges voice recognition processes specifically by narrowing the frequency band and centrally shifting frequencies at the upper and lower limits [13]. While differing microphones can introduce differences in audio data depending on specifications, smartphone microphones have been found to collect high-quality audio data suitable for acoustic analysis [14].

This exploratory research aims to investigate the effect of common audio data compression algorithms, such as MP3, AAC (compression algorithm for M4A), and WMA, on the vocal biomarker feature extraction process. Additionally, the effect of compression bitrate or encoder type will be analyzed to determine whether these factors make a difference within each format. Understanding the impact of popular data compression methods on acoustic vocal biomarker analysis is important as it can significantly affect the quality and interpretability of biomarker data [15,16]. Moreover, this knowledge can guide the development of best practices and inform the compression implementation process for the specific needs of health care applications, such as remote medical care involving telephone or video conferencing, thereby minimizing the risk of unintentional distortion of vocal biomarkers.

Objective

The objective of this research is to analyze the effect of several common audio data compression algorithms: MP3, M4A, and WMA, in 2 common bitrates, completed by 3 different conversion tools, on feature extraction from voice data for vocal biomarker detection.

Methods

Overview

In this research, acoustic features were derived from uncompressed voice samples, which were subsequently converted into MP3, M4A, and WMA formats using 3 distinct tools, namely MediaHuman (MH) Audio Converter, WonderShare (WS) UniConverter, and Fast Forward Moving Picture Experts Group (FFmpeg) across 2 different bitrates (320 and 128 kbps). MH, WS, and FFmpeg conversion tools were selected because of their accessibility as free, downloadable audio conversion software. Our goal was to explore how different audio conversion tools, formats, and 2 specific bitrates affect the data set used to develop a biomarker prediction model [7]. By focusing on these tools and bitrates, we aimed to provide insights into the potential impact of common audio compression methods on the extracted voice features. This approach allowed for a manageable analysis while paving the way for future

research to delve deeper into the nuances of audio compression effects on biomarker prediction models.

Data and Participants

This research was conducted using a data set of audio recordings that were collected from 505 participants (mean age 41.03, SD 13.29 years, 336 male participants) recruited between August 30, 2021, and June 30, 2022, for a study in India [7]. Participants were instructed to record a short English phrase up to 6 times daily using their smartphone for 14 consecutive days. As these data were originally recorded for research involving diabetes, the phrase was “Hello. How are you? What is my glucose level right now?” All audio files used in the research originated in the uncompressed waveform audio file format, 16-bit 44.1 kHz.

Participants in this study used a variety of smartphone models for data recording. While efforts were made to request recordings in quiet environments, the inherent difficulty in controlling recording conditions may have introduced variability in the recorded speech data. No preliminary tests were conducted to assess the recording quality across different smartphone models, and no preprocessing techniques were applied to address potential hardware variations in the recorded speech data. It is noteworthy that the intention of the prediction model was to be run on a smartphone; therefore, the recordings were made using smartphone uncompressed audio to align with the intended application context.

File Conversion

To explore the impact of diverse data compression methods, the original files underwent conversion using MH (version 2.2.2), WS (version 15), and FFmpeg (version 6.1.1) in Python (version 3.10.11; Python Software Foundation) on a PC. Three distinct compression algorithms—MP3, M4A, and WMA—at 2 bitrates—128 kbps and 320 kbps—were applied to simulate real-world scenarios where audio data are commonly subjected to different compression algorithms for storage and transmission purposes. The sample rate (44.1 kHz) and the channels (stereo) were kept consistent over all formats. The choice of encoders used in the research was not a primary consideration; rather, our focus was on comparing the results obtained from different compression methods. It is worth noting that the selected encoders were accessible, free, and capable of batch processing multiple files, which facilitated efficient experimentation. Despite maintaining consistency in factors such as bitrate, channels, and formats between the 3 encoders, there are features of the tools that remain hidden that could potentially cause differences in the converted files, such as the encoding mode (ie, constant or variable bitrate) or other encoding options. However, these hidden features are not a large concern because the objective of the study was to compare compressed and uncompressed data rather than comparing between compression. The incorporation of multiple encoders served the purpose of discerning whether factors beyond just bitrate and file format influenced feature values.

Feature Extraction and Comparison

We chose to use the same feature set ([Multimedia Appendix 1](#)) as in our previous research on developing a voice-to-type 2 diabetes model to maintain consistency and leverage their

established effectiveness in capturing relevant biological information from voice data [7]. Acoustic features were extracted from both the original waveform audio file format files and the compressed audio formats using Python (version 3.10.11; Python Software Foundation). The voice feature extraction process leveraged Parselmouth, a Python integration of Praat speech and voice analysis software [17,18], ensuring robustness and accuracy in feature extraction. The extracted features aimed to capture pertinent acoustic characteristics of the voice data, such as pitch, jitter, and intensity, as well as Mel-frequency cepstral coefficients (MFCCs) [19], which have demonstrated efficacy in capturing subtle variations in vocal properties associated with health conditions.

Notable perceived voice qualities such as breathiness, hoarseness, and roughness, which typically present with elevated levels of shimmer and jitter, were often associated with certain pathological conditions and were therefore included in the biomarker development as well as this research [7,20]. While acoustic analysis is mainly performed using sustained phonation of vowel sounds, recent studies have demonstrated the use of shimmer and jitter measurements in identifying dysphonia even when calculated from entire sentence recordings [20]. Thus, because the data set was originally studied for the purpose of biomarker development, we chose to include the evaluations of shimmer and jitter alongside traditional vocal parameters such as pitch, intensity, and harmonic noise ratio in this analysis of how audio data compression impacts feature values.

Given the non-Gaussian distribution of feature data, assessed via the Shapiro-Wilk test, a nonparametric approach—specifically, the Wilcoxon signed rank test—was adopted for statistical analysis. This paired test aimed to evaluate the impact of each compression method on audio features by comparing the features extracted from the original uncompressed files with those obtained from each compressed format individually. In this study, the Bonferroni correction method was used to account for multiple comparisons. Given our focus on assessing the impact of each conversion method relative to the original feature values rather than comparing between different treatments, this correction was deemed appropriate. This approach allowed us to effectively manage the potential for false positives while evaluating the stability of feature values across different compression methods.

Ethical Considerations

The protocol (ID MGCTS107) received ethics approval by Saanvi Ethical Research LLP, all participants signed informed consent, and data were stored in a secure cloud database with no identifying information. Participants were compensated for their time.

Results

Data and Participants

A total of 36,970 audio files were recorded from the 505 participants who completed the study. Speech-to-text screening ensured that the audio files adhered to the fixed sentence criteria and were devoid of substantial background noise, resulting in

a total of 17,298 recordings. All participants were native to India.

File Conversion

The noncustomizable differences between the audio conversion software MH Audio Converter and WS UniConverter manifested

in evident variations in the converted files. Table 1 displays the differences in compression ratio and data set size, highlighting these distinctions and emphasizing the impact of software-specific characteristics on the compression outcomes such as constant or variable bitrates.

Table 1. Compression specifications for each data compression method including the final size of the data set and compression ratio.

Format and tools	Bitrate (kbps)	Data set size (GB)	Compression ratio
MP3			
MediaHuman	• 128	• 1.29	• 5.42
	• 320	• 3.22	• 2.17
WonderShare	• 128	• 0.43	• 16.26
	• 320	• 1.05	• 6.66
FFmpeg ^a	• 128	• 1.29	• 5.42
	• 320	• 3.20	• 2.18
M4A			
MediaHuman	• 128	• 1.39	• 5.03
	• 320	• 5.10	• 1.37
WonderShare	• 128	• 0.47	• 14.87
	• 320	• 1.10	• 6.35
FFmpeg	• 128	• 1.26	• 5.55
	• 320	• 2.18	• 3.21
WMA			
MediaHuman	• 128	• 1.39	• 5.03
	• 320	• 5.49	• 1.27
WonderShare	• 128	• 0.70	• 9.99
	• 320	• 1.33	• 5.26
FFmpeg	• 128	• 1.39	• 5.03
	• 320	• 5.49	• 1.27

^aFFmpeg: Fast Forward Moving Picture Experts Group.

Feature Extraction and Comparison

This research investigated the influence of diverse data compression formats on an extensive array of voice features and MFCCs. The corresponding *P* values for each feature are provided in the subsequent table from the results of the 756 Wilcoxon signed rank tests. *P* values below the level of significance, 6.61×10^{-5} with the Bonferroni correction (Table S1-S3 in [Multimedia Appendix 2](#)), signify a notable difference in feature values between the original .wav format and the corresponding compressed format, indicating a significant alteration due to compression. Conversely, features with *P* values greater than 6.61×10^{-5} (Table S1-S3 in [Multimedia Appendix 2](#)) are deemed robust, suggesting their resilience to the compression process.

Discussion

Overview

This investigation illuminated the effects of diverse audio file compression methods on a broad spectrum of voice features and MFCCs. The results revealed that the impact of data compression is feature specific and varies across different encoders, formats, and bitrates.

Principal Findings

The encoder played a substantial role in influencing voice features, with MH- and FFmpeg-converted files demonstrating greater resilience to compression compared to WS-converted files, regardless of the format. For MH, WS, and FFmpeg, there were 15, 6, and 21 features, respectively, that had at least 1 format or bitrate combination that was unaffected by the conversion. A total of 59 compressed feature comparisons showed stability for MH, 8 for WS, and 67 for FFmpeg (Table S1-S3 in [Multimedia Appendix 2](#)). The conversion bitrate also

exhibited an impact on feature stability, with some features remaining consistent for both bitrates, while others were affected uniquely at either 128 kbps or 320 kbps. A total of 38 feature comparison cases (of the total of 134) were only affected by compression for a single bitrate. Of those 38, 15 feature comparisons were only unaffected with 128 kbps, while 23 were stable for compression at only 320 kbps. MH and FFmpeg conversions had more features unaffected when conversions were done with a bitrate of 320 kbps compared to 128 kbps. Additionally, the voice features were found to be more stable than the MFCCs. The findings indicate that not all voice features respond equally to audio file compression. Certain features exhibited robustness and remained consistent despite compression, holding promise for applications involving compressed voice data storage or transmission. For instance, in our previous work on type 2 diabetes prediction from voice, features such as mean fundamental frequency/pitch (meanF0), pitch SD (stdevF0), and relative average perturbation jitter (rapJitter) remained consistent across several compression methods, including MP3 from MH at 320 kbps and FFmpeg at both 320 and 128 kbps and WMA from MH and FFmpeg at both 128 kbps and 320 kbps (Table S1-S3 in [Multimedia Appendix 2](#)) [7]. For the male prediction model, 1 of the 2 features (meanI) was significantly affected by all conversion methods. The second feature (apq11) remained stable for conversions with MH and FFmpeg to WMA format at both bitrates, MP3 at 320 kbps, and MH-converted M4A at 320 kbps. (Table S1-S3 in [Multimedia Appendix 2](#)) [7]. However, this study also identified features significantly altered by compression (Table S1-S3 in [Multimedia Appendix 2](#)), emphasizing the need to understand the stability and sensitivity of individual features for maintaining accuracy and interpretability in applications like health care diagnostics and voice recognition.

Vocal biomarkers, being a relatively new concept, are predominantly situated within the realm of research rather than practical settings where considerations for data storage and transmission are paramount. The study's implications extend to various fields, particularly in health care, where voice data are increasingly used for disease detection and monitoring. When dealing with features significantly influenced by a specific compression algorithm, considerations should be made to preserve accuracy in applications requiring high diagnostic precision. The study suggests that certain voice features can

withstand common data compression formats, enabling the use of compressed data in medical applications without compromising diagnostic accuracy, depending on the features. This is crucial in scenarios involving limited bandwidth for audio data transmission or storage constraints, where choosing an appropriate compression format while considering feature resilience becomes pivotal. Conversely, for research applications where features are being investigated, the use of uncompressed or lossless compression is essential.

Limitations and Future Directions

This study has several limitations. First, while it focused on a specific set of voice features and how they were changed based on compression formats, future research could benefit from isolating compression settings to study their individual effects rigorously. Second, controlling microphone and recording settings could enhance data consistency and reliability, as variations in these factors may introduce confounding variables. Additionally, exploring different recording sentences could provide insights into how content variability influences the impact of compression on feature extraction. Finally, a broader exploration of diverse features beyond those examined in this study, such as spectral or temporal features, could offer a more comprehensive understanding of the impact of compression on acoustic vocal biomarkers.

Conclusions

In this research, we have provided insights into the influence of audio data compression on feature values used in biomarker prediction model development. Our findings underscore the importance of considering compression effects in the design and optimization of diagnostic tools reliant on voice-based biomarkers. Through analysis and statistical comparisons, we have demonstrated the nuanced impact of compression formats, bitrates, and conversion tools on the stability and reliability of extracted feature values. By revealing these effects, our research not only advances our understanding of the complex interplay between audio data processing and biomarker extraction but also offers practical implications for health care practitioners and researchers. Moving forward, the findings pave the way for future investigations aimed at refining compression strategies, exploring alternative extraction methodologies, and ultimately enhancing the accuracy and efficacy of biomarker-based diagnostic models in clinical practice.

Acknowledgments

This research was internally funded by Klick Inc.

Data Availability

Data were commissioned by Klick Health, a private business, and are owned by Klick who does not allow sharing them.

Authors' Contributions

Data processing and analysis was done by JO. Feature extraction method was developed by JK. All authors participated in manuscript editing.

Conflicts of Interest

All authors are employees of Klick Inc.

Multimedia Appendix 1

Voice feature set.

[[DOCX File, 14 KB - biomedeng_v9i1e56246_app1.docx](#)]

Multimedia Appendix 2

Pairwise comparisons for feature values between compressed and uncompressed audio data.

[[DOCX File, 51 KB - biomedeng_v9i1e56246_app2.docx](#)]

References

1. Sara JDS, Orbelo D, Maor E, Lerman LO, Lerman A. Guess what we can hear-novel voice biomarkers for the remote detection of disease. *Mayo Clin Proc* 2023;98(9):1353-1375 [[FREE Full text](#)] [doi: [10.1016/j.mayocp.2023.03.007](https://doi.org/10.1016/j.mayocp.2023.03.007)] [Medline: [37661144](https://pubmed.ncbi.nlm.nih.gov/37661144/)]
2. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for health: the use of vocal biomarkers from research to clinical practice. *Digit Biomark* 2021;5(1):78-88 [[FREE Full text](#)] [doi: [10.1159/000515346](https://doi.org/10.1159/000515346)] [Medline: [34056518](https://pubmed.ncbi.nlm.nih.gov/34056518/)]
3. Zhang Z. Mechanics of human voice production and control. *J Acoust Soc Am* 2016;140(4):2614 [[FREE Full text](#)] [doi: [10.1121/1.4964509](https://doi.org/10.1121/1.4964509)] [Medline: [27794319](https://pubmed.ncbi.nlm.nih.gov/27794319/)]
4. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun* 2015;71:10-49 [[FREE Full text](#)] [doi: [10.1016/j.specom.2015.03.004](https://doi.org/10.1016/j.specom.2015.03.004)]
5. Sara JDS, Maor E, Borlaug B, Lewis BR, Orbelo D, Lerman LO, et al. Non-invasive vocal biomarker is associated with pulmonary hypertension. *PLoS One* 2020;15(4):e0231441 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0231441](https://doi.org/10.1371/journal.pone.0231441)] [Medline: [32298301](https://pubmed.ncbi.nlm.nih.gov/32298301/)]
6. Maor E, Tsur N, Barkai G, Meister I, Makmel S, Friedman E, et al. Noninvasive vocal biomarker is associated with severe acute respiratory syndrome Coronavirus 2 infection. *Mayo Clin Proc Innov Qual Outcomes* 2021;5(3):654-662 [[FREE Full text](#)] [doi: [10.1016/j.mayocpiqo.2021.05.007](https://doi.org/10.1016/j.mayocpiqo.2021.05.007)] [Medline: [34007956](https://pubmed.ncbi.nlm.nih.gov/34007956/)]
7. Kaufman JM, Thommandram A, Fossat Y. Acoustic analysis and prediction of type 2 diabetes mellitus using smartphone-recorded voice segments. *Mayo Clin Proc Digit Health* 2023;1(4):534-544 [[FREE Full text](#)] [doi: [10.1016/j.mcpdig.2023.08.005](https://doi.org/10.1016/j.mcpdig.2023.08.005)]
8. Sharma G, Umapathy K, Krishnan S. Trends in audio signal feature extraction methods. *Appl Acoust* 2020;158:107020 [[FREE Full text](#)] [doi: [10.1016/j.apacoust.2019.107020](https://doi.org/10.1016/j.apacoust.2019.107020)]
9. Pan DY. Digital audio compression. *Digit Tech J* 1993;5(2):28-40 [[FREE Full text](#)] [doi: [10.4324/9780080495811-6](https://doi.org/10.4324/9780080495811-6)]
10. Cavalcanti JC, Englert M, Oliveira M, Constantini AC. Microphone and audio compression effects on acoustic voice analysis: a pilot study. *J Voice* 2023;37(2):162-172. [doi: [10.1016/j.jvoice.2020.12.005](https://doi.org/10.1016/j.jvoice.2020.12.005)] [Medline: [33451892](https://pubmed.ncbi.nlm.nih.gov/33451892/)]
11. Luo D, Luo W, Yang R, Huang J. Identifying compression history of wave audio and its applications. *ACM Trans Multimedia Comput Commun Appl* 2014;10(3):1-19. [doi: [10.1145/2575978](https://doi.org/10.1145/2575978)]
12. Sayood K. *Introduction to Data Compression*, 5th Edition. Cambridge, United Kingdom: Morgan Kaufmann; 2018.
13. Perepelytsia V, Dellwo V. Acoustic compression in Zoom audio does not compromise voice recognition performance. *Sci Rep* 2023;13(1):18742 [[FREE Full text](#)] [doi: [10.1038/s41598-023-45971-x](https://doi.org/10.1038/s41598-023-45971-x)] [Medline: [37907749](https://pubmed.ncbi.nlm.nih.gov/37907749/)]
14. Awan SN, Shaikh MA, Awan JA, Abdalla I, Lim KO, Misono S. Smartphone recordings are comparable to 'gold standard' recordings for acoustic measurements of voice. *J Voice* 2023 Apr 03. [doi: [10.1016/j.jvoice.2023.01.031](https://doi.org/10.1016/j.jvoice.2023.01.031)] [Medline: [37019804](https://pubmed.ncbi.nlm.nih.gov/37019804/)]
15. Ireland D, Knuepffer C, McBride SJ. Adaptive multi-rate compression effects on vowel analysis. *Front Bioeng Biotechnol* 2015;3:118 [[FREE Full text](#)] [doi: [10.3389/fbioe.2015.00118](https://doi.org/10.3389/fbioe.2015.00118)] [Medline: [26347863](https://pubmed.ncbi.nlm.nih.gov/26347863/)]
16. Sáenz-Lechón N, Osmá-Ruiz V, Godino-Llorente JI, Blanco-Velasco M, Cruz-Roldán F, Arias-Londoño JD. Effects of audio compression in automatic detection of voice pathologies. *IEEE Trans Biomed Eng* 2008;55(12):2831-2835. [doi: [10.1109/TBME.2008.923769](https://doi.org/10.1109/TBME.2008.923769)] [Medline: [19126465](https://pubmed.ncbi.nlm.nih.gov/19126465/)]
17. Jadoul Y, Thompson B, de Boer B. Introducing parselmouth: a Python interface to Praat. *J Phon* 2018;71:1-15. [doi: [10.1016/j.wocn.2018.07.001](https://doi.org/10.1016/j.wocn.2018.07.001)]
18. Boersma P, Weenink DJM, Van Heuven V. PRAAT, a system for doing phonetics by computer *Speak and unSpeak with PRAAT*. *Glott International* 2001;5:341-345 [[FREE Full text](#)]
19. Picone JW. Signal modeling techniques in speech recognition. *Proc IEEE* 1993;81(9):1215-1247. [doi: [10.1109/5.237532](https://doi.org/10.1109/5.237532)]
20. Ancillao A, Galli M, Mignano M, Dellavalle R, Albertini G. Quantitative analysis of pathological female human voice by processing complete sentences recordings. *J Laryngol Voice* 2013;3(2):46. [doi: [10.4103/2230-9748.132045](https://doi.org/10.4103/2230-9748.132045)]

Abbreviations**FFmpeg:** Fast Forward Moving Picture Experts Group**MFCC:** Mel-frequency cepstral coefficient**MH:** MediaHuman**WS:** WonderShare

Edited by T Leung; submitted 10.01.24; peer-reviewed by V Perepelytsia, V Despotovic; comments to author 07.02.24; revised version received 28.02.24; accepted 23.03.24; published 15.04.24.

Please cite as:

Oreskovic J, Kaufman J, Fossat Y

Impact of Audio Data Compression on Feature Extraction for Vocal Biomarker Detection: Validation Study

JMIR Biomed Eng 2024;9:e56246

URL: <https://biomedeng.jmir.org/2024/1/e56246>

doi: [10.2196/56246](https://doi.org/10.2196/56246)

PMID: [38875677](https://pubmed.ncbi.nlm.nih.gov/38875677/)

©Jessica Oreskovic, Jaycee Kaufman, Yan Fossat. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 15.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Agreement Between Apple Watch and Actical Step Counts in a Community Setting: Cross-Sectional Investigation From the Framingham Heart Study

Nicole L Spartano^{1,2}, PhD; Yuankai Zhang³, MS; Chunyu Liu³, PhD; Ariel Chernofsky³, PhD; Honghuang Lin⁴, PhD; Ludovic Trinquart^{5,6}, MPH, PhD; Belinda Borrelli⁷, PhD; Chathurangi H Pathiravasan³, PhD; Vik Kheterpal⁸, MD; Christopher Nowak⁸, PhD; Ramachandran S Vasani^{2,9,10}, MD; Emelia J Benjamin^{2,11,12}, MD, ScM; David D McManus^{13,14}, MD, ScD; Joanne M Murabito^{2,15}, MD, ScM

¹Section of Endocrinology, Diabetes, Nutrition, and Weight Management, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

²Boston University's and National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, United States

³Boston University School of Public Health, Boston, MA, United States

⁴University of Massachusetts Chan Medical School, Worcester, MA, United States

⁵Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, United States

⁶Tufts Clinical and Translational Science Institute, Tufts University, Boston, MA, United States

⁷Boston University Henry M. Goldman School of Dental Medicine, Center for Behavioral Science Research, Boston, MA, United States

⁸Care Evolution, Ann Arbor, MI, United States

⁹Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

¹⁰University of Texas School of Public Health and University of Texas Health Sciences Center, San Antonio, TX, United States

¹¹Section of Cardiology, Department of Medicine, Boston Medical Center, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

¹²Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

¹³Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

¹⁴Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, Worcester, MA, United States

¹⁵Section of General Internal Medicine, Department of Medicine, Boston University Chobanian and Avedisian School of Medicine and Boston Medical Center, Boston, MA, United States

Corresponding Author:

Nicole L Spartano, PhD

Section of Endocrinology, Diabetes, Nutrition, and Weight Management

Boston University Chobanian and Avedisian School of Medicine

72 E Concord St, Suite 301 Collamore

Boston, MA, 02118-2371

United States

Phone: 1 3154152040

Email: spartano@bu.edu

Abstract

Background: Step counting is comparable among many research-grade and consumer-grade accelerometers in laboratory settings.

Objective: The purpose of this study was to compare the agreement between Actical and Apple Watch step-counting in a community setting.

Methods: Among Third Generation Framingham Heart Study participants (N=3486), we examined the agreement of step-counting between those who wore a consumer-grade accelerometer (Apple Watch Series 0) and a research-grade accelerometer (Actical) on the same days. Secondly, we examined the agreement during each hour when both devices were worn to account for differences in wear time between devices.

Results: We studied 523 participants (n=3223 person-days, mean age 51.7, SD 8.9 years; women: n=298, 57.0%). Between devices, we observed modest correlation (intraclass correlation [ICC] 0.56, 95% CI 0.54-0.59), poor continuous agreement (29.7%, n=957 of days having steps counts with $\leq 15\%$ difference), a mean difference of 499 steps per day higher count by Actical, and wide limits of agreement, roughly ± 9000 steps per day. However, devices showed stronger agreement in identifying who meets various steps per day thresholds (eg, at 8000 steps per day, kappa coefficient=0.49), for which devices were concordant for 74.8% (n=391) of participants. In secondary analyses, in the hours during which both devices were worn (n=456 participants, n=18,760 person-hours), the correlation was much stronger (ICC 0.86, 95% CI 0.85-0.86), but continuous agreement remained poor (27.3%, n=5115 of hours having step counts with $\leq 15\%$ difference) between devices and was slightly worse for those with mobility limitations or obesity.

Conclusions: Our investigation suggests poor overall agreement between steps counted by the Actical device and those counted by the Apple Watch device, with stronger agreement in discriminating who meets certain step thresholds. The impact of these challenges may be minimized if accelerometers are used by individuals to determine whether they are meeting physical activity guidelines or tracking step counts. It is also possible that some of the limitations of these older accelerometers may be improved in newer devices.

(JMIR Biomed Eng 2024;9:e54631) doi:[10.2196/54631](https://doi.org/10.2196/54631)

KEYWORDS

accelerometer; mobile health; mHealth; wearable device; fitness tracker; physical activity; mobile phone; Apple Watch; step counts; Framingham Heart Study; Actical; digital health; tracker; wearable; wearables

Introduction

Physical inactivity is an important risk factor for many chronic diseases including obesity, diabetes mellitus, hypertension, cardiovascular disease, and dementia [1]. The 2018 Physical Activity Guidelines for Americans recommend 150 minutes of moderate to vigorous physical activity (MVPA) or more per week [1]. Despite many known benefits of physical activity, many Americans do not meet the Physical Activity Guidelines, the proportion of Americans meeting these guidelines changes drastically depending on whether physical activity levels are measured using accelerometers or self-report. Guideline achievement has been estimated to be as low as 15% of Americans using accelerometry in a nationally representative sample, or as high as 66% using self-reported data in the same individuals [2,3]. Furthermore, experts have expressed concern over whether these guidelines are appropriate and attainable, especially in older adults or those with mobility limitations [1,4,5].

Walking is a central component of physical activity and public health promotion efforts [6]. Public health messages focused on daily step counts may be a more appropriate target for achieving recommended amounts of physical activity in adults [6], which might have even more significance in older populations and those who have low MVPA levels. We are in a new paradigm in health care, in which 69% of US adults report tracking at least 1 health metric [7,8], including millions of individuals who track their steps using wearable accelerometer devices that are available commercially [9]. Despite the longstanding use of step counting in public health interventions [10], the Physical Activity Guidelines Committee has not yet created recommendations for the number of daily steps to target as a goal for health promotion [1]. The primary reason for this lack of step count guidelines has been a lack of evidence, but meta-analyses conducted from large cohort studies have recently reported that step count is associated with a lower risk of death and chronic disease [11,12]. Many accelerometers and

pedometers have been validated to accurately count steps in the laboratory setting [13-15], but a remaining concern is that it is unclear how the number of steps reported in studies using research-grade accelerometers compares to steps counted by consumer-grade wearable devices used by the public living in the community (ie, the free-living setting).

During a recent Framingham Heart Study (FHS) exam cycle, physical activity was measured using both a consumer or mobile health device (Apple Watch) and a research-grade accelerometer (Actical) at the same time in the same individuals. The purpose of this investigation was to assess the agreement between Apple Watch and Actical-derived daily step count in free-living environments. We primarily assessed whether step count agreed when devices were worn on the same day, even if wear times differed, because we acknowledge that wear time and behavior may differ when participants wear different devices in the real world. We secondarily assessed whether agreement differed when devices were worn for the same hour block and whether agreement differed by age, sex, height, BMI, or those with mobility disabilities. This report will enable a better interpretation of the Apple Watch's daily step count for research studies and consumers using these devices.

Methods

Study Cohort

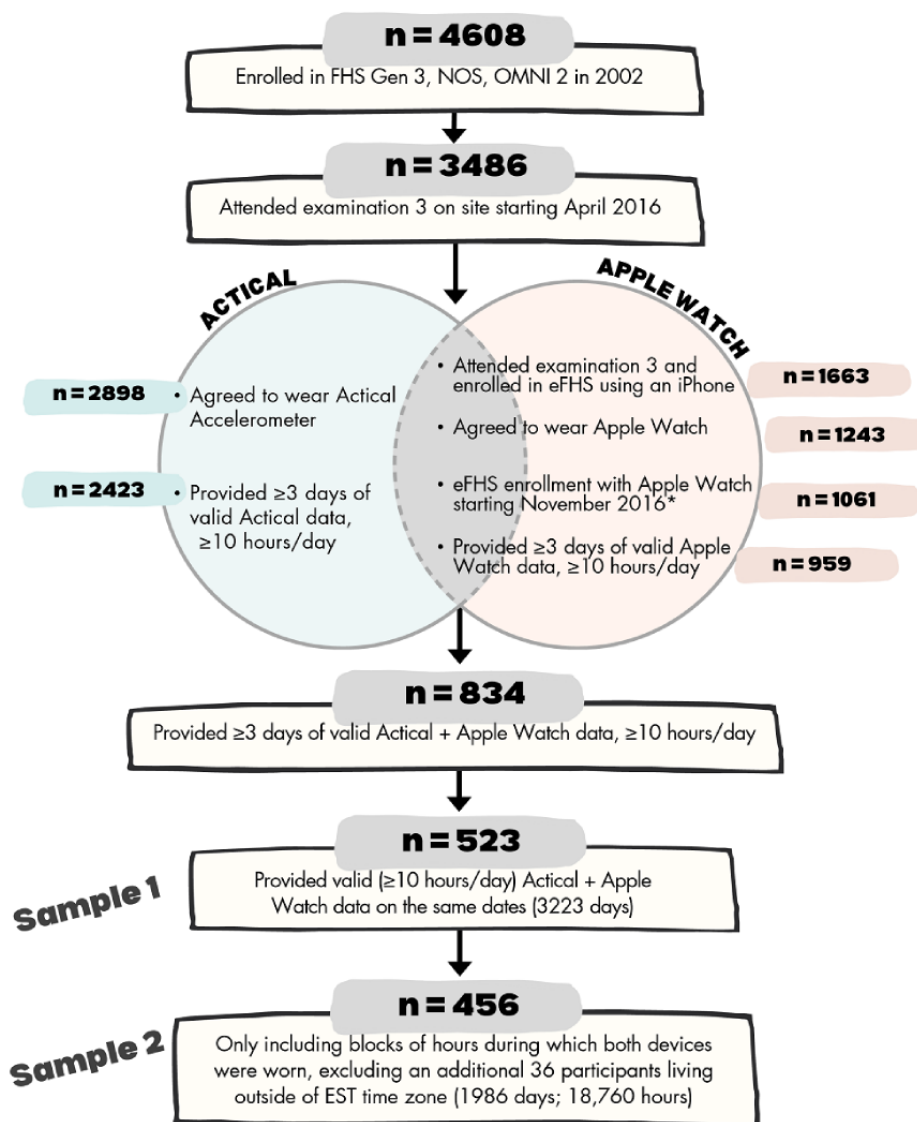
The FHS Third Generation-based (Gen 3) cohort was recruited in 2002-2005 (n=4095) [16], and consisted mostly of grandchildren of the Original FHS cohort [17], who were largely individuals of European descent. The Gen 3-based examinations also included spouses of the Original cohort's offspring (New Offspring Spouses [NOS], n=103) who were not already included in the Offspring (Generation 2) cohort and included a multiethnic Omni Group 2 (n=410). Participants from these cohorts have been examined every 6-8 years.

During the third in-person research examination of these cohorts (April 2016-March 2019), participants were asked to wear an

Actical accelerometer for 8 consecutive days on the hip. Beginning in November 2016, as part of the electronic FHS (eFHS) ancillary study [18], participants were also asked to wear an Apple Watch (Series 0) on their wrist for up to 1 year if they owned an iPhone with a compatible iOS (version 9 or higher). Of the 3486 FHS participants examined at the Research Center for exam 3, from April 2016 to March 2019, a total of 2898 (83%) agreed to take the Actical monitor, of which n=2423 (92% of those who took the device) had “valid” steps data, meaning they wore the monitor for at least 3 days, for at least 10 hours per day (Figure 1).

In total, 1061 eFHS enrollees (since November 2016) agreed to take the Apple Watch or use their own, of which 959 (90% of those who agreed to use an Apple Watch) wore the device for at least 3 days for at least 10 hours per day during the follow-up period. A total of 834 participants had at least 3 days of “valid” data from both devices (Actical and Apple Watch). Of those, 523 participants had at least 10 hours of wear time on both devices on the same day, providing a total of 3223 person-days for our primary study sample (sample 1).

Figure 1. Participant flow diagram for the analysis of agreement between Apple Watch and Actical step counts. eFHS: electronic Framingham Heart Study; EST: Eastern Standard Time; FHS: Framingham Heart Study; Gen: generation; NOS: New Offspring Spouse. *Enrollment in eFHS starting in November 2016 was necessary because this was the first date Apple Watches were given out at the FHS Research Center. Participants were able to enroll in eFHS prior to this, but they were given an Apple Watch to use later (starting in November), so their Apple Watch use would not align with the Actical monitor wearing dates.



Ethical Considerations

All participants provided written informed consent and the institutional review board at Boston University Medical Center approved the study protocols (H-32132).

Actical Physical Activity

During the 8-day wear period, participants were asked to remove the Actical accelerometer (Philips Respironics, model numbers 198-0302-xx, Respironics Co Inc) each night for sleep and when bathing or swimming. Actical data were recorded in 30-second epochs and expressed as counts (or steps) per 30 seconds. Actical step counting has been validated against hand counting

in a laboratory setting [19,20]. For sample 1, data were processed using a SAS program (SAS Institute) developed by Colley et al [21], and modified with input from collaborators [22], including nonwear time removal using the Choi algorithm [23], as explained in detail in Methods in [Multimedia Appendix 1](#) [23]. After processing, there remained 18 hours of possible wear time per day. A valid day was defined as ≥ 10 hours of wear time, with ≥ 3 days required for inclusion in the main analysis [24].

Apple Watch Series 0 Physical Activity

As part of the eFHS protocol, participants were asked to wear the smartwatch daily and were sent home with instructions on proper smartwatch use with advice to remove the smartwatch for charging every night. We also set up permissions for the Apple Watch app to access health information from other apps on the smartphone (ie, steps, heart rate, blood pressure, and weight) but we did not enter participant-specific data during Apple Watch setup. In contrast to data collected from the Actical, which had both counts and steps per 30-second interval, we were only able to collect Apple Watch data at the granularity of the number of steps per hour. For the Apple Watch, 1 wear hour was defined as an hour with at least 2 heart rates or at least 30 steps accumulated [25]. Unlike for Actical, there was no maximum number of wear hours chosen for the Apple Watch. A valid day was defined as ≥ 10 hours of wear time, with ≥ 3 days required for inclusion in the main analysis.

Covariates

The covariates that were measured during the examination when Actical and Apple Watch devices were provided to participants were current smoking status, self-reported health, BMI, hypertension stage II (systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg, or use of blood pressure medications) [26], diabetes mellitus (fasting plasma glucose ≥ 126 mg/dL and use of medications for diabetes mellitus), and prevalent cardiovascular disease. Depression status was defined as anyone with a score of 16 or greater on the Center for Epidemiological Studies Depression (CESD) scale. The physical activity index was a composite score constructed by weighing self-reported time spent in physical activity intensities over a 24-hour “typical” day [27]. Mobility limitation was defined as those self-reporting that they were unable to walk 0.5 miles without help or that they were limited a little or a lot when climbing several flights of stairs.

Statistical Analysis

After excluding participants who did not have at least 3 days of valid data from both devices and then excluding dates on which only 1 device was worn, we were left with 523 participants (3223 person-days, sample 1). We compared the number of hours participants wore each device on average days and average steps accumulated to determine device-specific differences, reporting means and SDs or medians and IQRs.

To examine the agreement between devices on days when both devices were worn for >10 hours (sample 1), we reported the intraclass correlation (ICC) using the random-effects model in our 2 study samples and used the Lin concordance coefficient (accounting for repeated observations). We also used kappa

coefficients to assess concordance between the devices in identifying participants meeting thresholds of average daily steps (at 3000, 6000, 8000, and 10,000 steps per day). Bland-Altman plots were also used to assess potential non-systematic differences between devices and provide a visual representation of these differences in steps per day and the percent differences (100 multiplied by [Apple Watch mean minus Actical steps] divided by mean steps). We assessed the limits of agreement for the Bland-Altman plot using repeated measures. Agreement of Apple Watch and Actical step counts per day was also assessed as the percent of days in which steps for each device fell within 15% agreement of one another. In personal communication with physical activity research experts (unpublished), most suggested that acceptable agreement should be set at a 5% difference level (Tudor-Locke et al [28]), with a few experts acknowledging that agreement within 15% may be considered acceptable (Breteler et al [14]). Experts polled were those who participated as an author in the meta-analysis of 15 international cohorts with accelerometer data published by Paluch et al [11]. We chose to report the more lenient agreement threshold in order to better detect variability in agreement among subsamples of our population, especially after observing the poor overall agreement within these ranges displayed in the results.

In secondary analyses, we also examined agreement between devices during hours when both devices were worn (sample 2) to account for potential differences in wear periods (by device) throughout the day. To create this sample, first, we identified blocks of time ≥ 3 hours each day (midnight to midnight) during which both devices were worn. We defined an hour of Actical wear as any hour with >0 step count. In this study, we defined an hour of Apple Watch wear as an hour with >30 step counts or 0-30 step counts with at least 2 heart rates recorded, but there does not seem to be an established threshold used in this research field. We excluded hours for which step counts were missing (shown as NA in Figure S1 in [Multimedia Appendix 1](#)). These definitions differed because of different device-wearing locations (hip vs wrist). When devices are worn on the hip, they can show 0 step counts for prolonged periods of time when a participant is wearing the device sitting, but this is less likely to occur with a wrist-worn device. A total of 30 participants had <3 hours of overlapping wear time and were excluded (Figure 1). These 30 participants had >10 hours of Apple Watch wear time on days when the Actical was worn for >10 hours, but the Apple Watch wear hours did not have at least 3 consecutive hours. When steps were counted, each hour or 2, they were broken up by hours with heart rate measurements, but they were often missing step counts. An example of 24 hours of Actical and Apple Watch data is shown in Figure S1 in [Multimedia Appendix 1](#). We provide further interpretation of these “interruptions” in wear time in the discussion section. Our next step was to remove the first and last hour of each ≥ 3 -hour block because we could not determine whether they were full or partial hours. The remaining hours in that block were each used as separate data points, to provide us with steps accumulated by the 2 devices for every hour that both devices were worn. As Apple Watch (but not Actical), changes time stamps during the collection period to be consistent as people move across different time zones, we additionally excluded

participants residing outside of the Eastern Standard Time Zone ($n=36$), which may have resulted in discordant hours being counted by each device. One extreme outlier (1 person-hour) was also removed (see Figure S2 in [Multimedia Appendix 1](#)), which did not affect results (data not shown). We repeated the analysis from sample 1.

Next, for each sample, we tested for interactions by age, sex, height, and BMI in the linear regression analysis to assess whether these factors influenced agreement between the Apple Watch and Actical device measures of total daily steps. Finally, we performed sensitivity analyses, repeating our agreement analysis in subsamples excluding participants with high or low step counts. All statistical analyses were performed with R (version 4.1.3; R Core Team), including ggplot2 (for plots), irr (for ICC), epiR (for Lin concordance correlation), and psych (for kappa coefficients) packages.

Results

Overview

Compared to the total FHS Gen 3, NOS, and Omni 2 cohort, participants who returned valid (ie, sufficient) data from the 2 wearable devices were on average younger, healthier (less smoking, diabetes, hypertension, cardiovascular disease, and depression), and were more likely to have completed college or received a graduate degree ([Table 1](#)). The average wear time for the Apple Watch was more than an hour longer each day than for the Actical (15.6 vs 14.4 hours; [Table 1](#)), which may be partially due to the removal of 6 hours of each 24 hours and other Actical data processing, as described in Methods in [Multimedia Appendix 1](#).

Table 1. Characteristics for all FHS^a Gen^b 3 participants who attended examination 3, compared to those with valid Actical and Apple Watch data on the same date.

	FHS Gen 3 (n=3521)	FHS Gen 3 with valid Actical +Apple Watch data on the same date (sample 1, n=523)
Age (years), mean (SD)	54.5 (9.4)	51.7 (8.9)
Women, n (%)	1896 (53.9)	298 (57.0)
Race and ethnicity, n (%)		
Non-Hispanic White	3233 (91.8)	478 (91.4)
Non-Hispanic Black	59 (1.7)	12 (2.3)
Hispanic or Latino	106 (3.0)	14 (2.7)
Asian	71 (2.0)	9 (1.7)
American Indian	1 (0.03)	1 (0.2)
Pacific Islander	2 (0.06)	0 (0)
More than 1 race	41 (1.2)	8 (1.5)
Unknown	8 (0.2)	1 (0.2)
BMI (kg/m ²), mean (SD)	28.6 (6.2)	28.2 (5.7)
Height (inches), mean (SD)	66.6 (3.7)	66.8 (3.6)
Mobility limitation, n (%)	703 (20)	85 (16.3)
Smoking, n (%)	234 (6.7)	27 (5.2)
Education, n (%)		
Less than HS ^c	48 (1.4)	3 (0.6)
Completed HS	470 (13.5)	43 (8.2)
Some college	489 (14)	114 (21.8)
Bachelor's degree	1222 (35.0)	214 (41.0)
Graduate or professional degree	843 (24.1)	148 (28.4)
Married, living as married, living with partner, n (%)	2454 (70.5)	397 (76.4)
Employed full-time, n (%)	2277 (65.4)	381 (73.1)
Self-reported health (excellent), n (%)	750 (21.4)	128 (24.5)
Diabetes mellitus, n (%)	310 (8.8)	26 (5.0)
Hypertension stage II, n (%)	1095 (31.1)	112 (21.4)
Cardiovascular disease, n (%)	164 (4.7)	18 (3.4)
Depression (CESD ^d >16), n (%)	449 (12.8)	55 (10.5)
Physical activity index (score), mean (SD)	33.9 (5.7)	33.2 (4.7)
Actical steps, median (IQR)	N/A ^e	7064 (4638-10,529)
Apple Watch steps, median (IQR)	N/A	7060 (4450-10,348)
Actical wear time (hours), mean (SD)	N/A	14.4 (1.8)
Apple Watch wear time (hours), mean (SD)	N/A	15.6 (2.6)

^aFHS: Framingham Heart Study.^bGen: generation.^cHS: high school.^dCESD: Center for Epidemiological Studies Depression.^eN/A: not applicable.

Primary Analysis (Sample 1, n=523): Step Agreement Per Day of Device Wear

We observed a modest correlation (ICC 0.56, 95% CI 0.54-0.59; Table 2), but poor agreement (29.7%, n=957 of days having steps counts with $\leq 15\%$ difference) between devices. Lin concordance coefficient, accounting for repeated observations, produced the same coefficients as traditional ICC for all results. The 2 devices demonstrated moderate agreement for distinguishing between participants meeting versus not meeting step per day thresholds by their average daily steps (kappa coefficient ≈ 0.5 ; Table 3). The Apple Watch and Actical devices were concordant 74.8% (n=391)-84.5% (n=442) of the time, depending on the threshold (3000, 6000, 8000, and 10,000 steps per day). This reliability for distinguishing between thresholds did not change greatly if we used average daily steps (as in Table 3) or steps per person-day (as in Table S1 in Multimedia Appendix 1), but improved slightly to 77.2% (n=889) to 85.3% (n=982) if we excluded person-days in which wear time was >1 hour different between.

On average, we observed more steps per day counted by the Actical device, with a mean difference of 499 more steps per day counter compared to the Apple Watch (Figure 2, Table 2). Limits of agreement were -9543, 8544 steps per day, meaning

that differences in step counting between devices are expected to be roughly ± 9000 steps in a given day of device wear. The differences in step counting between devices tended to increase with higher average steps counted, but the percent differences did not (average limits of agreement were -134.6% to 118.2% difference between step counts). There also did not appear to be a major under- or overestimation of steps by 1 device compared to the other. We observed an interaction (Table S2 in Multimedia Appendix 1; $P < .001$) between wear time and device type in their association with daily step count.

Each point represents data from 1 participant on a single date (1 person-day). In the scatterplot, dashed lines are set at 1000 and 30,000 step thresholds. Days on which participants accumulated 1000-30,000 steps are dark green and days outside that threshold are presented in light green. Sections separated by the dashed lines include the following number of person-days divided by participants: A=17/5, B=205/68, C=2963/512, D=3/2, E=4/4, and F=31/29. The Bland-Altman plots on the right show the mean difference or mean % difference (red dashed line) and the limits of agreement 95% CI (blue dashed lines). The mean % difference was calculated as 100 multiplied by (Apple Watch steps minus Actical steps) divided by (average Apple Watch and Actical steps).

Table 2. Agreement between steps accumulated on Actical versus Apple Watch device by participants wearing both devices on the same date^a.

Sample and sample description	Adjusted linear regression, β (95% CI)	ICC ^b (95% CI)	Lin concordance correlation, r (95% CI)	Mean difference ^c (Bland-Altman limits of agreement)	Mean % difference ^d (Bland-Altman limits of agreement)	Percent of Apple Watch days with a step count within 15% agreement compared to Actical, n (%)
Sample 1 (n=523 participants; n=3223 person-days): includes all days when both devices were worn for >10 hours	0.67 (0.65-0.70)	0.56 (0.54-0.59)	0.56 (0.54-0.58)	-499 (-9543 to 8544)	-8.2 (-134.6 to 118.2)	957 (29.7)
Sample 2 (n=456 participants; n=1986 person-days; n=18,760 person-hours): only includes blocks of hours during which both devices were worn^e	0.97 (0.96-0.97)	0.86 (0.85-0.86)	0.86 (0.85-0.86)	20 (-844 to 884)	16.6 (-98.0 to 131.3)	5115 (27.3)
Sample 2A (n=151 participants; n=5397 person-hours): with obesity	0.94 (0.92-0.95)	0.85 (0.84-0.86)	0.85 (0.84-0.85)	33 (-844 to 909)	18.2 (-95.5 to 131.9)	1397 (25.9)
Sample 2B (n=79 participants; n=2967 person-hours): with mobility limits	0.86 (0.84-0.88)	0.86 (0.83-0.88)	0.86 (0.85-0.87)	98 (-953 to 1148)	29.8 (-83.8 to 143.5)	709 (23.9)
Sample 2C (n=266 participants; n=11699 person-hours): without obesity or mobility limitations	0.98 (0.97-0.99)	0.85 (0.85-0.86)	0.85 (0.85-0.86)	3 (-1089 to 1096)	14.4 (-101.4 to 130.3)	3275 (28.0)

^aThe adjusted linear regression model includes age, sex, cohort type, BMI, height, (and the difference in wear time for sample 1).

^bICC: intraclass correlation.

^cMean difference was Apple Watch steps minus Actical steps

^dMean % difference was 100 multiplied by (Apple Watch steps minus Actical steps) divided by (average Apple Watch and Actical steps).

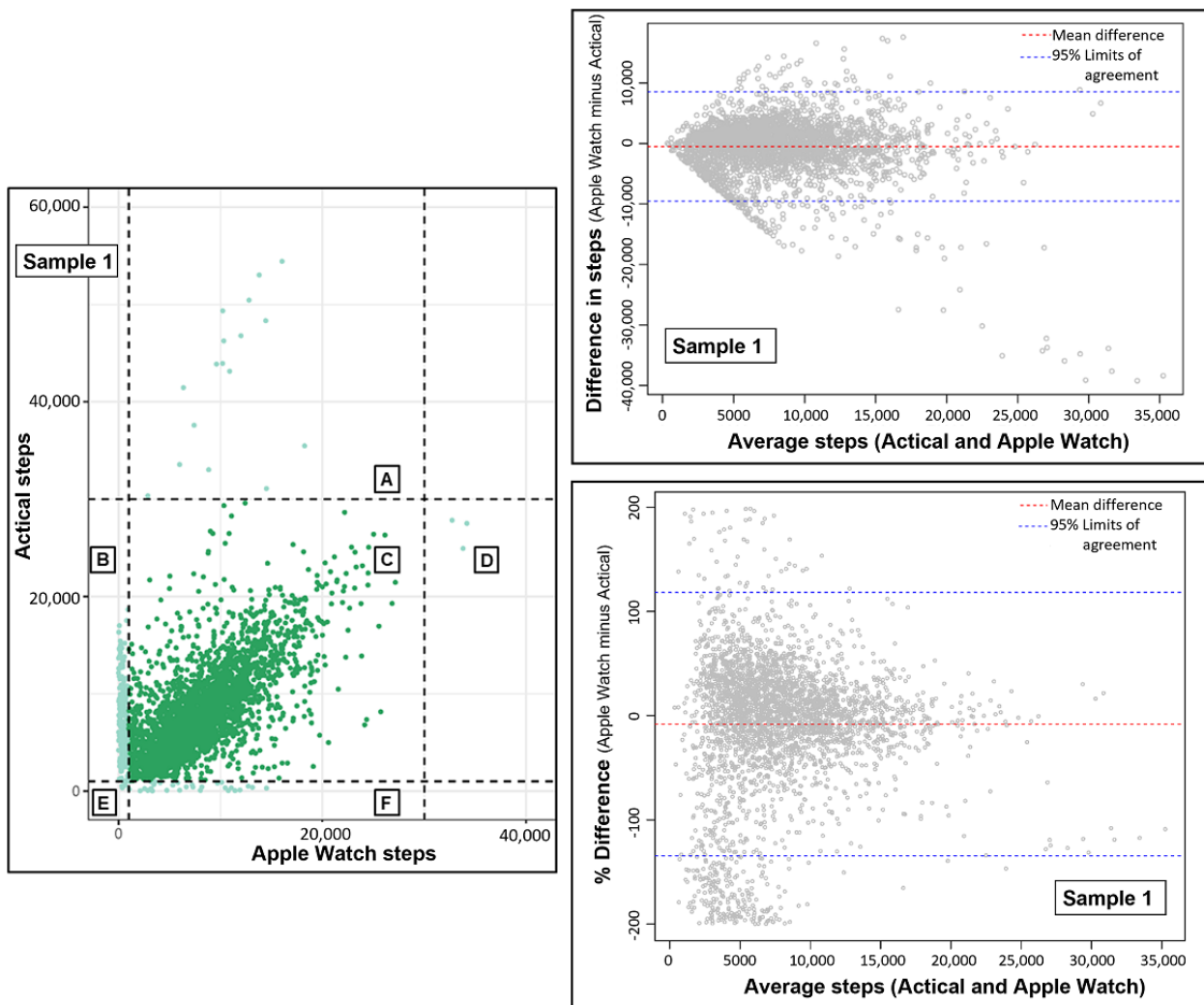
^eSample 2: after removing hours when both devices were not being worn, we removed the first and last hours of remaining blocks of hours. We additionally excluded participants who lived outside Eastern Standard Time Zone and removed 1 data point that was an extreme outlier (Figure S1 in [Multimedia Appendix 1](#)). We used each remaining hour as a separate data point.

Table 3. Agreement of Actical and Apple Watch devices to identify participants meeting average daily step thresholds (sample 1, n=523 participants; 3223 person-days).

Step per day threshold	Percent concordance for "meets the PA ^a threshold" as measured by the 2 devices, n (%)	Kappa coefficients (95% CI) for "meets the PA threshold" as measured by the 2 devices
3000 steps per day	442 (84.5)	0.12 (0.01-0.22)
6000 steps per day	396 (75.7)	0.46 (0.38-0.54)
8000 steps per day	391 (74.8)	0.49 (0.41-0.56)
10,000 steps per day	426 (81.5)	0.49 (0.40-0.58)

^aPA: physical activity.

Figure 2. Scatterplot and Bland-Altman plots (difference and % difference) of Apple Watch steps by Actical steps accumulated on the same date (sample 1, all data, 3223 person-days, 523 participants).



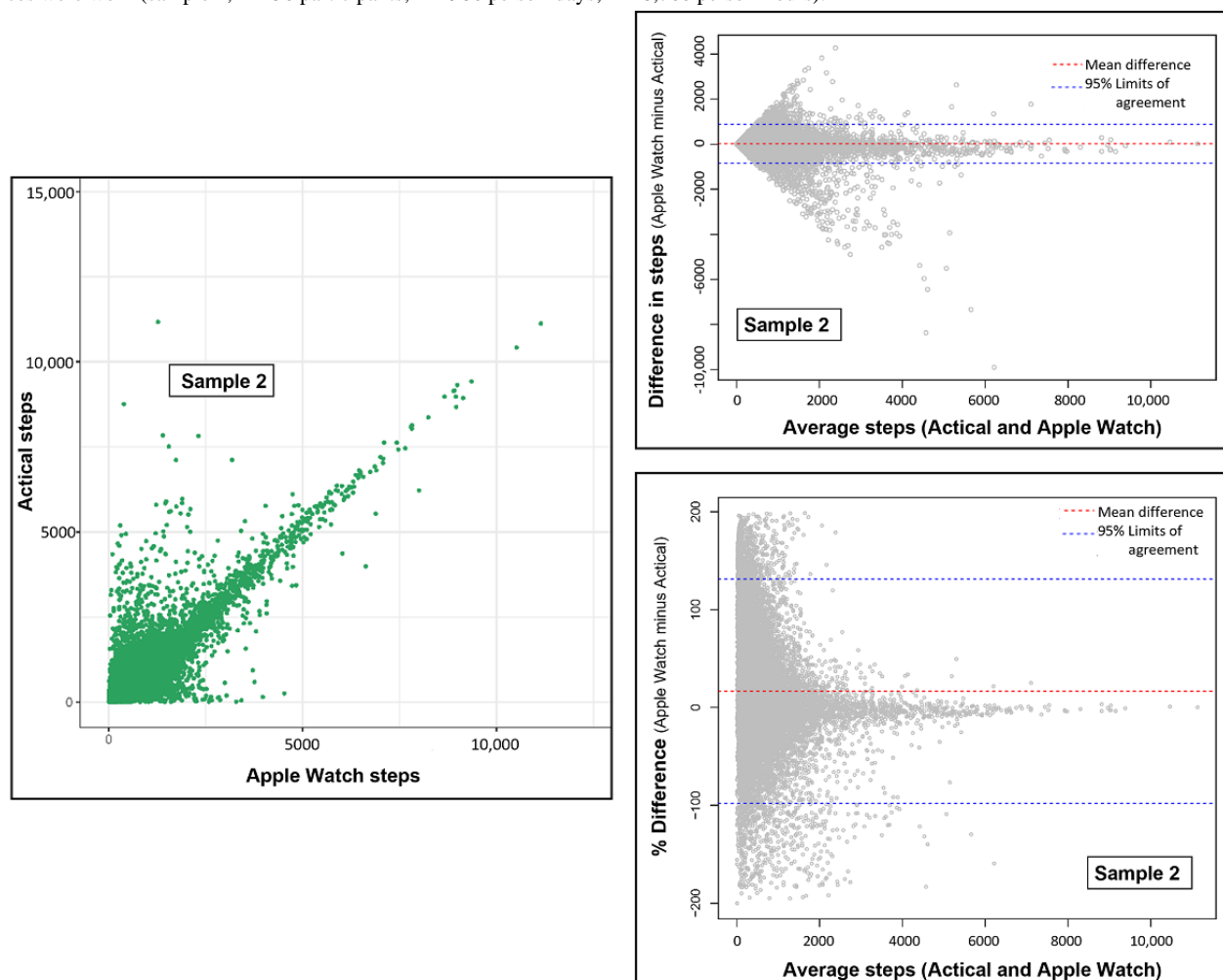
Secondary Analysis (Sample 2, n=456): Step Agreement Per Hour of Device Wear

We conducted secondary analyses to explore the agreement between devices, with differences in wear time minimized. We assessed agreement between devices for each hour during which both devices were worn (456 participants, 1986 person-days, 18,760 person-hours, Table 2 and Figure 3). Among hours when both devices were worn, the correlation of absolute step counts between devices was much stronger (ICC 0.86, 95% CI 0.85-0.86, Table 2) than it was for sample 1, but the agreement of steps counted per hour was still poor (only 27.3%, n=5115 of hours having step counts with $\leq 15\%$ difference) between devices. The mean difference in step count between devices was only 20 steps per hour, but limits of agreement were large (-844, 884 steps per hour) and a 16.6% difference (-98, 131.3% limits of agreement) between Apple Watch and Actical step counting on hours when both devices were worn.

Each point represents data from a single hour (1 person-hour). The Bland-Altman plots on the right show the mean difference or mean % difference (red dashed line) and the limits of agreement 95% CI (blue dashed lines). The mean % difference was calculated as 100 multiplied by (Apple Watch steps minus Actical steps) divided by (average Apple Watch and Actical steps).

Next, we assessed potential interactions (in sample 2, Table S2 in Multimedia Appendix 1), observing interactions by obesity status and mobility status ($P < .001$). We observed that correlations were similar regardless of these factors (samples 2A-2C, Table 2), but the agreement of step counts with $\leq 15\%$ difference between devices was slightly worse for participants with obesity (n=1397, 25.9% agreement) or self-reported mobility limitations (n=709, 23.9% agreement), compared to those with neither (n=3275, 28.0% agreement).

Figure 3. Scatterplot and Bland-Altman plot (difference and % difference) of Apple Watch steps by Actical steps accumulated during hours when both devices were worn (sample 2, n=456 participants; n=1986 person-days; n=18,760 person-hours).



Sensitivity Analyses: Exploration of Days With Low Step Counts and Large Differences in Step Count

Despite a strong correlation in step counts, there was substantial variability between devices in terms of device agreement, as demonstrated in the total sample 1, Figure 2. We observed 17 person-days with >30,000 steps per day by Actical but <20,000 steps per day by Apple Watch, representing days from 5 participants (section A in Figure 2, Figure S3 and Table S3 in Multimedia Appendix 1). In sensitivity analyses excluding data from these 5 participants, the ICC improved slightly for samples 1 and 2, but the percent of days or hours during which the devices agreed within 15% only improved by <1% (Table S4 in Multimedia Appendix 1). Very few days (~10%) met the stricter 5% threshold for agreement between devices and agreement was further reduced when only observing days when the Apple Watch was worn 5-10 hours and Actical was worn >10 hours.

Other substantial variability we observed between device counting by Actical compared to Apple Watch was observed in a large number of days during which 1 device counted <1000 steps and the other device counted >1000 steps (sections B and F in Figure 2). In Figures S4 and S5 in Multimedia Appendix 1, we show scatterplots for hours when both devices were worn. During most hours represented in sections B and F in Figure 2,

the devices were either being worn at different times of the day or there were interruptions in step counting. Furthermore, when observing hours from “other days” of those participants who had days that fell into sections B or F (Figures S4 and S5 in Multimedia Appendix 1), the pattern appears similar to the scatterplot for the overall sample 2 (Figure 3). Exclusion of participants who contributed days that fell into sections A, B, D, E, or F in Figure 2 (those with step counts <1000 or >30,000 by either device) did not improve agreement results greatly either, improving the percent of days on which the devices agreed within 15% only up to 32.2% of all days (Table S4 in Multimedia Appendix 1).

Discussion

Principal Findings

Consumer accelerometer devices are being used by millions of people to track their physical activity levels and progress toward public health recommendations or personal goals. These devices have been validated in laboratory settings against research-grade devices, but few studies have explored how consumer and research-grade accelerometer step counting compares when participants are living out in the community.

In this study, we observed poor overall agreement between steps counted by Actical and Apple Watch (Series 0) devices. Larger between-device differences were seen when the step count was higher, but the percent difference did not increase. However, our results suggest that we can expect the 2 devices to classify individuals into the same step thresholds about 75%-85% of the time. Results such as these may be important to consider when translating future step guidelines to the public using consumer brand devices. The limitations in agreement among accelerometer devices may be less important when they are used by individuals to determine the achievement of a recommended number of steps or for the purposes of tracking their step count over time.

The agreement we observed in this study in a free-living environment was worse than previous laboratory-based studies of consumer-grade devices comparing them to hand-counted steps or research accelerometer devices [13,15]. However, 1 study observed that even when testing the consistency of step counting in the same device, wearing the device at different locations (wrist vs hip) can result in inconsistencies in device step counting [29]. The difference in device location alone may have contributed greatly to the poor agreement of step counting between devices in our study.

Similar to our design, 1 study by Breteler et al [14], examined Apple Watch step counting in a free-living setting (wrist-worn) in comparison to other accelerometers worn on the hip. In this study, 30 healthy participants (mean age 40 years) wearing multiple devices over a 3-day period observed a median absolute relative difference of 7.7% comparing the Apple Watch to the ActiGraph (similar to our mean relative difference results comparing Apple Watch to Actical). However, they did not report the limits of agreement for this relative difference. Other devices they tested had a median absolute relative difference >15% [14]. A low mean or median relative difference indicates low bias (lack of systematic over- or undercounting by 1 device), but only limits of agreement can inform about the precision of agreement. Breteler et al [14] reported the mean difference in step counting was 968 more steps per day counted by the Apple Watch with limits of agreement ± 6000 steps per day (compared to ActiGraph), which is almost as high as the limits of the agreement we observed in our study sample 1 (compared to Actical). Investigators in that study observed that Apple Watch devices added steps overnight when other devices were not counting any steps, which could have been due to the delayed transmission of step count data. We did not observe the same phenomenon in our data, which may be due to us using an older Apple Watch device model. In our study, we observed that some participants had long periods of consecutive hours with heart rate data, but 0 step counts (meaning the Apple Watch device was being worn) and had long periods of consecutive Actical step data >0 during this time. We suspect that the Apple Watch step data were either not being recorded or transmitted during these time periods or were delayed by many hours. In order for Apple Watch Series 0 data to be recorded or transmitted, the participant's smartphone needed to be charged, connected to the internet, and unlocked. This finding has important implications for future research teams when analyzing data from other mobile health devices.

Although we observed poor overall agreement (due to wide limits of agreement) in our study and in Breteler et al [14] we also reported low bias due to low mean difference and percent difference. However, individual differences in gait, which may be in part due to older age, mobility limitations, or body stature (influenced by sex, height, and body composition), might introduce systematic bias into the measurement of steps in the community and should be considered in future studies [30-32]. Accelerometers have different sensitivities to slow gait speeds or low-frequency movement [29], even when tested in a laboratory in which gait differences are minimized. We observed slightly worse agreement between devices for individuals who were obese or self-reported mobility limitations. An individual's usual cadence and the amount of time they spend participating in movement activities other than ambulation (such as household chores or other multidirectional activities) may also influence step detection in certain accelerometer devices [33-35]. Although individuals with mobility limitations and other conditions that alter gait (eg, obesity) only worsened agreement slightly, the overall influence of gait on how these devices count steps may partially explain the poor agreement between devices.

Agreement of step-counting devices has implications for future updates of the Physical Activity Guidelines. Advancements in technology and the widespread availability of consumer wearable devices make physical activity monitoring feasible in research or clinical settings and for individuals in the community. During the development of the 2018 Physical Activity Guidelines, it was determined that there was insufficient evidence to create a guideline for health promotion based on step count [1]. However, an estimation by Tudor-Locke et al [36] suggests that the MVPA guidelines can be met by adults who walked a minimum of ~7000-8000 steps per day. Furthermore, a 2022 meta-analysis of 15 observational cohort studies (including FHS) using research-grade physical activity monitors (eg, ActiGraph, Actical, and Activ-PAL), reported that individuals achieving ≥ 8000 (vs < 8000) steps per day in middle age had the lowest risk of mortality [11]. In older adults, a lower threshold of ≥ 6000 steps per day was associated with almost 50% lower risk of death compared to older adults who walked less. The study, which was the largest meta-analysis of its kind, may serve as evidence to support future guidelines as to the number of steps adults should walk each day.

Although we now have some evidence that achieving step thresholds from 6000 to 8000 steps per day is associated with lower mortality [11], creating step guidelines is complicated by the observation that individuals in the community do not use the same research-grade devices as used in many prior studies. Instead, the public uses popular consumer activity trackers, such as Fitbit (Google), Apple Watch, and Garmin among other devices. Although these consumer devices have been well validated for the measurement of steps in laboratory settings [13-15], it has not been clear whether the steps counted by these consumer devices are comparable to steps counted by research-grade devices in free-living settings [37]. Unfortunately, it does not appear to be a simple fix to "convert" steps measured by a research device to those measured by a consumer device, based on the wide limits of agreement. Despite the poor overall agreement of step counting between devices,

favorably, the devices had a substantially better agreement in identifying who meets thresholds between 6000 and 10,000 steps per day, with an agreement for ~75%-82% of individuals. These thresholds may serve as targets for future public health recommendations.

Strengths and Limitations

Strengths of our investigation include the large sample size and the study being conducted in a community setting, which increases the generalizability of the findings. However, the homogeneous nature of our study cohort, who were mostly from 1 geographic location, were generally healthier and more highly educated than the general US population and were mostly of European ancestry, may limit generalizability to more diverse populations. Another strength was our use of different person-day samples to examine different questions such as comparing step counts between devices when worn for a comparable number of hours and observing the influence of different wearing behaviors on step count agreement. The lack of control of wear time and differences in wear time observed between the devices can be viewed as both a weakness (because wear time affects step accumulation) and strength (in that it demonstrates the differences that may be inherent in real-world device use). Similarly, as noted earlier, another difference between these devices was their placement on the wrist versus the hip, which may have also contributed to the variation. However, the device placement locations add another real-world element to our study design.

Wear time per day was longer, on average, for the Apple Watch, which may, in part, be due to wearing during sleeping hours. In our analysis, sleeping hours were removed from the Actical total wear time, but not from the Apple Watch. We asked participants to remove the Actical device when they bathed, swam, or slept. These instructions were not given to participants for the Apple Watch, although participants may have chosen to do so. The Apple Watch is waterproof, but the battery does not typically last much longer than 24 hours, so most participants likely took off the Apple Watch to charge at night. If the Apple Watch battery was not charged, a participant might not have worn the device and may have missed opportunities to record steps walked. The Actical battery did not need to be charged during the week that participants wore the device, which may have affected when it was worn compared to the Apple Watch. On the other hand, the Actical device was worn on a belt around the waist or hip, compared to the Apple Watch, worn on the wrist, either of which can be cumbersome, causing some participants to remove the device or wear it improperly (eg, loosely). It is unclear which placement site is preferred by the research community [38]. Although we sent participants home with instructions for when to take on and off the devices and the location where they should be worn, we did not emphasize that they should ensure a snug fit. Another possibility is that there could be calibration issues with some of the devices (Actical and Apple Watch could have drifted from factory calibration). A comparison of agreement results between samples 1 and 2 makes it clear that it is unlikely that the poor agreement was explained by participants wearing the devices during different times of day or activities. But it is also evident that some of the differences in step counting by these devices

may have been due to Bluetooth connectivity errors in the recording or transmission of step data, which led to very low steps counted by the Apple Watch.

Features of physical activity monitors are also important considerations. The Apple Watch device used in this investigation has many applications, including allowing participants to see step counts as they were accumulated (there was no visual display on the Actical device) and other functionalities. The availability of these features may also influence when the device is worn and how many steps are taken. We did observe that of the participants who agreed to either device, a roughly equal proportion of participants (~90%) wore those devices for ≥ 3 days for at least 10 hours per day. However, studies have shown that features such as a display showing step progress and encouragement (ie, nudges) to stand or move may increase both wearing and stepping behavior, especially over the short term, which may influence results from studies using consumer devices that tend to have these features.

Our study provided us with many lessons that we hope to communicate with investigators using accelerometers. An unexpected finding was that the agreement between these physical activity monitors only improved slightly after we limited differences in wear time between the devices. When experts develop public health guidelines for the number of steps to walk each day, they must consider that devices do not all record steps equivalently and that the type of device, wearing location, or mode (ie, watch, belt, or smartphone app), battery life, Bluetooth connectivity issues, other features of the device, and gait differences of participants may all influence when the device is worn and how many steps are counted. Moreover, we did not enter participant-specific data (ie, height, weight, age, and sex) when setting up the Apple Watch or Actical devices. However, the Apple Watch may have accessed this type of data from other health-related apps on a participant's smartphone. It is also important to note that we studied older versions of the devices, both of which are no longer supported by their manufacturers. Hopefully, newer device models may have overcome some of the limitations of the accelerometers we studied; we used Apple Watch Series 0 during data collection, but they have already transitioned to Series 8. In future research, it may also be important to emphasize proper wear of devices and input relevant participant-specific information during device setup for improved precision.

Conclusions

Our investigation suggests that overall agreement between steps counted by the Actical and Apple Watch Series 0 devices was poor, but agreement between devices was much stronger for distinguishing who meets certain step thresholds. Many large cohort studies have used the Actical device and other research and consumer devices to observe thresholds of physical activity (steps per day) that are associated with health outcomes [11,39-41]. Lessons learned from our investigation should be considered when translating thresholds of steps counted using the Actical to guidelines for members of the community using consumer devices, including the Apple Watch. Future studies should explore the agreement among other devices in the community setting and explore the role of interruptions in

connectivity, calibration, and factors affecting gait, such as age, sex, frailty or mobility status, BMI, and height on the accuracy of step count and agreement among devices. However, another important future direction should be the increased use of consumer accelerometer devices in research in order to replicate recent meta-analyses reporting the higher risk of mortality among physically inactive individuals (measured using research-grade devices) [11,42]. Studies such as All of Us and the Risk Underlying Rural Areas Longitudinal (RURAL) Heart and Lung Study that use Fitbit devices, for example, will be

extremely useful in the development and translation of future physical activity step guidelines [41].

The good news is that the impact of these challenges in measuring steps may be minimized when accelerometers are used by individuals for the purposes of tracking the changes in their physical activity over time, which eliminates the impact of gait differences (unless gait changes), factory calibration issues (if the same device is used), and presumably connectivity issues would remain consistent, limiting their impact too.

Acknowledgments

This study was supported by an award from the Robert Wood Johnson Foundation (74624) and grants from the National Heart Lung and Blood Institute (R01HL141434 and R01HL131029), National Institute on Aging (R01AG047645), and American Heart Association (15GPSGC24800006). The Framingham Heart Study was supported by a contract from the National Heart Lung and Blood Institute (principal investigator RSV 75N92019D00031); and investigator time from the grants (R01HL126911, R01HL092577, R01AG066010, and U54HL120163; EJB), (R01HL155343, R01HL141434, R33HL158541, U54HL143541, U54HL143541-05S1, and UG3NS135168; DDM), American Heart Association (18SFRN34110082; EJB). RSV was supported in part by the Evans Medical Foundation and the Jay and Louis Coffman Endowment from the Department of Medicine, Boston University Chobanian and Avedisian School of Medicine. The Apple Watches were provided to Boston University by Apple Inc at no cost to the study. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request. The Framingham Heart Study makes data available through the online repositories BioLINCC and dbGap.

Authors' Contributions

NLS, YZ, JMM, CL, and LT conceptualized the study design. YZ and AC contributed to statistical analysis. NLS drafted the paper. All authors interpreted results, edited the paper, and agreed to the final version.

Conflicts of Interest

Apple was not involved in the study design, analysis, interpretation, or reporting of study results. DDM has received research support from Fitbit, Apple Inc, Bristol–Myers Squibb, Boehringer–Ingelheim, Pfizer, Flexcon, Samsung, Philips Healthcare, and Biotronik, and he has received consultancy fees from Heart Rhythm Society, Bristol–Myers Squibb, Pfizer, Fitbit, Flexcon, Boston Biomedical Associates, VentureWell, Avania, NAMSA and Rose Consulting. DDM also declares financial support for serving on the Steering Committee for the GUARD-AF study (NCT04126486) and the advisory committee for the Fitbit Heart study (NCT04176926). VK is a principal, and CN is an employee of CareEvolution, Inc, a health care technology company. NLS received funding from Novo Nordisk for an investigator-initiated research grant unrelated to this paper. JMM received funding as a guest lecturer for Merck unrelated to this work. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1

Additional methods and results.

[[DOCX File, 206 KB - biomedeng_v9i1e54631_appl.docx](#)]

References

1. Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services; 2018.
2. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the physical activity guidelines for Americans. *Am J Prev Med* 2011;40(4):454-461. [doi: [10.1016/j.amepre.2010.12.016](https://doi.org/10.1016/j.amepre.2010.12.016)] [Medline: [21406280](https://pubmed.ncbi.nlm.nih.gov/21406280/)]
3. Zenko Z, Willis EA, White DA. Proportion of adults meeting the 2018 physical activity guidelines for Americans according to accelerometers. *Front Public Health* 2019;7:135 [FREE Full text] [doi: [10.3389/fpubh.2019.00135](https://doi.org/10.3389/fpubh.2019.00135)] [Medline: [31231627](https://pubmed.ncbi.nlm.nih.gov/31231627/)]
4. Lorbergs AL, Prorok JC, Holroyd-Leduc J, Bouchard DR, Giguere A, Gramlich L, et al. Nutrition and physical activity clinical practice guidelines for older adults living with frailty. *J Frailty Aging* 2022;11(1):3-11. [doi: [10.14283/jfa.2021.51](https://doi.org/10.14283/jfa.2021.51)] [Medline: [35122084](https://pubmed.ncbi.nlm.nih.gov/35122084/)]

5. Fanning J, Nicklas BJ, Rejeski WJ. Intervening on physical activity and sedentary behavior in older adults. *Exp Gerontol* 2022;157:111634. [doi: [10.1016/j.exger.2021.111634](https://doi.org/10.1016/j.exger.2021.111634)] [Medline: [34826574](https://pubmed.ncbi.nlm.nih.gov/34826574/)]
6. Office of the Surgeon G. The Surgeon General's Call to Action to Promote Walking and Walkable Communities. Washington DC: US Department of Health and Human Services; 2015.
7. Fox S, Duggan M. Tracking for health. Pew Research Center. 2013. URL: <http://www.pewinternet.org/2013/01/28/tracking-for-health/> [accessed 2023-06-27]
8. Abril EP. Tracking myself: assessing the contribution of mobile technologies for self-trackers of weight, diet, or exercise. *J Health Commun* 2016;21(6):638-646. [doi: [10.1080/10810730.2016.1153756](https://doi.org/10.1080/10810730.2016.1153756)] [Medline: [27168426](https://pubmed.ncbi.nlm.nih.gov/27168426/)]
9. Connected consumer survey 2023. Deloitte Insights. 2022. URL: <https://www2.deloitte.com/us/en/insights/industry/telecommunications/connectivity-mobile-trends-survey.html> [accessed 2023-06-27]
10. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. *Sports Med* 2017;47(7):1303-1315 [FREE Full text] [doi: [10.1007/s40279-016-0663-1](https://doi.org/10.1007/s40279-016-0663-1)] [Medline: [28005190](https://pubmed.ncbi.nlm.nih.gov/28005190/)]
11. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, Steps for Health Collaborative. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *Lancet Public Health* 2022;7(3):e219-e228 [FREE Full text] [doi: [10.1016/S2468-2667\(21\)00302-9](https://doi.org/10.1016/S2468-2667(21)00302-9)] [Medline: [35247352](https://pubmed.ncbi.nlm.nih.gov/35247352/)]
12. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. *Int J Behav Nutr Phys Act* 2020;17(1):78 [FREE Full text] [doi: [10.1186/s12966-020-00978-9](https://doi.org/10.1186/s12966-020-00978-9)] [Medline: [32563261](https://pubmed.ncbi.nlm.nih.gov/32563261/)]
13. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthc Inform Res* 2015;21(4):315-320 [FREE Full text] [doi: [10.4258/hir.2015.21.4.315](https://doi.org/10.4258/hir.2015.21.4.315)] [Medline: [26618039](https://pubmed.ncbi.nlm.nih.gov/26618039/)]
14. Breteler MJ, Janssen JH, Spiering W, Kalkman CJ, van Solinge WW, Dohmen DA. Measuring free-living physical activity with three commercially available activity monitors for telemonitoring purposes: validation study. *JMIR Form Res* 2019;3(2):e11489 [FREE Full text] [doi: [10.2196/11489](https://doi.org/10.2196/11489)] [Medline: [31017587](https://pubmed.ncbi.nlm.nih.gov/31017587/)]
15. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. *JMIR Mhealth Uhealth* 2018;6(4):e94 [FREE Full text] [doi: [10.2196/mhealth.9754](https://doi.org/10.2196/mhealth.9754)] [Medline: [29650506](https://pubmed.ncbi.nlm.nih.gov/29650506/)]
16. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 2007;165(11):1328-1335. [doi: [10.1093/aje/kwm021](https://doi.org/10.1093/aje/kwm021)] [Medline: [17372189](https://pubmed.ncbi.nlm.nih.gov/17372189/)]
17. Dawber TR, Kannel WB. The Framingham study an epidemiological approach to coronary heart disease. *Circulation* 1966;34(4):553-555. [doi: [10.1161/01.cir.34.4.553](https://doi.org/10.1161/01.cir.34.4.553)] [Medline: [5921755](https://pubmed.ncbi.nlm.nih.gov/5921755/)]
18. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and preliminary findings from a new electronic cohort embedded in the Framingham Heart Study. *J Med Internet Res* 2019;21(3):e12143 [FREE Full text] [doi: [10.2196/12143](https://doi.org/10.2196/12143)] [Medline: [30821691](https://pubmed.ncbi.nlm.nih.gov/30821691/)]
19. Esliger DW, Probert A, Connor Gorber S, Bryan S, Laviolette M, Tremblay MS. Validity of the actical accelerometer step-count function. *Med Sci Sports Exerc* 2007;39(7):1200-1204. [doi: [10.1249/mss.0b013e3181ed61a3](https://doi.org/10.1249/mss.0b013e3181ed61a3)] [Medline: [17596790](https://pubmed.ncbi.nlm.nih.gov/17596790/)]
20. Johnson M, Meltz K, Hart K, Schudlach M, Clarkson L, Borman K. Validity of the actical activity monitor for assessing steps and energy expenditure during walking. *J Sports Sci* 2015;33(8):769-776. [doi: [10.1080/02640414.2014.964747](https://doi.org/10.1080/02640414.2014.964747)] [Medline: [25356920](https://pubmed.ncbi.nlm.nih.gov/25356920/)]
21. Colley RC, Tremblay MS. Moderate and vigorous physical activity intensity cut-points for the actical accelerometer. *J Sports Sci* 2011;29(8):783-789. [doi: [10.1080/02640414.2011.557744](https://doi.org/10.1080/02640414.2011.557744)] [Medline: [21424979](https://pubmed.ncbi.nlm.nih.gov/21424979/)]
22. Evenson KR, Sotres-Alvarez D, Deng YU, Marshall SJ, Isasi CR, Esliger DW, et al. Accelerometer adherence and performance in a cohort study of US Hispanic adults. *Med Sci Sports Exerc* 2015;47(4):725-734 [FREE Full text] [doi: [10.1249/MSS.0000000000000478](https://doi.org/10.1249/MSS.0000000000000478)] [Medline: [25137369](https://pubmed.ncbi.nlm.nih.gov/25137369/)]
23. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc* 2011;43(2):357-364 [FREE Full text] [doi: [10.1249/MSS.0b013e3181ed61a3](https://doi.org/10.1249/MSS.0b013e3181ed61a3)] [Medline: [20581716](https://pubmed.ncbi.nlm.nih.gov/20581716/)]
24. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? *Int J Behav Nutr Phys Act* 2011;8:62 [FREE Full text] [doi: [10.1186/1479-5868-8-62](https://doi.org/10.1186/1479-5868-8-62)] [Medline: [21679426](https://pubmed.ncbi.nlm.nih.gov/21679426/)]
25. Lin H, Sardana M, Zhang Y, Liu C, Trinquart L, Benjamin EJ, et al. Association of habitual physical activity with cardiovascular disease risk. *Circ Res* 2020;127(10):1253-1260 [FREE Full text] [doi: [10.1161/CIRCRESAHA.120.317578](https://doi.org/10.1161/CIRCRESAHA.120.317578)] [Medline: [32842915](https://pubmed.ncbi.nlm.nih.gov/32842915/)]
26. Whelton PK, Carey R, Aronow W, Casey DJ, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Hypertension* 2018;71(6):1269-1324 [FREE Full text] [doi: [10.1161/HYP.000000000000066](https://doi.org/10.1161/HYP.000000000000066)] [Medline: [29133354](https://pubmed.ncbi.nlm.nih.gov/29133354/)]

27. Kannel WB, Belanger A, D'Agostino R, Israel I. Physical activity and physical demand on the job and risk of cardiovascular disease and death: the Framingham study. *Am Heart J* 1986;112(4):820-825 [FREE Full text] [doi: [10.1016/0002-8703\(86\)90480-1](https://doi.org/10.1016/0002-8703(86)90480-1)] [Medline: [3766383](https://pubmed.ncbi.nlm.nih.gov/3766383/)]
28. Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A. Evaluation of quality of commercial pedometers. *Can J Public Health* 2006;97 Suppl 1(Suppl 1):S10-15,SS10-16 [FREE Full text] [doi: [10.1007/BF03405359](https://doi.org/10.1007/BF03405359)] [Medline: [16676833](https://pubmed.ncbi.nlm.nih.gov/16676833/)]
29. Mora-Gonzalez J, Gould ZR, Moore CC, Aguiar EJ, Ducharme SW, Schuna JJ, et al. A catalog of validity indices for step counting wearable technologies during treadmill walking: the CADENCE-adults study. *Int J Behav Nutr Phys Act* 2022;19(1):117 [FREE Full text] [doi: [10.1186/s12966-022-01350-9](https://doi.org/10.1186/s12966-022-01350-9)] [Medline: [36076265](https://pubmed.ncbi.nlm.nih.gov/36076265/)]
30. Tyo BM, Fitzhugh EC, Bassett DJ, John D, Feito Y, Thompson DL. Effects of body mass index and step rate on pedometer error in a free-living environment. *Med Sci Sports Exerc* 2011;43(2):350-356. [doi: [10.1249/MSS.0b013e3181e9b133](https://doi.org/10.1249/MSS.0b013e3181e9b133)] [Medline: [20543755](https://pubmed.ncbi.nlm.nih.gov/20543755/)]
31. Pomeroy J, Brage S, Curtis JM, Swan PD, Knowler WC, Franks PW. Between-monitor differences in step counts are related to body size: implications for objective physical activity measurement. *PLoS One* 2011;6(4):e18942 [FREE Full text] [doi: [10.1371/journal.pone.0018942](https://doi.org/10.1371/journal.pone.0018942)] [Medline: [21556140](https://pubmed.ncbi.nlm.nih.gov/21556140/)]
32. Feito Y, Bassett DR, Thompson DL, Tyo BM. Effects of body mass index on step count accuracy of physical activity monitors. *J Phys Act Health* 2012;9(4):594-600. [doi: [10.1123/jpah.9.4.594](https://doi.org/10.1123/jpah.9.4.594)] [Medline: [21946229](https://pubmed.ncbi.nlm.nih.gov/21946229/)]
33. 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;13(2):145-153. [doi: [10.1123/jpah.2015-0203](https://doi.org/10.1123/jpah.2015-0203)] [Medline: [26107045](https://pubmed.ncbi.nlm.nih.gov/26107045/)]
34. Dall PM, McCrorie PRW, Granat MH, Stansfield BW. Step accumulation per minute epoch is not the same as cadence for free-living adults. *Med Sci Sports Exerc* 2013;45(10):1995-2001. [doi: [10.1249/MSS.0b013e3182955780](https://doi.org/10.1249/MSS.0b013e3182955780)] [Medline: [23568091](https://pubmed.ncbi.nlm.nih.gov/23568091/)]
35. Fokkema T, Kooiman TJM, Krijnen WP, VAN DER Schans CP, DE Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. *Med Sci Sports Exerc* 2017;49(4):793-800. [doi: [10.1249/MSS.0000000000001146](https://doi.org/10.1249/MSS.0000000000001146)] [Medline: [28319983](https://pubmed.ncbi.nlm.nih.gov/28319983/)]
36. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. *Int J Behav Nutr Phys Act* 2011;8:79 [FREE Full text] [doi: [10.1186/1479-5868-8-79](https://doi.org/10.1186/1479-5868-8-79)] [Medline: [21798015](https://pubmed.ncbi.nlm.nih.gov/21798015/)]
37. Middelweerd A, VAN DER Ploeg HP, VAN Halteren A, Twisk JWR, Brug J, Te Velde SJ. A validation study of the fitbit one in daily life using different time intervals. *Med Sci Sports Exerc* 2017;49(6):1270-1279 [FREE Full text] [doi: [10.1249/MSS.0000000000001225](https://doi.org/10.1249/MSS.0000000000001225)] [Medline: [28511193](https://pubmed.ncbi.nlm.nih.gov/28511193/)]
38. Schrack JA, Cooper R, Koster A, Shiroma EJ, Murabito JM, Rejeski WJ, et al. Assessing daily physical activity in older adults: unraveling the complexity of monitors, measures, and methods. *J Gerontol A Biol Sci Med Sci* 2016;71(8):1039-1048 [FREE Full text] [doi: [10.1093/gerona/glw026](https://doi.org/10.1093/gerona/glw026)] [Medline: [26957472](https://pubmed.ncbi.nlm.nih.gov/26957472/)]
39. Cuthbertson CC, Moore CC, Sotres-Alvarez D, Heiss G, Isasi CR, Mossavar-Rahmani Y, et al. Associations of steps per day and step intensity with the risk of diabetes: the hispanic community health study / study of latinos (HCHS/SOL). *Int J Behav Nutr Phys Act* 2022;19(1):46 [FREE Full text] [doi: [10.1186/s12966-022-01284-2](https://doi.org/10.1186/s12966-022-01284-2)] [Medline: [35428253](https://pubmed.ncbi.nlm.nih.gov/35428253/)]
40. Spartano NL, Demissie S, Himali JJ, Dukes KA, Murabito JM, Vasani RS, et al. Accelerometer-determined physical activity and cognitive function in middle-aged and older adults from two generations of the Framingham heart study. *Alzheimers Dement (N Y)* 2019;5:618-626 [FREE Full text] [doi: [10.1016/j.trci.2019.08.007](https://doi.org/10.1016/j.trci.2019.08.007)] [Medline: [31660424](https://pubmed.ncbi.nlm.nih.gov/31660424/)]
41. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the all of US research program. *Nat Med* 2022;28(11):2301-2308 [FREE Full text] [doi: [10.1038/s41591-022-02012-w](https://doi.org/10.1038/s41591-022-02012-w)] [Medline: [36216933](https://pubmed.ncbi.nlm.nih.gov/36216933/)]
42. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. *BMJ* 2019;366:14570 [FREE Full text] [doi: [10.1136/bmj.14570](https://doi.org/10.1136/bmj.14570)] [Medline: [31434697](https://pubmed.ncbi.nlm.nih.gov/31434697/)]

Abbreviations

CESD: Center for Epidemiological Studies Depression

eFHS: electronic Framingham Heart Study

FHS: Framingham Heart Study

Gen: generation

ICC: intraclass correlation

MVPA: moderate to vigorous physical activity

NOS: New Offspring Spouses

RURAL: Risk Underlying Rural Areas Longitudinal

Edited by T Leung; submitted 16.11.23; peer-reviewed by C Sakal, P Bergman; comments to author 12.02.24; revised version received 17.05.24; accepted 31.05.24; published 24.07.24.

Please cite as:

Spartano NL, Zhang Y, Liu C, Chernofsky A, Lin H, Trinquart L, Borrelli B, Pathiravasan CH, Kheterpal V, Nowak C, Vasan RS, Benjamin EJ, McManus DD, Murabito JM

Agreement Between Apple Watch and Actical Step Counts in a Community Setting: Cross-Sectional Investigation From the Framingham Heart Study

JMIR Biomed Eng 2024;9:e54631

URL: <https://biomedeng.jmir.org/2024/1/e54631>

doi: [10.2196/54631](https://doi.org/10.2196/54631)

PMID:

©Nicole L Spartano, Yuankai Zhang, Chunyu Liu, Ariel Chernofsky, Honghuang Lin, Ludovic Trinquart, Belinda Borrelli, Chathurangi H Pathiravasan, Vik Kheterpal, Christopher Nowak, Ramachandran S Vasan, Emelia J Benjamin, David D McManus, Joanne M Murabito. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 24.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Assessing the Accuracy of Smartwatch-Based Estimation of Maximum Oxygen Uptake Using the Apple Watch Series 7: Validation Study

Polona Caserman¹, Dr-Ing; Sungsoo Yum¹, BSc; Stefan Göbel¹, PD, Dr-Ing; Andreas Reif², Prof Dr; Silke Matura², PD, Dr

¹Serious Games Research Group, Technical University of Darmstadt, Darmstadt, Germany

²Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Goethe University Frankfurt, University Hospital, Frankfurt am Main, Germany

Corresponding Author:

Polona Caserman, Dr-Ing

Serious Games Research Group

Technical University of Darmstadt

Rundeturmstraße 10

Darmstadt, 64289

Germany

Email: polona.caserman@tu-darmstadt.de

Abstract

Background: Determining maximum oxygen uptake (VO_{2max}) is essential for evaluating cardiorespiratory fitness. While laboratory-based testing is considered the gold standard, sports watches or fitness trackers offer a convenient alternative. However, despite the high number of wrist-worn devices, there is a lack of scientific validation for VO_{2max} estimation outside the laboratory setting.

Objective: This study aims to compare the Apple Watch Series 7's performance against the gold standard in VO_{2max} estimation and Apple's validation findings.

Methods: A total of 19 participants (7 female and 12 male), aged 18 to 63 (mean 28.42, SD 11.43) years were included in the validation study. VO_{2max} for all participants was determined in a controlled laboratory environment using a metabolic gas analyzer. Thereby, they completed a graded exercise test on a cycle ergometer until reaching subjective exhaustion. This value was then compared with the estimated VO_{2max} value from the Apple Watch, which was calculated after wearing the watch for at least 2 consecutive days and measured directly after an outdoor running test.

Results: The measured VO_{2max} (mean 45.88, SD 9.42 mL/kg/minute) in the laboratory setting was significantly higher than the predicted VO_{2max} (mean 41.37, SD 6.5 mL/kg/minute) from the Apple Watch ($t_{18}=2.51$; $P=.01$) with a medium effect size (Hedges $g=0.53$). The Bland-Altman analysis revealed a good overall agreement between both measurements. However, the intraclass correlation coefficient $ICC(2,1)=0.47$ (95% CI 0.06-0.75) indicated poor reliability. The mean absolute percentage error between the predicted and the actual VO_{2max} was 15.79%, while the root mean square error was 8.85 mL/kg/minute. The analysis further revealed higher accuracy when focusing on participants with good fitness levels (mean absolute percentage error=14.59%; root-mean-square error=7.22 ml/kg/minute; $ICC(2,1)=0.60$ 95% CI 0.09-0.87).

Conclusions: Similar to other smartwatches, the Apple Watch also overestimates or underestimates the VO_{2max} in individuals with poor or excellent fitness levels, respectively. Assessing the accuracy and reliability of the Apple Watch's VO_{2max} estimation is crucial for determining its suitability as an alternative to laboratory testing. The findings of this study will apprise researchers, physical training professionals, and end users of wearable technology, thereby enhancing the knowledge base and practical application of such devices in assessing cardiorespiratory fitness parameters.

(JMIR Biomed Eng 2024;9:e59459) doi:[10.2196/59459](https://doi.org/10.2196/59459)

KEYWORDS

maximal oxygen uptake; oxygen consumption; cardiorespiratory fitness; physical fitness; physical activity; fitness tracker; wearables; wearable; exercise; fitness; tracker; trackers; cardiorespiratory; wrist worn device; devices; validation study; VO₂max; sport watch; fitness level; mobile phone

Introduction

The concept of the maximum oxygen uptake (VO₂max), established in 1923 by Hill and Lupton [1] is a fundamental measure in assessing cardiorespiratory fitness [2] and is also often used to determine an individual's physical fitness level [3,4]. Cardiorespiratory fitness is defined as the ability of the circulatory and respiratory systems to supply oxygen to the muscles during sustained physical activity [3]. VO₂max is also often used as a performance measure [5,6]. Previous research concludes that VO₂max is closely related to all-cause mortality and underscores the importance of enhancing VO₂max to reduce the risks of developing cardiovascular diseases [7-10].

Typically, VO₂max is measured in a controlled laboratory setting using a metabolic gas analyzer during an incremental exercise test, commonly administered on a motorized treadmill or a cycle ergometer [7]. During the test, either the speed on the treadmill or the resistance on the ergometer is gradually increased, until participants reach maximum exhaustion. Such tests are typically directed toward special populations, for example, individuals with known or suspected cardiovascular diseases or endurance athletes. Laboratory tests require expensive equipment (ie, a metabolic gas analyzer) and trained personnel and are therefore often costly and time-consuming. As the maximal exercise test necessitates participants to achieve maximal exertion, it may not always be safe for everyone, especially not without medical supervision and emergency equipment [11]. Accordingly, given the impracticality of VO₂max assessments for everyday application and their limited accessibility by the general population, the emergence of fitness trackers has provided a convenient and accessible alternative for estimating VO₂max in real-world settings. A recent survey shows that 21% of Americans already use a smartwatch or a fitness tracker such as the Garmin, Fitbit, or Apple Watch [12]. According to another recent survey, wearable technology has also been identified as the number one fitness trend in 2022 [13].

Prior investigations have already assessed the reliability and validity of various wearables, using heart rate (HR) as a metric for quantifying individual physiological exertion [14]. Further studies have explored the potential of biometric monitoring technologies in estimating users' cardiovascular fitness levels, using algorithms like those developed by Firstbeat Analytics [15] and used by prominent brands such as Garmin and Huawei [16]. Additionally, researchers developed their methodologies to calculate oxygen uptake using wearable devices or smartphones [17-20]. Previous research further validated various fitness tests carried out using smartphones, offering additional insights into the accuracy of these devices in evaluating physical metrics [21,22]. Despite the promising potential of wrist-worn devices in facilitating fitness assessments, concerns have been raised regarding the accuracy and reliability of estimating parameters, such as VO₂max or VO₂ peak, with particular

concern about their potential misuse by consumers for making medical decisions [23]. While several studies have shown that wearables are very accurate [15,24-29], contradictory evidence suggests potential overestimation or underestimation in VO₂max measurements [30-33]. Notably, only little research has been conducted on the accuracy of VO₂max predictions among participants with varying fitness levels, particularly those with lower or higher fitness levels [34,35].

Given the Apple Watch's dominant position in the global smartwatch market with the largest share of shipments [36] and being the primary choice for the majority of users [12], assessing the accuracy and reliability of its VO₂max estimation becomes critical in determining its potential as a dependable alternative to traditional laboratory testing. However, only a little research has been conducted evaluating the accuracy of the Apple Watch in estimating cardiorespiratory fitness indicators. Most of the studies that validated the accuracy of the Apple Watch focused on fitness parameters such as energy expenditure, HR, HR variability, or oxygen consumption reserve [37-41]. There remains a gap in the literature regarding the specific evaluation of the Apple Watch to predict VO₂max. While Apple has conducted an extensive study to validate its VO₂max estimation algorithm [42], concerns exist regarding potential bias and the limited medical representativeness of their findings.

To address these concerns and contribute to the understanding of wearable technology in fitness assessment, this study aims to assess the accuracy and reliability of VO₂max estimation using the Apple Watch Series 7. Toward this end, we conducted a comparative analysis between the VO₂max estimation of the Apple Watch 7 and the gold-standard testing in a laboratory setting, using a metabolic gas analyzer. The level of agreement was evaluated using Bland-Altman plots. We calculated the error in terms of mean absolute percentage error (MAPE) and root-mean-square error (RMSE), and further assessed the reliability by calculating the intraclass correlation coefficient (ICC). The outcomes of this study will hopefully provide valuable insights into the performance of the Apple Watch Series 7 relative to other validation studies of wrist-worn devices and Apple's validation results.

Methods

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the Technical University of Darmstadt (approval EK 11/2023; March 20, 2023). In the first session, all participants were informed about the specific purpose of the study. We informed them that all collected data are confidential and solely used in anonymized form. To ensure anonymity, each participant was assigned a pseudonym. Participants were informed about the risks and their right to terminate the experiment at any point

without the need for an explanation. Afterward, participants provided written informed consent, completed a demographics questionnaire, and responded to inquiries regarding their physical activity.

Study Design

The study used a repeated measures design with each participant completing 2 sessions on separate days, with a minimum resting period of 48 hours in between. Before undergoing the tests, participants were advised to refrain from consuming alcohol or any other substances that could potentially influence their respiratory system and HR. This precautionary measure aimed to ensure accurate readings and mitigate the risk of any potential false results during the testing procedure. The initial session was conducted in a controlled laboratory setting to establish a reference value for VO_{2max} . The subsequent session took place on the university's stadium track field, using the Apple Watch Series 7 to obtain an estimated VO_{2max} value. Following the completion of both sessions, the VO_{2max} values obtained from the 2 methods were compared against each other for analysis.

Measurement of VO_{2max} in a Laboratory Setting—Cycle Test

The performance test in the laboratory setting was assessed through an endurance test using a cycle ergometer. Such tests are widely used in sports science to measure VO_{2max} , serving as a crucial indicator of aerobic endurance performance [43]. Due to the lack of medical expertise to conduct a maximal exercise test, we alternatively conducted a graded exercise test until subjective exhaustion. This decision was influenced by our ability to adhere to a rigorous protocol within the controlled environment of the laboratory, as well as the availability of the necessary equipment to monitor respiratory parameters and promptly terminate the session if the participant's safety was compromised. Submaximal exercise prediction was also used in the field test using the Apple Watch, which facilitates comparison of the values derived from sessions 1 and 2.

Accordingly, the reference VO_{2max} value was determined through a graded exercise test conducted on a cycle ergometer, using the portable metabolic gas analyzer (VO_2 Master Health Sensors Inc [44]). Evidence of the measurement accuracy of the hardware used can be found in references [45,46]. The gas analyzer was calibrated prior to each test (ie, for each participant), using a 3-L syringe for both flow and gas calibration. Furthermore, the supervisor entered the participants' age, sex, height, and weight in the VO_2 Master Manager app (installed on an iPhone 13 Mini), which was paired with the gas analyzer. After the calibration, participants put on the electrocardiogram chest strap (Polar H10 Heart Rate Sensor [47]) and the gas analyzer while the supervisor (SY) checked the plausibility of the system (ie, both sensors connected to the smartphone via Bluetooth and transmitting the data via VO_2 Master Manager app). Once participants successfully put on the equipment, they were instructed to sit on the cycle ergometer (ERGO-FIT Cycle 4073 [48]) after adjusting the seat height according to their height.

Once the setup was completed, the endurance test was conducted. The laboratory protocol was equal for male and

female participants. Throughout the test, vital parameters (ie, the HR and breathing) and the participant's current state were continuously monitored. Participants started with a 3-minute warm-up phase, riding on the cycle ergometer at a workload of 50 W at a speed of 60 rotations per minute. Afterward, the ergometer's resistance was increased by 50 W every 2 minutes until one of the termination criteria was met (based on the criteria by Klingenheben et al [49]):

- Maximum HR, based on age and sex, individually calculated for each participant using the Fairbairn equation [50], was exceeded for 10 consecutive seconds:

$$HR_{\max\text{Fairban}}=208-0.8\times\text{age, for male participants}$$

$$HR_{\max\text{Fairban}}=201-0.6\times\text{age, for female participants}$$

We intentionally used the Fairbairn equation to predict the maximum HR, instead of using the Fox equation $HR_{\max\text{Fox}}=220-\text{age}$ [51], which is only dependent on age. According to the analysis by Cleary et al [52], the Fairbairn equation, which considers the age and sex of the participants, is more accurate.

- Inability to maintain a pedal rate of 60 rotations per minute for more than 3 seconds
- An abnormally rapid acceleration or deceleration in HR that is not consistent with physiological norms
- Plateau in VO_2 , despite increasing resistance on the ergometer (increase <1 mL/kg/minute)
- Symptoms of angina pectoris (ie, pain behind the breastbone, tightness, numbness, nausea, vomiting, sweating, and shortness of breath, and anxiety)
- Other conspicuous findings, such as malaise, dizziness, headache, conspicuous pallor, and other complaints
- Signs of respiratory insufficiency could be observed, that is, participants' ventilation reached a dangerous level (around 150 L/minute) in the VO_2 Master Manager app
- Self-reported volitional exhaustion or fatigue
- Failure of monitoring equipment

At the end of the session, protocol outcomes were saved for each participant. In addition to VO_{2max} , the gas analyzer provided the following parameters in real time:

- Metabolism:
 - Absolute oxygen consumption (VO_2 [mL/minute])
 - Oxygen consumption relative to weight (VO_2 [mL/kg/minute])
 - Energy expenditure (Kcal/day)
 - Calories (kcal/hour)
- Pulmonary function:
 - Ventilation; air moved by lungs (V_e [L/minute])
 - Respiratory frequency; breaths per minute (beats per minute)
 - Tidal volume; volume breathed in a breath (L)
- Respiratory efficiency:
 - (V_e/VO_2)
 - Fraction of oxygen in expired breath (FeO_2 [%])
- Cardiac function:
 - HR (beats per minute)

- RR Intervals (RR [milliseconds])

Estimation of VO₂max Using the Apple Watch—Track Field Test

Within 1 week after the initial laboratory session, participants were provided with an iPhone SE 2020 and an Apple Watch (Series 7, 41 mm). The Apple Watch was paired with an iPhone that had been reset to factory setting to ensure data privacy. To complete the setup of the Apple Watch, the supervisor (SY) ensured that participants entered their age, sex, height, and weight in the iPhone.

Participants were instructed to wear the Apple Watch continuously, including during sleep and showers, for at least 48 hours prior to the second session. This prolonged wearing duration was essential as the Apple Watch required at least 24 hours of continuous wear time to reliably estimate VO₂max.

The precise algorithm for VO₂max estimation is not publicly disclosed; however, discussions with Apple technical support revealed that it incorporates resting HR measurements, exercise HR measurements, and GPS-derived velocity data from outdoor runs. To ensure a valid VO₂max from the Apple Watch, we consulted with the manufacturer and adhered to the following procedure: participants needed to complete at least 1 training prior to the track field test, that is, an outdoor walk for 15-20 minutes. They needed to manually measure the HR every hour (using the preinstalled Health app), in addition to the passive measurements of the Apple Watch itself. Throughout the process, participants needed to ensure that the Apple Watch was always connected to the iPhone, which maintained an internet connection.

Only participants who followed the instructions and completed the outdoor walk were permitted to proceed with the run test. The run test was conducted at the university stadium at the Technical University of Darmstadt. Consistent with our laboratory protocol, we used a submaximal exercise test to mitigate the risk of injury; however, in this session, the test was conducted outdoors. The outdoor setting was necessary to ensure a sufficient GPS signal.

Before the run test, participants were given brief instructions. Particularly, they were instructed to activate the outdoor running app on their Apple Watch prior to starting the track run. To minimize the risk of injury, the protocol included a 5-minute warm-up phase, during which participants ran at a moderate pace. Following the warm-up, participants continued at a self-selected running pace, ensuring a minimum duration of 15 minutes. Once participants completed the run and returned to the starting point, they stopped the recording on their Apple Watch and proceeded with a cool-down phase. Subsequently, the supervisor accessed relevant metrics from the Health app on the paired iPhone, specifically the estimated VO₂max in the cardio fitness section.

Recruitment

Participants were recruited among students and employees of the Technical University of Darmstadt through the Discord server from the IT department and the university's mailing list. To ensure a diverse range of fitness levels, we also recruited

members of a local fitness studio. Eligibility criteria required participants to be older than 18 years and in good health. To streamline the selection process, the Physical Activity Readiness Questionnaire [53] was administered. As a result, individuals with any preexisting heart disease, cardiovascular conditions, orthopedic injuries, or current use of medication were deemed ineligible for participation.

To determine the required sample size, we conducted a priori power analysis using G*Power (version 3.1; Heinrich-Heine-Universität Düsseldorf) [54] with a power of 0.8, a significance level of 0.05, and a medium effect size of 0.5. This analysis indicated a minimum sample size of 27 participants. Therefore, considering expected dropouts, we initially aimed for a larger sample size of at least 30 participants. Recruitment took place over a 4-week period in the spring of 2023.

Statistical Analysis

All data were analyzed using MATLAB (MathWorks, Inc), including external code [55,56].

We first assessed the limit of agreement between the values obtained from laboratory measurements and those provided by the Apple Watch using the Bland-Altman plot. The Bland-Altman plot enables us to evaluate if the 2 methods of measurement show a sufficient level of agreement [57]. It displays the limits of agreement by using the mean and SD of the differences between the 2 methods. As recommended by the authors themselves, 95% of the data points should lie within ± 2 SD of the mean difference [57,58]. Additionally, the plot also allows us to spot outliers and to see whether there is any trend in overestimating or underestimating.

Second, in addition to the Bland-Altman plots, we calculated the ICC(2,1) to test for bias and absolute agreement in VO₂max estimation. ICC is different from correlations such as Pearson or Spearman correlation. Calculating correlation is not appropriate to evaluate the measure of agreement, especially as the correlation coefficient depends on both the variation between individuals (ie, between the true values) and the variation within individuals (measurement error) [57]. ICC is suitable for reliability analyses, where a value less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 indicate poor, moderate, good, and excellent reliability, respectively [59].

Third, similar to other validation studies, we used the MAPE and RMSE to calculate the overall measurement error between the VO₂max value derived from the Apple Watch and the metabolic gas analyzer. MAPE was calculated as the average absolute difference between the actual and the predicted measure divided by the actual measure and multiplied by 100 [60]. Furthermore, RMSE was calculated as the square root of the average of the squared differences between predicted and observed values [61].

Finally, to determine any significant differences between the predicted and measured VO₂max, we used statistical tests, specifically the paired 1-tailed *t* test. We tested the assumption of normally distributed data using the Anderson-Darling test

($P=.65$). Furthermore, we calculate the effect size using Hedges g , taking the sample size into account [62], with a value of 0.2 representing a small, 0.5 a medium, and 0.8 a large effect size [63].

Data Analysis and Fitness Level Categorization

In the first step, we analyzed the entire data set to assess the overall performance of the Apple Watch. Additionally, we aimed to get better insights regarding its performance across varying user fitness levels. To achieve this, participants were categorized into 3 groups based on their reference $VO_2\text{max}$ obtained from the laboratory setting. Hence, based on the fitness categories outlined by the Fitness Registry and the Importance of Exercise National Database [64], participants were divided into poor, good, and excellent fitness levels, allowing us a more nuanced investigation of the Apple Watch's estimations.

Table 1. Participant characteristics.

	Male (n=12, 63%), mean (SD)	Female (n=7, 37%), mean (SD)	Total (n=19), mean (SD)
Age (in years)	28.17 (12.40)	28.71 (10.63)	28.37 (11.48)
BMI (kg/m^2)	23.92 (3.79)	23.04 (2.11)	23.60 (3.23)

Limit of Agreement

The detailed results are presented in Table 2. The mean $VO_2\text{max}$ determined in the laboratory setting was 45.88 (SD 9.42) mL/kg/minute, ranging from 32 to 64 mL/kg/minute. Furthermore, the mean estimated $VO_2\text{max}$ from the Apple Watch was 41.37 (SD 6.50) mL/kg/minute, ranging from 29 to 52 mL/kg/minute. Our analysis revealed that the measured $VO_2\text{max}$ is significantly higher than the predicted value from the Apple Watch ($t_{18}=2.51$; $P=.01$) with a medium effect size

Results

Participants

Out of the 30 (14 female and 16 male) initially recruited participants, 6 participants withdrew from the study before the first session due to health and personal reasons. Additionally, after the initial session, 4 participants were deemed ineligible for the study due to health concerns and recommendations from their respective health care providers, and 1 participant did not attend the second session due to personal reasons.

A total of 19 participants successfully completed the initial session in the laboratory setting, which involved a cycle test until subjective exhaustion and metabolic gas analysis, followed by the second session including an outdoor running test. Among the participants, 7 participants were female (mean age 28.86, SD 10.48 years; mean BMI 23.09, SD 2.31 kg/m^2) and 12 participants were male (mean age 28.17, SD 12.40 years; mean BMI 23.76, SD 3.99 kg/m^2). Participant characteristics are further detailed in Table 1.

(Hedges $g=0.53$). These findings are consistent with observations from the Bland-Altman plot (Figure 1A), showing an overall underestimation of $VO_2\text{max}$ by the Apple Watch. Specifically, the mean difference (bias) between the laboratory value and the estimated $VO_2\text{max}$ value from the Apple Watch is -4.51 (SD 7.82) mL/kg/minute. Although all data points fall within the limits of agreement, indicating "good agreement" between the 2 methods, the ICC(2,1) of 0.47 (95% CI 0.06-0.75) suggests only poor to moderate reliability.

Figure 1. Bland-Altman plot of mean (x-axis) and difference (y-axis) between measured VO₂max in the laboratory and predicted VO₂max from the Apple Watch. The solid line represents the mean difference and the dashed lines present the 95% limit of agreement.

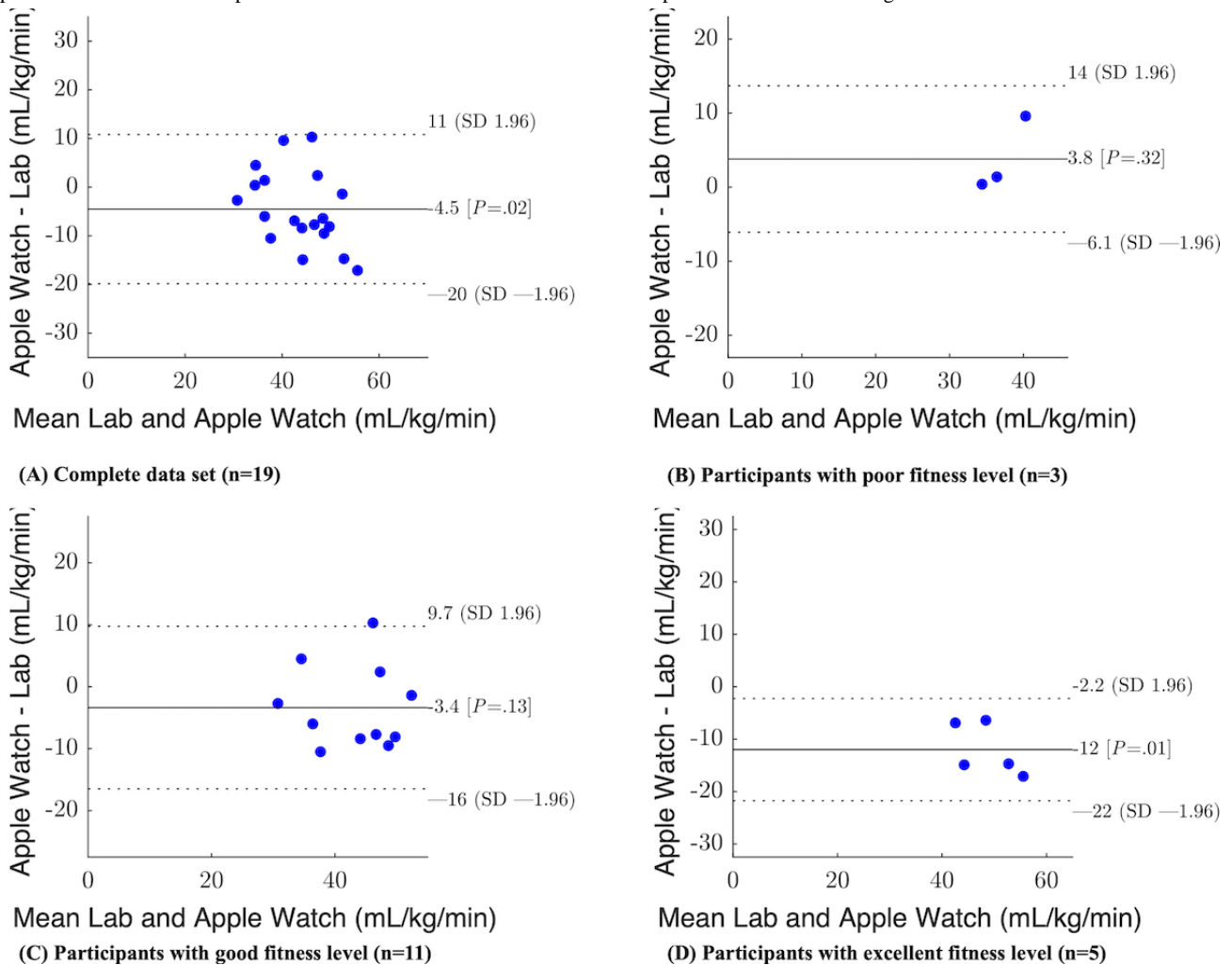


Table 2. Descriptive examination of the differences between the measured and predicted VO₂max.

Fitness level ^a	Participant pool (n=19), n (%)	VO ₂ max—Lab ^b (mL/kg/minute), mean (SD)	VO ₂ max—Apple Watch ^c		MAPE ^e (%)	RMSE ^f (mL/kg/minute)	ICC (2,1) ^g ICC (95% CI)
			VO ₂ max—Lab ^b (mL/kg/minute), mean (SD)	VO ₂ max delta ^d (mL/kg/minute), mean (SD)			
Poor	3 (16)	35.13 (.81)	38.93 (5.48)	3.8 (5.05)	10.71	5.61	0.14 (-0.61 to 0.96)
Good	11 (58)	44.81 (7.97)	41.44 (7.70)	-3.37 (6.69)	14.59	7.22	0.60 (0.09 to 0.87)
Excellent	5 (26)	54.70 (7.28)	42.70 (4.46)	-12 (4.98)	21.47	12.80	0.23 (-0.07 to 0.79)
Combined	19 (100)	45.88 (9.42)	41.37 (6.5)	-4.51 (7.82)	15.79	8.85	0.47 (0.06 to 0.75)

^aCategorized according to sex and age based on the Fitness Registry and the Importance of Exercise National Database [64] criteria.

^bVO₂max—Lab: measured VO₂max in the laboratory.

^cVO₂max—Apple Watch: estimated VO₂max from the Apple Watch.

^dVO₂max delta: Apple Watch estimate versus laboratory measurement.

^eMAPE: mean absolute percentage error.

^fRMSE: root mean square error.

^gICC (2,1): intraclass correlation coefficient.

We furthermore analyzed the limit of agreement for participants with lower and higher fitness levels. When the participants were split into groups of poor (n=3), good (n=11), and excellent (n=5) fitness levels, the smartwatch showed a bias of mean 3.80 (SD 5.05) mL/kg/minute, mean -3.37 (SD 6.69) mL/kg/minute, and

mean -12.00 (SD 4.98) mL/kg/minute, respectively. As depicted in Figures 1B-1D, the Apple Watch tends to overestimate VO₂max for participants with a poor fitness level while underestimating it for those with a higher fitness level.

Moreover, the ICC for poor and excellent fitness levels was 0.14 and 0.23, respectively, indicating poor reliability. Only for participants with good ($n=11$) fitness levels, an ICC(2,1) of 0.60 indicates moderate reliability. However, it is important to highlight the limitations associated with interpreting the results for subgroups due to the small sample size.

Error Between Predicted and Actual VO₂max

The MAPE in the cohort of all participants ($n=19$) was 15.78%, with an RMSE of 8.85 mL/kg/minute. Upon dividing the VO₂max values into categories based on poor, good, and excellent fitness levels, the smartwatch showed MAPEs of 10.71%, 14.59%, and 21.47%, respectively. Regarding RMSE, the smartwatch showed values of 5.61, 7.22, and 12.80 mL/kg/minute for participants with poor, good, and excellent fitness levels, respectively. However, as already mentioned before, it is important to emphasize the limitation in interpreting results for subgroups due to the limited sample size.

Discussion

Principal Results

The purpose of this study was to assess the accuracy of the VO₂max estimation of the Apple Watch Series 7. Other validation studies using the Apple Watch focused on evaluating the accuracy of measuring oxygen consumption reserve [41], HR [38,39], HR variability [40], or energy expenditure [37]. To the best of our knowledge, this is the first study validating the VO₂max using the Apple Watch, aside from Apple's validation study [42].

Overall, our findings reveal a significant underestimation of the estimated VO₂max value from the Apple Watch ($t_{18}=2.51$; $P=.01$; bias: mean -4.51 , SD 7.82 mL/kg/minute; Hedges $g=0.53$). These results deviate from the original validation study by Apple [42], which reported a smaller bias of mean 1.2 (SD 4.4) mL/kg/minute and mean 1.4 (SD 4.7) mL/kg/minute for the design and validation groups, respectively. However, it is important to acknowledge that our VO₂max value from the Apple Watch was obtained after only 1 outdoor walking and running session. According to Apple's explanation, increasing the number of outdoor workouts enhances the accuracy of the VO₂max estimate [42]. In contrast to our study, Apple's validation study was designed as a longitudinal study, extending over an average of 441 days for the design group and 390 days for the validation group. The researchers computed the mean and SD for differences between the last estimated VO₂max from the Apple Watch and the mean VO₂max value determined in up to 6 maximal or submaximal cardiopulmonary exercise tests while wearing the Apple Watch Series 4. However, it remains unclear how exactly the cardiopulmonary exercise test was conducted. Therefore, a direct comparison of our results with theirs is not feasible as they estimated VO₂max from multiple workouts. It is plausible that our results would show also a smaller error if the participants in our study wore the watch for a longer duration. Apple's statement that the VO₂max estimation by the Apple Watch is accurate and reliable compared to conventional methods of VO₂max measurement [42] can

therefore not be contradicted on the basis of the available findings.

Our findings regarding intraclass correlation reveal that ICC(2,1)=0.47, indicating relatively poor reliability, as outlined in reference [59]. Upon excluding participants with poor and excellent fitness levels and focusing solely on those with good fitness levels, we observed an improved ICC(2,1) value of 0.60, suggesting moderate reliability. These results underscore the influence of fitness levels on the reliability of VO₂max estimation through the Apple Watch. The validation study conducted by Apple calculating ICC(A,1), yielded values of 0.89 and 0.86 for the design and validation groups, respectively, indicative of good reliability [42]. Notably, Apple's evaluation involved assessing absolute agreement per participant by comparing the last valid VO₂max estimate with the value estimated at least 28 days prior. This methodology differs from our approach, where we aimed to evaluate the reliability between laboratory-measured values and Apple Watch estimates without a significant time gap.

Comparison With Prior Work

There is no standardized threshold for high or low MAPE, but we consider an error below 5% to be a good indicator for an accurate measurement. Regarding our results from the Apple Watch, we can conclude that regardless of the fitness level of the participants, the MAPE exceeded 10%. Unfortunately, related studies do not consistently report MAPE values. Nevertheless, 1 study using Polar [30] showed MAPE values above 10% (specifically 13.2%). In addition, studies with Fitbit devices showed MAPE around 10% [27,65]. Conversely, studies on Garmin devices [25,30,33,35], using algorithms developed by Firstbeat Technologies [15], consistently reported MAPE values well below 10%, highlighting their superior accuracy compared to other smartwatches.

We furthermore attempted to compare our results on ICC with those of other studies. Since not all studies provided comprehensive information regarding ICC forms used, making direct comparisons proved to be challenging. Nevertheless, studies on the Garmin Watch have indicated high reliability, with ICC(2,1)=0.87 [29] or ICC(3,1)=0.94 [35], although it is important to note that the latter study validated the estimation of VO₂ peak rather than VO₂max.

In terms of fitness levels, this study aligns with findings from related research using various smartwatches. Consistent with observations from references [30,31,33-35,65], our results suggest a tendency for the Apple Watch to overestimate VO₂max values among users with poor fitness levels (mean 3.80, SD 5.05 mL/kg/minute) and underestimate them among those with higher fitness levels (mean -3.37 , SD 6.69 mL/kg/minute and mean -12.00 , SD 4.98 mL/kg/minute for good and excellent fitness levels, respectively). However, it should be noted that this study involved a relatively small sample size, and classifying participants based on their fitness levels further reduced the sample size in each group ($n=3$ for participants with lower fitness, $n=11$ for those with good fitness, and $n=5$ for those with excellent fitness). Despite this limitation, our findings suggest that the Apple Watch may provide more accurate

VO₂max estimates for users with poor or good fitness levels. This conclusion is further supported by MAPE, which shows a smaller error for users with poorer fitness levels while the error increases in participants with higher fitness levels (see also Table 2). This could be attributed to the potential influence of fitness levels on the accuracy of physiological measurements obtained through wearable devices. Nonetheless, further research with larger sample sizes is necessary to validate and elucidate these observations. Such investigations could shed light on the factors influencing the performance of wearable devices in estimating VO₂max across various fitness levels, thereby enhancing our understanding of their use in health and fitness monitoring.

Limitations

The major limitation of this study is the small sample size. Although we aimed to recruit at least 30 participants, we ultimately obtained complete data from only 19 participants. To address this limitation, we reported effect sizes alongside our statistical tests, ensuring that our results remain reliable despite the smaller sample size. Nevertheless, further studies with larger and more varied populations are recommended to build on these findings and enhance the statistical power of the conclusions. It would also be beneficial to extend the duration during which participants consistently wear a smartwatch, as we believe that longer wear periods may enhance the accuracy of VO₂max estimation by the Apple Watch.

Although VO₂max measurement is considered the gold standard among sports medicine professionals for determining an individual's fitness level, prior research has suggested that VO₂max is constrained by the variability in an individual's effort and is highly reliant on VO₂max extent to which participants are properly motivated to achieve their true maximum [66]. Furthermore, as VO₂max criteria are not standardized, there is some uncertainty regarding whether the true VO₂max has actually been attained and if a maximum effort has been exerted [67]. To address these concerns, Edvardsen et al [68] proposed revised termination criteria for VO₂max tests that consider sex and age. Furthermore, as the true VO₂max value can differ, depending on whether the cardiopulmonary exercise testing was done on a treadmill or cycle ergometer, it would be important to use both tests independently to achieve optimal fitness assessment [69]. Nevertheless, varying termination criteria, testing methodologies, and participant populations across studies continue to pose challenges [67]. Despite these challenges, our aim involved making selective comparisons between our study and related research, diligently acknowledging the notable differences between the studies.

Another limitation we encountered was related to calibration error. Our attempt to compare the approximate prediction method of the Apple Watch with a gas analyzer was conducted using a graded exercise test until subjective exhaustion, potentially leading to an underestimation of the true VO₂max value. Noonan and Dean [70] outlined the advantages of submaximal exercise tests over maximal exercise tests, citing factors such as requirements for trained personnel and safety concerns. They conclude that submaximal exercise tests are reliable if an appropriate protocol is selected and the protocol is followed. However, it is crucial to note the potential influence of different protocols or increased participant motivation, as these factors could impact the measured VO₂max.

An additional limitation of our study is the lack of medical equipment. Ideally, we would have conducted periodic blood samples to measure the lactate threshold, allowing us to detect the point when the participant's respiratory system attained its maximum capacity. The lactate concentration in blood is a valuable metric to monitor because an increase in blood lactate indicates a transition from aerobic to anaerobic exercise, suggesting that the body has surpassed its capacity for oxygen uptake to supply the muscles adequately [71]. Unfortunately, due to the unavailability of suitable equipment and the lack of medical professionals capable of carrying out such data collection, we were unable to include blood lactate as a termination criterion in our study. Additionally, it would have been ideal to monitor the volume of carbon dioxide produced; however, this capability is not provided by the VO₂ Master Analyzer.

Conclusions

Overall, the Apple Watch Series 7 underestimated VO₂max compared to the values obtained using the gold standard assessment methods within a laboratory setting. This underestimation was even pronounced in participants with very high fitness levels. On the contrary, VO₂max values were overestimated by the Apple watch in participants with comparably low fitness levels. These findings highlight the importance of calibrating consumer-grade fitness trackers for greater accuracy across a diverse range of fitness levels. As consumer-grade technology continues to evolve, there is an opportunity for ongoing research and development to close the gap between the accuracy of portable devices and laboratory-grade equipment. This would not only enhance individual training and health monitoring but could also expand the use of such wearables in professional sports and clinical settings.

Acknowledgments

Generative artificial intelligence, specifically ChatGPT-3.5, was not used to generate any new content for this manuscript. Its use was limited to proofreading for grammatical errors and enhancing the clarity and flow of the writing.

Authors' Contributions

PC and SM contributed to the conceptualization, formal analysis, methodology, validation, and visualization. SY contributed to data curation, formal analysis, investigation, methodology, and software. SG and AR provided supervision. PC, SY, SG, AR, and SM wrote the original draft of the manuscript and further reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

References

1. Hill AV, Lupton H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. *QJM* 1923;os-16(62):135-171. [doi: [10.1093/qjmed/os-16.62.135](https://doi.org/10.1093/qjmed/os-16.62.135)]
2. American College of Sports Medicine. ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription. Philadelphia, PA: Lippincott Williams & Wilkins; 2012:864.
3. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985;100(2):126-131 [FREE Full text] [Medline: [3920711](https://pubmed.ncbi.nlm.nih.gov/3920711/)]
4. Raghuvveer G, Hartz J, Lubans DR, Takken T, Wiltz JL, Mietus-Snyder M, et al. Cardiorespiratory fitness in youth: an important marker of health: a scientific statement from the American Heart Association. *Circulation* 2020;142(7):e101-e118 [FREE Full text] [doi: [10.1161/CIR.0000000000000866](https://doi.org/10.1161/CIR.0000000000000866)] [Medline: [32686505](https://pubmed.ncbi.nlm.nih.gov/32686505/)]
5. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol* 1955;8(1):73-80. [doi: [10.1152/jappl.1955.8.1.73](https://doi.org/10.1152/jappl.1955.8.1.73)]
6. Bassett DR, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med Sci Sports Exerc* 2000;32(1):70-84. [doi: [10.1097/00005768-200001000-00012](https://doi.org/10.1097/00005768-200001000-00012)] [Medline: [10647532](https://pubmed.ncbi.nlm.nih.gov/10647532/)]
7. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. *Circulation* 2016;134(24):e653-e699. [doi: [10.1161/CIR.0000000000000461](https://doi.org/10.1161/CIR.0000000000000461)] [Medline: [27881567](https://pubmed.ncbi.nlm.nih.gov/27881567/)]
8. McMurray RG, Ainsworth BE, Harrell JS, Griggs TR, Williams OD. Is physical activity or aerobic power more influential on reducing cardiovascular disease risk factors? *Med Sci Sports Exerc* 1998;30(10):1521-1529. [doi: [10.1097/00005768-199810000-00009](https://doi.org/10.1097/00005768-199810000-00009)] [Medline: [9789853](https://pubmed.ncbi.nlm.nih.gov/9789853/)]
9. Myers J, McAuley P, Lavie CJ, Despres J, Arena R, Kokkinos P. Physical activity and cardiorespiratory fitness as major markers of cardiovascular risk: their independent and interwoven importance to health status. *Prog Cardiovasc Dis* 2015;57(4):306-314. [doi: [10.1016/j.pcad.2014.09.011](https://doi.org/10.1016/j.pcad.2014.09.011)] [Medline: [25269064](https://pubmed.ncbi.nlm.nih.gov/25269064/)]
10. Fogelholm M. Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors. a systematic review. *Obes Rev* 2010;11(3):202-221. [doi: [10.1111/j.1467-789X.2009.00653.x](https://doi.org/10.1111/j.1467-789X.2009.00653.x)] [Medline: [19744231](https://pubmed.ncbi.nlm.nih.gov/19744231/)]
11. Sartor F, Vernillo G, de Morree HM, Bonomi AG, La Torre A, Kubis H, et al. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. *Sports Med* 2013;43(9):865-873. [doi: [10.1007/s40279-013-0068-3](https://doi.org/10.1007/s40279-013-0068-3)] [Medline: [23821468](https://pubmed.ncbi.nlm.nih.gov/23821468/)]
12. Holko M, Litwin TR, Munoz F, Theisz KI, Salgin L, Jenks NP, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. *NPJ Digit Med* 2022;5(1):53 [FREE Full text] [doi: [10.1038/s41746-022-00593-x](https://doi.org/10.1038/s41746-022-00593-x)] [Medline: [35469045](https://pubmed.ncbi.nlm.nih.gov/35469045/)]
13. Thompson WR. Worldwide survey of fitness trends for 2022. *ACSM's Health Fit J* 2022;26(1):11-20. [doi: [10.1249/FIT.0000000000000732](https://doi.org/10.1249/FIT.0000000000000732)]
14. Ludwig M, Hoffmann K, Endler S, Asteroth A, Wiemeyer J. Measurement, prediction, and control of individual heart rate responses to exercise-basics and options for wearable devices. *Front Physiol* 2018;9:778 [FREE Full text] [doi: [10.3389/fphys.2018.00778](https://doi.org/10.3389/fphys.2018.00778)] [Medline: [29988588](https://pubmed.ncbi.nlm.nih.gov/29988588/)]
15. Automated fitness level (VO₂max): estimation with heart rate and speed data. Firstbeat Technologies Ltd. 2014. URL: https://assets.firstbeat.com/firstbeat/uploads/2017/06/white_paper_VO2max_30.6.2017.pdf [accessed 2024-07-12]
16. Year in review 2019: firstbeat sports, fitness, and lifestyle wearables. Firstbeat. URL: <https://www.firstbeat.com/en/blog/year-in-review-2019-firstbeat-sports-fitness-and-lifestyle-wearables/> [accessed 2024-07-12]
17. Kwon SB, Ahn JW, Lee SM, Lee J, Lee D, Hong J, et al. Estimating maximal oxygen uptake from daily activity data measured by a watch-type fitness tracker: cross-sectional study. *JMIR Mhealth Uhealth* 2019;7(6):e13327 [FREE Full text] [doi: [10.2196/13327](https://doi.org/10.2196/13327)] [Medline: [31199336](https://pubmed.ncbi.nlm.nih.gov/31199336/)]
18. Cook AJ, Ng B, Gargiulo GD, Hindmarsh D, Pitney M, Lehmann T, et al. Instantaneous VO₂ from a wearable device. *Med Eng Phys* 2018;52:41-48. [doi: [10.1016/j.medengphy.2017.12.008](https://doi.org/10.1016/j.medengphy.2017.12.008)] [Medline: [29373233](https://pubmed.ncbi.nlm.nih.gov/29373233/)]
19. Bonomi AG, Ten Hoor GA, de Morree HM, Plasqui G, Sartor F. Cardiorespiratory fitness estimation from heart rate and body movement in daily life. *J Appl Physiol* (1985) 2020;128(3):493-500 [FREE Full text] [doi: [10.1152/jappphysiol.00631.2019](https://doi.org/10.1152/jappphysiol.00631.2019)] [Medline: [31999530](https://pubmed.ncbi.nlm.nih.gov/31999530/)]

20. Webster DE, Tummalacherla M, Higgins M, Wing D, Ashley E, Kelly VE, et al. Smartphone-based VO₂max measurement with heart snapshot in clinical and real-world settings with a diverse population: validation study. *JMIR Mhealth Uhealth* 2021;9(6):e26006 [FREE Full text] [doi: [10.2196/26006](https://doi.org/10.2196/26006)] [Medline: [34085945](https://pubmed.ncbi.nlm.nih.gov/34085945/)]
21. Muntaner-Mas A, Martinez-Nicolas A, Quesada A, Cadenas-Sanchez C, Ortega FB. Smartphone app (2kmFIT-App) for measuring cardiorespiratory fitness: validity and reliability study. *JMIR Mhealth Uhealth* 2021;9(1):e14864 [FREE Full text] [doi: [10.2196/14864](https://doi.org/10.2196/14864)] [Medline: [33416503](https://pubmed.ncbi.nlm.nih.gov/33416503/)]
22. Muntaner-Mas A, Martinez-Nicolas A, Lavie CJ, Blair SN, Ross R, Arena R, et al. A systematic review of fitness apps and their potential clinical and sports utility for objective and remote assessment of cardiorespiratory fitness. *Sports Med* 2019;49(4):587-600 [FREE Full text] [doi: [10.1007/s40279-019-01084-y](https://doi.org/10.1007/s40279-019-01084-y)] [Medline: [30825094](https://pubmed.ncbi.nlm.nih.gov/30825094/)]
23. Bent B, Dunn JP. Wearables in the SARS-CoV-2 pandemic: what are they good for? *JMIR Mhealth Uhealth* 2020;8(12):e25137 [FREE Full text] [doi: [10.2196/25137](https://doi.org/10.2196/25137)] [Medline: [33315580](https://pubmed.ncbi.nlm.nih.gov/33315580/)]
24. Kraft GL, Roberts RA. Validation of the Garmin Forerunner 920XT fitness watch VO₂peak test. *Int J Innov Educ Res* 2017;5(2):63-69. [doi: [10.31686/ijer.vol5.iss2.619](https://doi.org/10.31686/ijer.vol5.iss2.619)]
25. Helm M, Carrier B, Davis D, Cruz K, Barrios B, Navalta J. Validation of the Garmin Fenix 6S maximal oxygen consumption (VO₂max) estimate. *Int J Exerc Sci Conf Proc* 2021;14(1):29 [FREE Full text]
26. Kraft GL, Dow M. Validation of the Garmin Forerunner 920XT VO₂max estimation and the polar RS300X fitness test. *Int J Innov Educ Res* 2019;7(9):22-28. [doi: [10.31686/ijer.vol7.iss9.1658](https://doi.org/10.31686/ijer.vol7.iss9.1658)]
27. Klepin K, Wing D, Higgins M, Nichols J, Godino JG. Validity of cardiorespiratory fitness measured with fitbit compared to VO₂max. *Med Sci Sports Exerc* 2019;51(11):2251-2256 [FREE Full text] [doi: [10.1249/mss.0000000000002041](https://doi.org/10.1249/mss.0000000000002041)]
28. Cooper KD, Shafer AB. Validity and reliability of the Polar A300's fitness test feature to predict VO₂max. *Int J Exerc Sci* 2019;12(4):393-401 [FREE Full text] [Medline: [30899351](https://pubmed.ncbi.nlm.nih.gov/30899351/)]
29. Muthusamy S, Subramaniam A, Balasubramanian K, Purushothaman VK, Vasanthi RK. Assessment of VO₂ max reliability with Garmin smart watch among swimmers. *Int J Life Sci Pharm Res* 2021;11:42-46. [doi: [10.22376/ijpbs/lpr.2021.11.4.142-46](https://doi.org/10.22376/ijpbs/lpr.2021.11.4.142-46)]
30. Passler S, Bohrer J, Blöching L, Senner V. Validity of wrist-worn activity trackers for estimating VO₂max and energy expenditure. *Int J Environ Res Public Health* 2019;16(17):3037 [FREE Full text] [doi: [10.3390/ijerph16173037](https://doi.org/10.3390/ijerph16173037)] [Medline: [31443347](https://pubmed.ncbi.nlm.nih.gov/31443347/)]
31. Esco MR, Snarr RL, Williford HN. Monitoring changes in VO₂max via the Polar FT40 in female collegiate soccer players. *J Sports Sci* 2014;32(11):1084-1090. [doi: [10.1080/02640414.2013.879672](https://doi.org/10.1080/02640414.2013.879672)] [Medline: [24506090](https://pubmed.ncbi.nlm.nih.gov/24506090/)]
32. Wagner M, Engel F, Klier K, Klughardt S, Wallner F, Wiecek A. On the reliability of wearable devices using the example of a premium multisport smartwatch. *Ger J Exerc Sport Res* 2021;51(1):49-62. [doi: [10.1007/s12662-020-00682-7](https://doi.org/10.1007/s12662-020-00682-7)]
33. Carrier B, Creer A, Williams L. Validation of Garmin Fenix 3 HR fitness tracker biomechanics and metabolics (VO₂max). *J Meas Phys Behav* 2020;3(4):331-337. [doi: [10.1123/jmpb.2019-0066](https://doi.org/10.1123/jmpb.2019-0066)]
34. Foley R. Assessing the validity of the Garmin Venu SQ for estimating VO₂max [Thesis].: Northwest University; 2022. URL: <http://archives.northwestu.edu/handle/nu/58749> [accessed 2024-03-18]
35. Düking P, Van Hooren B, Sperlich B. Assessment of peak oxygen uptake with a smartwatch and its usefulness for training of runners. *Int J Sports Med* 2022;43(7):642-647 [FREE Full text] [doi: [10.1055/a-1686-9068](https://doi.org/10.1055/a-1686-9068)] [Medline: [35094376](https://pubmed.ncbi.nlm.nih.gov/35094376/)]
36. Market share of smartwatch unit shipments worldwide from the 2nd quarter 2014 to 2nd quarter 2021, by vendor. Statista. URL: <https://www.statista.com/statistics/524830/global-smartwatch-vendors-market-share/> [accessed 2024-03-18]
37. Le S, Wang X, Zhang T, Lei SM, Cheng S, Yao W, et al. Validity of three smartwatches in estimating energy expenditure during outdoor walking and running. *Front Physiol* 2022;13:995575 [FREE Full text] [doi: [10.3389/fphys.2022.995575](https://doi.org/10.3389/fphys.2022.995575)] [Medline: [36225296](https://pubmed.ncbi.nlm.nih.gov/36225296/)]
38. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. *JMIR Mhealth Uhealth* 2019;7(3):e10828 [FREE Full text] [doi: [10.2196/10828](https://doi.org/10.2196/10828)] [Medline: [30855232](https://pubmed.ncbi.nlm.nih.gov/30855232/)]
39. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol* 2017;2(1):104-106. [doi: [10.1001/jamacardio.2016.3340](https://doi.org/10.1001/jamacardio.2016.3340)] [Medline: [27732703](https://pubmed.ncbi.nlm.nih.gov/27732703/)]
40. Hernando D, Roca S, Sancho J, Alesanco Á, Bailón R. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Sensors (Basel)* 2018;18(8):2619 [FREE Full text] [doi: [10.3390/s18082619](https://doi.org/10.3390/s18082619)] [Medline: [30103376](https://pubmed.ncbi.nlm.nih.gov/30103376/)]
41. Abt G, Bray J, Benson AC. Measuring moderate-intensity exercise with the apple watch: validation study. *JMIR Cardio* 2018;2(1):e6 [FREE Full text] [doi: [10.2196/cardio.8574](https://doi.org/10.2196/cardio.8574)] [Medline: [31758766](https://pubmed.ncbi.nlm.nih.gov/31758766/)]
42. Using apple watch to estimate cardio fitness with VO₂max. Apple Inc. 2021. URL: https://www.apple.com/healthcare/docs/site/Using_Apple_Watch_to_Estimate_Cardio_Fitness_with_VO2_max.pdf [accessed 2024-07-12]
43. Pokan R, Förster H, Hofmann P, Hörtnagl H, Ledl-Kurkowski E, Wonisch M, editors. *Kompandium der Sportmedizin: Physiologie, Innere Medizin und Pädiatrie*. Austria: Springer-Verlag; 2004.
44. VO₂ Master. URL: <https://vo2master.com/> [accessed 2024-03-18]
45. Montoye AHK, Vondrasek JD, Hancock JB. Validity and reliability of the VO₂ master pro for oxygen consumption and ventilation assessment. *Int J Exerc Sci* 2020;13(4):1382-1401 [FREE Full text] [Medline: [33042375](https://pubmed.ncbi.nlm.nih.gov/33042375/)]

46. Webber J. Validity and Reliability of the VO2 Master. VO2 Master. 2019. URL: <https://vo2master.com/blog/validity-and-reliability-of-the-vo2-master/> [accessed 2024-03-21]
47. Polar H10. Polar Electro. URL: <https://www.polar.com/en/sensors/h10-heart-rate-sensor> [accessed 2024-03-18]
48. Cycle 407/457. ERGOFIT. URL: <https://www.ergo-fit.de/en/professional/products/cardio/cycle-407/457/> [accessed 2018-03-24]
49. Klingenheben T, Löllgen H, Bosch R, Trappe H. Manual on the clinical use of ergometry. *Kardiologie* 2018;12(5):342-355. [doi: [10.1007/s12181-018-0265-2](https://doi.org/10.1007/s12181-018-0265-2)]
50. Fairbairn MS, Blackie SP, McElvaney NG, Wiggs BR, Paré PD, Pardy RL. Prediction of heart rate and oxygen uptake during incremental and maximal exercise in healthy adults. *Chest* 1994;105(5):1365-1369. [doi: [10.1378/chest.105.5.1365](https://doi.org/10.1378/chest.105.5.1365)] [Medline: [8181321](https://pubmed.ncbi.nlm.nih.gov/8181321/)]
51. Fox SM, Naughton JP. Physical activity and the prevention of coronary heart disease. *Prev Med* 1972;1(1):92-120. [doi: [10.1016/0091-7435\(72\)90079-5](https://doi.org/10.1016/0091-7435(72)90079-5)] [Medline: [5069016](https://pubmed.ncbi.nlm.nih.gov/5069016/)]
52. Cleary M, Hetzler R, Wages J, Lentz M, Stickley C, Kimura I. Comparisons of age-predicted maximum heart rate equations in college-aged subjects. *J Strength Cond Res* 2011;25(9):2591-2597. [doi: [10.1519/JSC.0b013e3182001832](https://doi.org/10.1519/JSC.0b013e3182001832)] [Medline: [21691228](https://pubmed.ncbi.nlm.nih.gov/21691228/)]
53. Adams R. Revised physical activity readiness questionnaire. *Can Fam Physician* 1999;45:992, 995, 1004-992, 995, 1005 [FREE Full text] [Medline: [10216799](https://pubmed.ncbi.nlm.nih.gov/10216799/)]
54. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39(2):175-191. [doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146)] [Medline: [17695343](https://pubmed.ncbi.nlm.nih.gov/17695343/)]
55. Klein R. Bland-altman and correlation plot. MATLAB. URL: <https://de.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot> [accessed 2024-03-18]
56. Matthew R. f_ICC. MATLAB. URL: https://de.mathworks.com/matlabcentral/fileexchange/66866-f_icc [accessed 2024-03-18]
57. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *J R Stat Soc D* 1983;32(3):307-317. [doi: [10.2307/2987937](https://doi.org/10.2307/2987937)]
58. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 2015;25(2):141-151 [FREE Full text] [doi: [10.11613/BM.2015.015](https://doi.org/10.11613/BM.2015.015)] [Medline: [26110027](https://pubmed.ncbi.nlm.nih.gov/26110027/)]
59. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155-163 [FREE Full text] [doi: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)] [Medline: [27330520](https://pubmed.ncbi.nlm.nih.gov/27330520/)]
60. Mean absolute percentage error between arrays. MATLAB. URL: <https://de.mathworks.com/help/matlab/ref/mape.html> [accessed 2024-03-18]
61. Root-mean-square error between arrays. MATLAB. URL: <https://de.mathworks.com/help/matlab/ref/rmse.html> [accessed 2024-03-18]
62. Rosnow RL, Rosenthal R. Effect sizes for experimenting psychologists. *Can J Exp Psychol* 2003;57(3):221-237. [doi: [10.1037/h0087427](https://doi.org/10.1037/h0087427)] [Medline: [14596479](https://pubmed.ncbi.nlm.nih.gov/14596479/)]
63. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. UK: Routledge; 1988.
64. Kaminsky LA, Arena R, Myers J. Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing: data from the fitness registry and the importance of exercise national database. *Mayo Clin Proc* 2015;90(11):1515-1523 [FREE Full text] [doi: [10.1016/j.mayocp.2015.07.026](https://doi.org/10.1016/j.mayocp.2015.07.026)] [Medline: [26455884](https://pubmed.ncbi.nlm.nih.gov/26455884/)]
65. Freeberg KA, Baughman BR, Vickey T, Sullivan JA, Sawyer BJ. Assessing the ability of the fitbit charge 2 to accurately predict VO2max. *mHealth* 2019;5:39 [FREE Full text] [doi: [10.21037/mhealth.2019.09.07](https://doi.org/10.21037/mhealth.2019.09.07)] [Medline: [31620466](https://pubmed.ncbi.nlm.nih.gov/31620466/)]
66. Carlson DJ. VO2max: the gold standard? *Chest* 1995;108(3):602-603. [doi: [10.1378/chest.108.3.602](https://doi.org/10.1378/chest.108.3.602)]
67. Midgley AW, McNaughton LR, Polman R, Marchant D. Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Med* 2007;37(12):1019-1028. [doi: [10.2165/00007256-200737120-00002](https://doi.org/10.2165/00007256-200737120-00002)] [Medline: [18027991](https://pubmed.ncbi.nlm.nih.gov/18027991/)]
68. Edvardsen E, Hem E, Anderssen SA. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *PLoS One* 2014;9(1):e85276 [FREE Full text] [doi: [10.1371/journal.pone.0085276](https://doi.org/10.1371/journal.pone.0085276)] [Medline: [24454832](https://pubmed.ncbi.nlm.nih.gov/24454832/)]
69. Price S, Wiecha S, Cieśliński I, Śliż D, Kasiak PS, Lach J, et al. Differences between treadmill and cycle ergometer cardiopulmonary exercise testing results in triathletes and their association with body composition and body mass index. *Int J Environ Res Public Health* 2022;19(6):3557 [FREE Full text] [doi: [10.3390/ijerph19063557](https://doi.org/10.3390/ijerph19063557)] [Medline: [35329246](https://pubmed.ncbi.nlm.nih.gov/35329246/)]
70. Noonan V, Dean E. Submaximal exercise testing: clinical application and interpretation. *Phys Ther* 2000;80(8):782-807. [doi: [10.1093/ptj/80.8.782](https://doi.org/10.1093/ptj/80.8.782)]
71. Goodwin ML, Harris JE, Hernández A, Gladden LB. Blood lactate measurements and analysis during exercise: a guide for clinicians. *J Diabetes Sci Technol* 2007;1(4):558-569 [FREE Full text] [doi: [10.1177/193229680700100414](https://doi.org/10.1177/193229680700100414)] [Medline: [19885119](https://pubmed.ncbi.nlm.nih.gov/19885119/)]

Abbreviations

HR: heart rate

ICC: intraclass correlation coefficient

MAPE: mean absolute percentage error

RMSE: root-mean-square error

VO2max: maximum oxygen uptake

Edited by T Leung; submitted 15.04.24; peer-reviewed by J Shaikh-Mohammed; comments to author 23.06.24; revised version received 28.06.24; accepted 30.06.24; published 31.07.24.

Please cite as:

Caserman P, Yum S, Göbel S, Reif A, Matura S

Assessing the Accuracy of Smartwatch-Based Estimation of Maximum Oxygen Uptake Using the Apple Watch Series 7: Validation Study

JMIR Biomed Eng 2024;9:e59459

URL: <https://biomedeng.jmir.org/2024/1/e59459>

doi: [10.2196/59459](https://doi.org/10.2196/59459)

PMID:

©Polona Caserman, Sungsoo Yum, Stefan Göbel, Andreas Reif, Silke Matura. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 31.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

A Deep Learning Framework for Predicting Patient Decannulation on Extracorporeal Membrane Oxygenation Devices: Development and Model Analysis Study

Joshua Fuller¹, BSc; Alexey Abramov², MD; Dana Mullin³, MS, CCP; James Beck³, MS, CCP; Philippe Lemaitre², MD, PhD; Elham Azizi^{4,5,6,7}, PhD

¹Vagelos College of Physicians and Surgeons, Columbia University, New York City, NY, United States

²Department of Surgery, Columbia University Irving Medical Center, New York, NY, United States

³Clinical Perfusion, New York Presbyterian Hospital, New York, NY, United States

⁴Department of Biomedical Engineering, Columbia University, New York City, NY, United States

⁵Irving Institute for Cancer Dynamics, Columbia University, New York, NY, United States

⁶Department of Computer Science, Columbia University, New York, NY, United States

⁷Data Science Institute, Columbia University, New York, NY, United States

Corresponding Author:

Elham Azizi, PhD

Department of Biomedical Engineering

Columbia University

500 W 120th St

Engineering Terrace 351

New York City, NY, 10027

United States

Phone: 1 (212) 851 0271

Email: ea2690@columbia.edu

Abstract

Background: Venovenous extracorporeal membrane oxygenation (VV-ECMO) is a therapy for patients with refractory respiratory failure. The decision to decannulate someone from extracorporeal membrane oxygenation (ECMO) often involves weaning trials and clinical intuition. To date, there are limited prognostication metrics to guide clinical decision-making to determine which patients will be successfully weaned and decannulated.

Objective: This study aims to assist clinicians with the decision to decannulate a patient from ECMO, using Continuous Evaluation of VV-ECMO Outcomes (CEVVO), a deep learning-based model for predicting success of decannulation in patients supported on VV-ECMO. The running metric may be applied daily to categorize patients into high-risk and low-risk groups. Using these data, providers may consider initiating a weaning trial based on their expertise and CEVVO.

Methods: Data were collected from 118 patients supported with VV-ECMO at the Columbia University Irving Medical Center. Using a long short-term memory-based network, CEVVO is the first model capable of integrating discrete clinical information with continuous data collected from an ECMO device. A total of 12 sets of 5-fold cross validations were conducted to assess the performance, which was measured using the area under the receiver operating characteristic curve (AUROC) and average precision (AP). To translate the predicted values into a clinically useful metric, the model results were calibrated and stratified into risk groups, ranging from 0 (high risk) to 3 (low risk). To further investigate the performance edge of CEVVO, 2 synthetic data sets were generated using Gaussian process regression. The first data set preserved the long-term dependency of the patient data set, whereas the second did not.

Results: CEVVO demonstrated consistently superior classification performance compared with contemporary models ($P < .001$ and $P = .04$ compared with the next highest AUROC and AP). Although the model's patient-by-patient predictive power may be too low to be integrated into a clinical setting (AUROC 95% CI 0.6822-0.7055; AP 95% CI 0.8515-0.8682), the patient risk classification system displayed greater potential. When measured at 72 hours, the high-risk group had a successful decannulation rate of 58% (7/12), whereas the low-risk group had a successful decannulation rate of 92% (11/12; $P = .04$). When measured at 96 hours, the high- and low-risk groups had a successful decannulation rate of 54% (6/11) and 100% (9/9), respectively ($P = .01$). We hypothesized that the improved performance of CEVVO was owing to its ability to efficiently capture transient temporal

patterns. Indeed, CEVVO exhibited improved performance on synthetic data with inherent temporal dependencies ($P < .001$) compared with logistic regression and a dense neural network.

Conclusions: The ability to interpret and integrate large data sets is paramount for creating accurate models capable of assisting clinicians in risk stratifying patients supported on VV-ECMO. Our framework may guide future incorporation of CEVVO into more comprehensive intensive care monitoring systems.

(*JMIR Biomed Eng* 2024;9:e48497) doi:[10.2196/48497](https://doi.org/10.2196/48497)

KEYWORDS

extracorporeal membrane oxygenation; ECMO; venovenous; VV; machine learning; supervised learning; dynamic data; time series; clinical decision support; artificial intelligence; AI; clinical AI; health informatics

Introduction

Background

Extracorporeal life support (ECLS) is a suite of resource-intensive therapies indicated in patients with refractory respiratory failure or cardiogenic shock [1]. This intervention involves cannulation of central or peripheral arteries and veins to provide forward flow through a circuit with a mechanical pump and gas exchange device, also called a membrane oxygenator. Air is connected to the membrane oxygenator to deliver oxygen and remove carbon dioxide from the circulating blood. Established indications for venovenous extracorporeal membrane oxygenation (VV-ECMO) exist in the literature, and the use of this technology was expanded during the COVID-19 pandemic [2]. The VV-ECMO configuration is specifically used for patients experiencing severe lung injury. This setup is designed to provide oxygenation and decarboxylation support without offering the additional hemodynamic assistance found in the venoarterial configuration. VV-ECMO is considered a last resort therapy for patients with end-stage respiratory failure [3], with an overall survival rate of 60% [4].

Decannulating a patient from VV-ECMO is a clinical challenge that requires considerable training and expertise from provider teams in the intensive care unit. Clinicians assess trends in the patient's vital signs, physical examination, response to various therapies, laboratory biochemistries, and radiographic studies. When the decision is made to proceed, decannulation is usually accomplished through a weaning trial during which VV-ECMO is gradually reduced. To date, there are limited prognostication scores that successfully predict when patients are ready to undergo a weaning trial. In this study, we present an artificial intelligence model capable of running in real time that incorporates discrete and continuous variables that clinicians may use in their assessment of patients for decannulation from VV-ECMO support.

Related Work

Multiple predictive scores have been developed to help clinicians prognosticate before cannulation. The 6 most common prognostication scores for adult respiratory failure supported on extracorporeal membrane oxygenation (ECMO) are ECMOnet, Predicting Death for Severe Ards on VV-ECMO, Respiratory ECMO Survival Prediction, Roch, Venovenous ECMO mortality score, and Prediction of Survival on ECMO Therapy score [5] (Table S1 in [Multimedia Appendix 1](#) [6-11]). Although these 6 scores are commonly used, they have 2 main

drawbacks. First, all input information is recorded before cannulation to ECLS because the primary intent of the models is to be used to determine which candidates were most likely to benefit from the intervention. Second, all scores use logistic regression to predict outcomes or identify significant variables. Logistic regression requires high-quality data from static variables, which limits the types of data that can be inputted. Thus, sequential or time-series data such as laboratory values and vital signs must be limited to a single time point or summarized. Furthermore, these statistical models are limited in terms of capturing nonlinear effects and interactions between variables.

To date, no studies have mitigated both issues to improve the prediction of successful decannulation in patients supported on VV-ECMO. However, some researchers have attempted to use deep learning to predict specific clinical events.

Abbasi et al [12] used clinical and ECLS data to compare 2 approaches, deep learning and traditional statistical methodology, to develop a model to predict hemorrhage and thrombosis events. The deep learning model outperformed linear regression in both hemorrhage and thrombosis data sets, suggesting that more complex models may achieve better predictive power. Other authors have applied deep learning and modified logistic regression to predict survival on venoarterial-ECMO (VA-ECMO) only. Ayers et al [13] used 48 hours of laboratory values after VA-ECMO cannulation to predict survival to discharge using a deep neural network.

Similarly, Loyaga et al [14] used clinical, echocardiographic, laboratory, and hemodynamic characteristics to predict 30-day mortality in patients on VA-ECMO using the elastic-net method. None of these studies used data obtained from the ECMO devices, and instead used laboratory values, clinical scores, and disease severity to train their models. These approaches leave a large amount of valuable information unused. In the hospital, clinicians adjust the parameters of ECMO support in real time according to the patient's condition and pathophysiology. Modern devices capture the interplay between the patient and ECMO by continuously collecting perfusion data [15]. Analysis of information-dense perfusion data may be leveraged to improve the prediction accuracy of clinically meaningful outcomes in ECLS care.

Incorporating more granular data requires a new model that is capable of integrating categorical and time-series data. The prevalence of recurrent neural networks (RNNs) in health care data science has increased recently. The ability of RNNs to

efficiently understand time dependencies makes this approach beneficial in certain types of medical data, such as ventilator settings [16], vital signs [17], medication administration [18], imaging studies [19], and radiology reports [20]. One type of RNN, long short-term memory (LSTM), is specifically designed for long time series, such as our data set with weeks-long hospital courses. LSTM can encode these time series into a compressed latent space, which can be concatenated with static variables, such as age, gender, and other clinical characteristics.

Novelty

The innovation of our study is two-fold: (1) data source and (2) algorithm design. Perfusion data were collected from the ECMO devices and recorded at highly granular intervals. Our analysis sheds new light on the effectiveness of ECMO. Second, unlike prior work using laboratory values, clinical scores, and other static data, the patient information used in our study was both dynamic and static. Using a 2-headed neural network, our predictive algorithm efficiently incorporates static information, such as sex and clinical scores, along with dynamic data. LSTM networks encode the perfusion time-series data into a latent space, which is then concatenated with an encoding of the static variables. This new latent space was used to classify patients.

We present the Continuous Evaluation of VV-ECMO Outcomes (CEVVO) predictive model for determining successful decannulation from VV-ECMO using both pre- and postcannulation data. When using both, the model can continuously update its prediction, providing a running measure for patient potential recovery. Such a measure may help clinicians and patient families make more informed decisions about care. Using synthetic data sets, we demonstrate that understanding time dependence is the essential ingredient to accurate predictions. Our framework also guides the categorization of patients into high-risk and low-risk groups, alerting care providers about which patients may be better candidates for weaning trials and decannulation.

Methods

Problem Formulation

Health care data of this type can be presented in two components: (1) clinical information that remains unchanged over the ECMO course, such as age and sex, which are considered static features, and (2) variables that change over time, such as laboratory values and perfusion data, which are considered temporal variables. This study follows the conventions presented in the study by Yoon et al [21]. We define S as a vector space of static features, and X as a vector space of temporal features. Let $S \in S$ and $X \in X$ be random vectors with specific values denoted by s and x . Each patient is a tuple of $(s, x_{1:T})$, where T is the number of time steps. For clarity, patients in our training set were indexed by $n \in 1, \dots, N$. Therefore, the training data set is denoted as $D = (s_n, x_{n,1:T})_{n=1}^N$. Each patient also had a categorical outcome $y \in \{0, 1\}$, which forms vector Y across all patients, with 0 representing unsuccessful decannulation and 1 representing success. We define the probability distribution $p(Y|S, X_{1:T})$, and our goal is to use training data D to learn a density $\hat{p}(Y|S, X_{1:T})$ that best

approximates $p(Y|S, X_{1:T})$. This is achieved through the optimization in equation 1:

$$\text{Min}_{\hat{p}} D_{KL}(p(Y|S, X_{1:T}) \parallel \hat{p}(Y|S, X_{1:T})) \quad (1)$$

The abovementioned Kullback-Leibler divergence can be calculated through the loss function in equation 2. This is identical to the cross-entropy because the entropy of the ground truth distribution is 0. The model can best approximate the true distribution by using backpropagation to minimize equation 2:

$$L = (-1/N) \sum_{n=1}^N (y_n \log(\hat{y}_n) + (1-y_n) \log(1-\hat{y}_n)) \quad (2)$$

Synthetic Data Set

We hypothesized that the high performance of the LSTM-based architecture is owing to its superior ability to capture long-term dependencies in the data set. To test this notion, 2 synthetic data sets of size $N=234$ and $t=2054$ were generated using a Gaussian process regression (GPR) model [22]. As GPR is nonparametric, it can generate synthetic data without making assumptions about the underlying relationships between variables and dynamics over time. By tuning parameters of the generative model, we can adjust the strength of long-term dependencies in the data. Using the GPR model, we sample data from a multivariable normal distribution, in which the covariance encodes dependencies between time points as shown in equation 3:

$$f \sim \mathcal{N}(\mu, \Sigma) \quad (3)$$

where μ denotes the expected values of the inputs and Σ denotes the covariance. The covariance is encoded by a radial basis function (RBF) kernel, as shown in equation 4. The length scale parameter L of the RBF adjusts the local smoothing. Higher values for this parameter encode dependencies over a longer period, leading to smoother dynamics.

$$k(x_i, x_j) = \exp(-(d(x_i, x_j)^2 / (2L))) \quad (4)$$

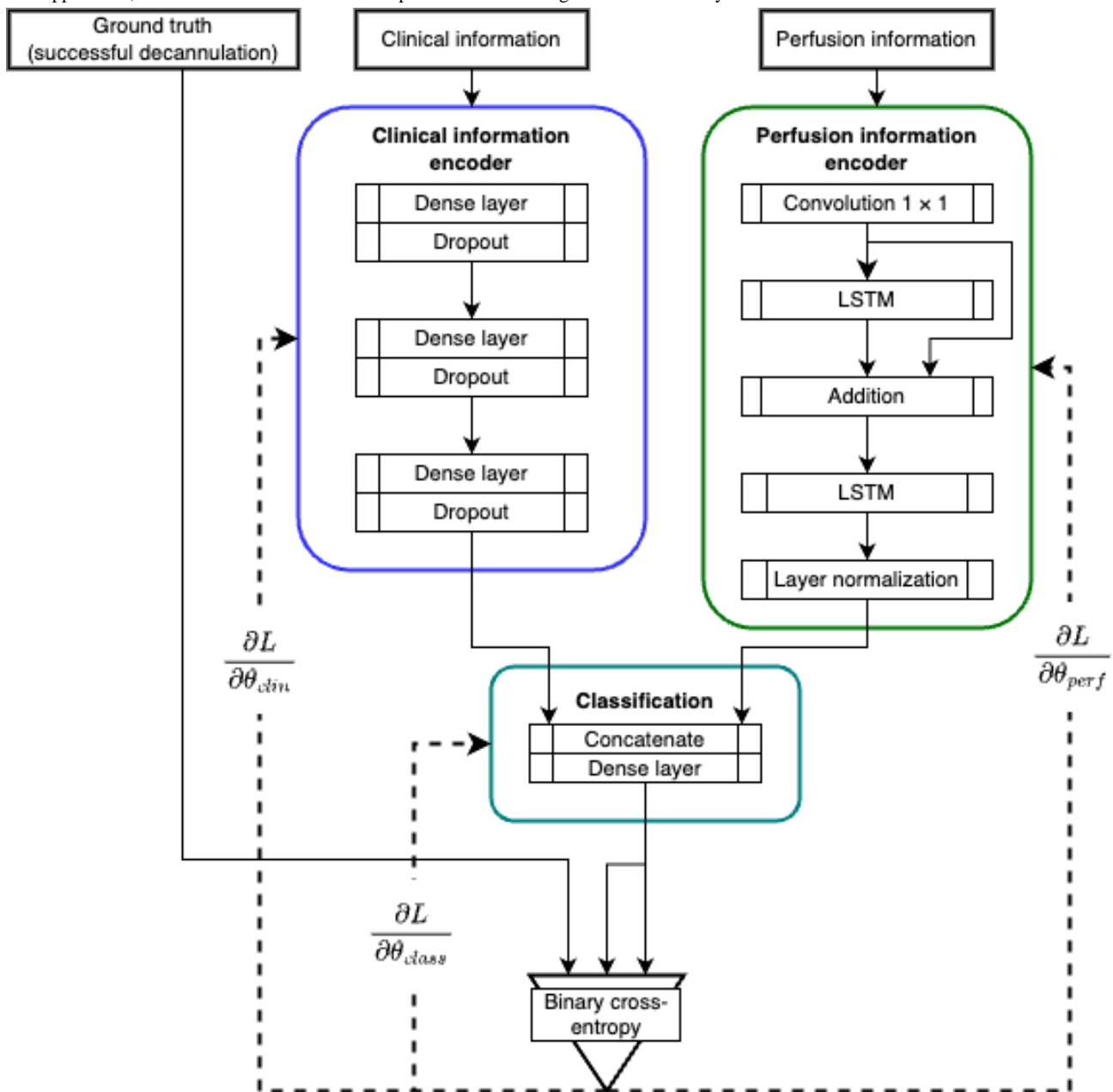
where $d(x_i, x_j)$ denotes the Euclidean distance. The long-term dependencies are captured by the probability of observing specific values conditioned on earlier time points. This assumption is reasonable in our application to VV-ECMO and not necessarily held in previous models such as logistic regression and some deep neural networks. The performance of previous models on GPR data is thus not affected by different choices of length scale, whereas the LSTM-based model should lose its advantage with increasing length scale.

Two groups of synthetic data were created: the first with $L=1$ and the second with $L=100$, and it was expected that CEVVO would be the only one to perform substantially better on $L=1$. The other models should have similar performance between $L=1$ and $L=100$. The length scale had to be larger than or equal to each time step; therefore, $L=1$ was close to the minimum allowable length scale.

Model Design

A general overview of this framework is presented in Figure 1. It is composed of 2 independent heads: a static (clinical) data encoder and a temporal feature (perfusion) encoder. Theoretically, each distills the relevant information from the 2 data sets (clinical and perfusion) before concatenating them in the classification block.

Figure 1. The overall architecture of the model. The double-headed approach allows the model to integrate static and dynamic data. Solid lines denote function application, and dashed lines denote loss computation. LSTM: long short-term memory.



The static information encoder is based on an autoencoding scheme along with an additional final dense layer. The first dense layer had 32 nodes, the second layer had 33, and the final layer had 25 nodes. These dimensions were chosen via Bayesian optimization hyperparameter tuning implemented through the *Keras* Python package by Chollet et al [23].

The perfusion information encoder was based on LSTM layers. These recurrent networks were found to work exceedingly well, as they were built on the assumption that earlier time points have marginal effects on later time points. A 1×1 convolutional layer was first used to expand the feature map before the LSTM to create a projection shortcut and act as a filter. The *tanh* activation function allowed the convolution layer to increase, decrease, or negate certain input values. Although an additional LSTM layer could do this processing, the convolution layer contained significantly fewer parameters. The filter size for the

convolution and the LSTMs was 1024, which was also chosen via Bayesian optimization hyperparameter tuning.

The classification block concatenated the final outputs of the clinical information encoder and the perfusion encoder. By this point, the original clinical inputs were reduced from 32 to 25, and the original perfusion inputs were reduced from 16,432 to 1024. These 2 final layers were concatenated into a final layer of 1049. This led to a single output neuron with a *sigmoid* activation, which acted as the final prediction. This prediction was then compared with the ground truth, and the loss was calculated using the binary cross-entropy. The average of all the losses was calculated with equation 2. These losses were backpropagated through the network to make the probability distribution generated by the model resemble the reality.

Defining Patient Risk Grouping

For the risk groups to have meaning, the calibration of the model must be assessed. A calibration plot for the training set was created and showed an S-shaped misalignment. The misalignment was corrected using Platt scaling.

Four clinical groups were defined with respect to the calibrated mean and SD of the model's predictions on D . Let M_D and S_D be the mean and SD of the sigmoid output values of training data D . The grouping was determined according to equation 5:

$$f_{\text{group}}(x) = \begin{cases} 0, & \text{if } x \leq M_D - S_D \\ 1, & \text{if } M_D - S_D < x \leq M_D \\ 2, & \text{if } M_D < x \leq M_D + S_D \\ 3, & \text{if } M_D + S_D < x \end{cases} \quad (5)$$

Ethical Considerations

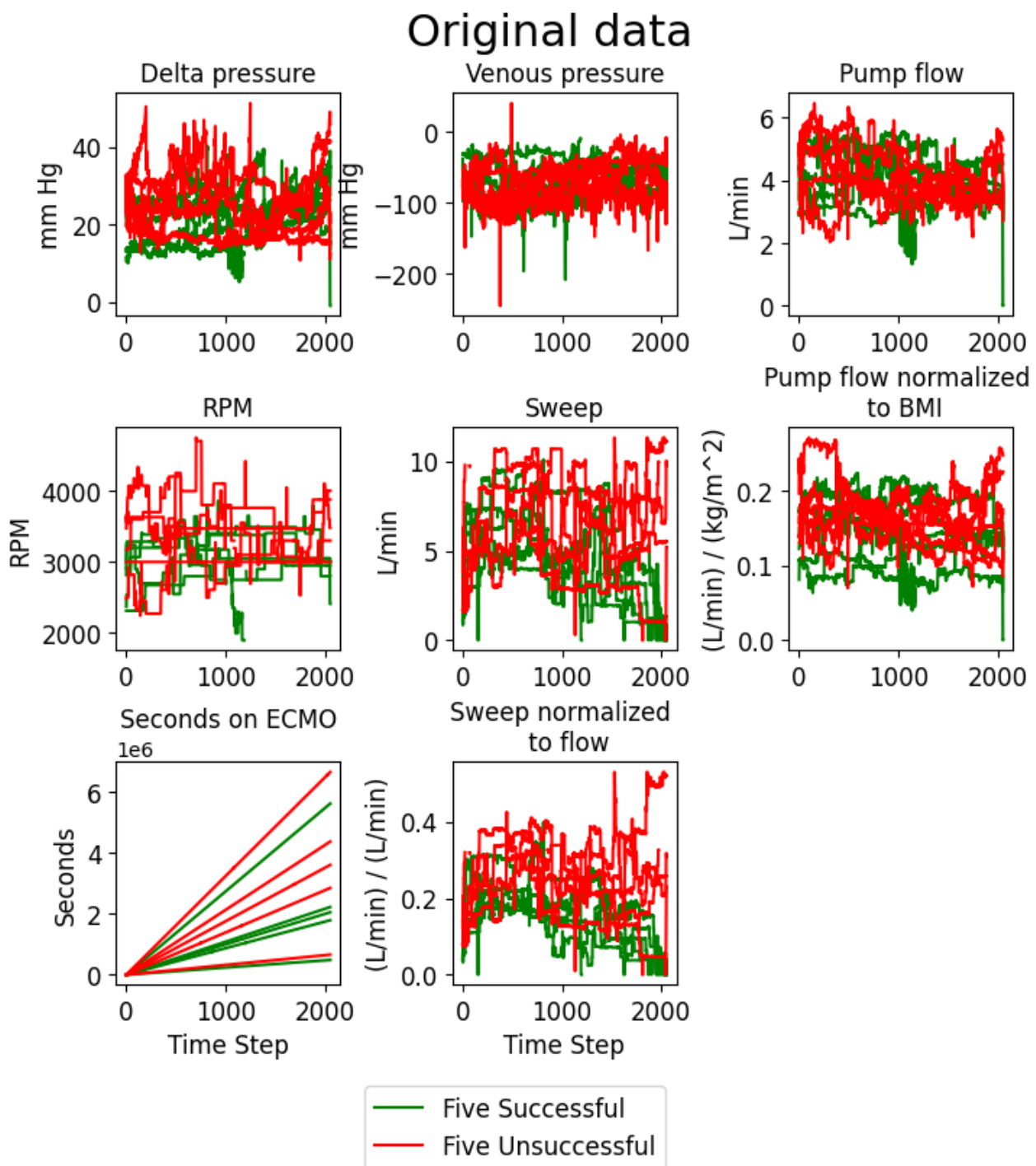
This study was conducted in accordance with the institutional review board of the Columbia University (#AAAT0563).

Data

A retrospective chart review was performed, and continuous perfusion data and clinical information were collected from 118 patients cannulated to VV-ECMO at a high-volume ECMO center's intensive care unit between January 1, 2020, and December 31, 2021. Patients reconfigured to venoarterial-venous or venoarterial were excluded.

Patient data were collected from Spectrum Medical software (Quantum Informatics), which records data from each patient's ECMO machine. Six relevant perfusion variables were selected with expert insight and were collected at 120-second intervals. These were the pressure change across the membrane lung, the venous drainage pressure, the blood flow across the ECMO circuit, the pump head rotation speed (needed to generate the blood flow), the sweep gas flow (rate of oxygenated gas flowing through the membrane lung), and length of time the patient was supported on ECMO. Two additional perfusion variables were created to account for differences between patients: the flow across the pump divided by the patient's BMI and the sweep gas flow divided by the flow across the pump. In addition to these 8 perfusion variables, 12 clinical variables were selected: decannulation result, age, sex, cause of respiratory distress, BMI, cardiac arrest before ECMO, shock (ie, hemodynamic instability) before ECMO, reinfusion and drainage cannulation location, reinfusion and drainage cannula size, and the type of ventilation provided (Table S2 in [Multimedia Appendix 1](#)). The 12 clinical variables included the outcome label, which was not included in the input data. Further clinical information that was not included in the model can be found in Table S3 in [Multimedia Appendix 1](#). An example of 5 successful and 5 unsuccessful patients is shown in [Figure 2](#). The chaotic nature of the perfusion variables helps to justify more advanced machine learning methods.

Figure 2. Extracorporeal membrane oxygenation (ECMO) perfusion data for 5 example patients with successful (green) or unsuccessful (red) decannulation. RPM: revolutions per minute.



To enable incorporation of all time points in each VV-ECMO run, the first preprocessing step involved truncation, which refers to clipping the perfusion data set at different percentages of the total run. For each patient, in addition to the 100% of the ECMO run (ie, the full run), the first 90%, the first 80%,..., the first 10% of the run were appended as additional runs. Thus, the full data set involved 1180 sequences of data points, 10 for each patient. Each data point consisted of a 3D perfusion time series (patient deidentified ID code, time step, and variable) and 2D clinical data (patient deidentified ID code and variable).

Owing to varying ECMO run lengths, each time-series sequence was standardized to 2054 time steps. This length was the largest size possible, given the GPU constraints. Standardization was performed by averaging dense time steps and forward-filling empty steps. The remaining empty time steps were set to 0. Truncations were treated as full runs, that is, the final values for the 10% and the 100% truncation occurred at the same time step—2053. Each truncation is, in effect, stretched over the 2054 time steps. This ensures that the model is not given hints about which truncation it is seeing.

The performance of the model was evaluated through cross-validation. In each iteration, the list of patient IDs was randomized and split into 5 groups of 23 patients, each with 18 successful and 5 unsuccessful patients. Three random patients were excluded to have 5 groups of the same size. Three groups of patients (69/115, 60% of the total) were chosen as the training data set, one group (23/115, 20%) was chosen as the validation set, and one group (23/115, 20%) was chosen as the test set. This process was repeated 5 times until each group had been included in the test set once. The patient list was then randomized again to begin the next cross-fold validation. This ensured that the training set, validation set, test set, and unused patients differed each time. In total, there were 12 iterations of this 5-fold cross-validation.

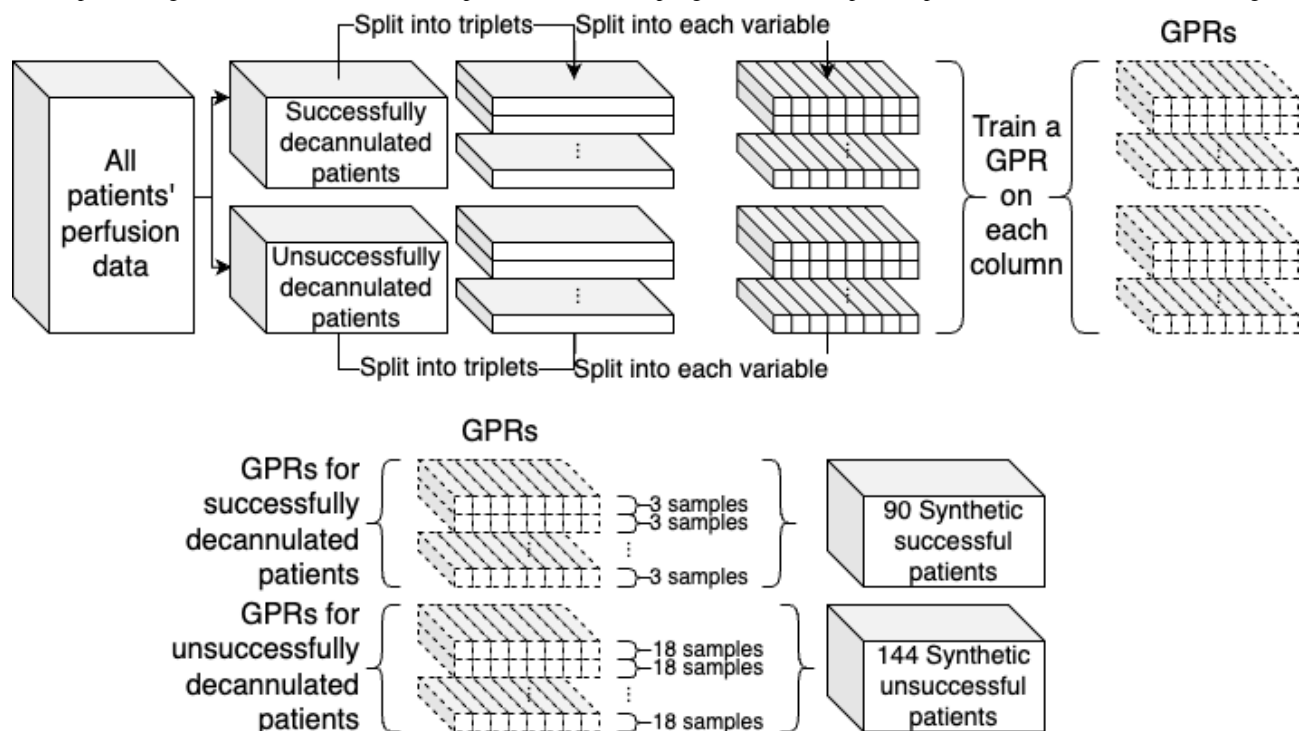
Each set of training data consisted of 69 patients, and the validation and test sets had 23 patients. Including all truncations, the training set had 690 data points, and the validation and test sets had 230 data points each.

The data sets were then scaled using `MinMaxScaler` from the `sklearn` Python package by Pedregosa et al [24]. The scaler was trained on the training data and then used to transform all 3 sets.

Synthetic Data Set

The GPRs were generated using the Gaussian process regressor Python package `sklearn` Pedregosa et al [24]. Two different data sets were generated with different values for the length scale of the RBF kernel (1, 100). A GPR model was first fit to unnormalized perfusion data from all patients. To generate more realistic synthetic data, patients were divided into successful and unsuccessful decannulation groups and sorted according to the ECMO run time. They were then grouped into triplets based on these criteria, which resulted in 30 *successful* triplets and 8 *unsuccessful* triplets. To create the training data for the GPRs, each patient's age, gender, and BMI were extracted and normalized using the `StandardScaler` from the `sklearn` Python package by Pedregosa et al [24]. These, in addition to a time-step value, were treated as independent variables. The dependent variable was the unnormalized perfusion data from the triplet. Each GPR was trained only on a single perfusion variable, so each triplet had 8 GPRs, 1 for each perfusion variable. For each of the 30 *successful* triplets, each GPR model was sampled 3 times for a total of 90 synthetic *successful* patients. For the 8 *unsuccessful* triplets, each GPR was sampled 18 times for a total of 144 synthetic *unsuccessful* patients. A diagram of this process is shown in Figure 3.

Figure 3. Diagram of the process of generating synthetic patient data. Solid boxes indicate patient data (including synthetic), and dotted boxes indicate a Gaussian process regression (GPR) model fit to real patient data. The sampling of the GPRs step was repeated 2 times, 1 for each kernel length scale.



The triplets were then split into training, validation, and test sets in the same manner as the original patient data. For each iteration of the 5 cross-folds validation, the triplets were randomized and split into groups of 7, each with 6 *successful* triplets and 1 *unsuccessful* triplet. A random set of 3 (38%) *unsuccessful* triplets, out of 8, were not included in each interaction of the 5 cross-folds. This was done to guarantee that each group had the same number of triplets. Inside the groups, each of the 6 *successful* triplets yielded 3 synthetic patients,

whereas the 1 *unsuccessful* triplet yielded 18 synthetic patients. This balanced out each group, with a total of 36 synthetic patients per group. Three groups were assigned to the training set, one group was the validation set, and the last group was the test set. Similar to the real data, the test set was rotated until each synthetic patient was tested.

Model Assessment

After each model was trained, predictions were calculated for the test sets. Each prediction varied between 0 and 1 owing to

the *sigmoid* activation in the final neurone. To assess performance, area under the receiver operating characteristic curve (AUROC) and the average precision (AP) were calculated using the *sklearn* package. AP approximated the area under the precision-recall curve. The predictions and ground truths were sampled 5000 times with replacement to create the AUROC and AP CIs. The bootstrapping pseudocode for estimating the AUROC CIs can be found in algorithm S1 in [Multimedia Appendix 1](#). This bootstrapping code was then repeated for different subsets of the data. The AUROC and AP CIs were calculated for each day after cannulation between 0 and 24 (eg, the AUROC and AP for all data points ending on day 10, including truncations). These values were plotted along with their CIs. For the synthetic data, the bootstrapping method was only used on the entirety of each data set.

A successful model is expected to provide accurate and reliable insight into whether a patient will be decannulated.

Results

Model Performance on Real Data

The model's overall performance on the real data achieved an average AUROC of 0.6937 (95% CI 0.6822-0.7055). The mean AP was 0.8599 (95% CI 0.8515-0.8682).

A clinically relevant breakdown is AUROC and AP by day, as shown in [Figures 4 and 5](#). Therefore, we observed that tight CIs begin to expand after day 11 as the number of data points decreased. By limiting the time frame to only include patient data points sampled between days 3 and 11, the AUROC 95% CI was 0.7048-0.7428, and the AP 95% CI was 0.9074-0.9261.

Figure 4. The area under the receiver operating characteristic curve (AUROC; in green) computed from all samples within a 1-day time frame, for example, AUROC for samples collected between days 0 and 1 are shown on day 1. Purple bars indicate the number of data points occurring on that day (right y-axis).

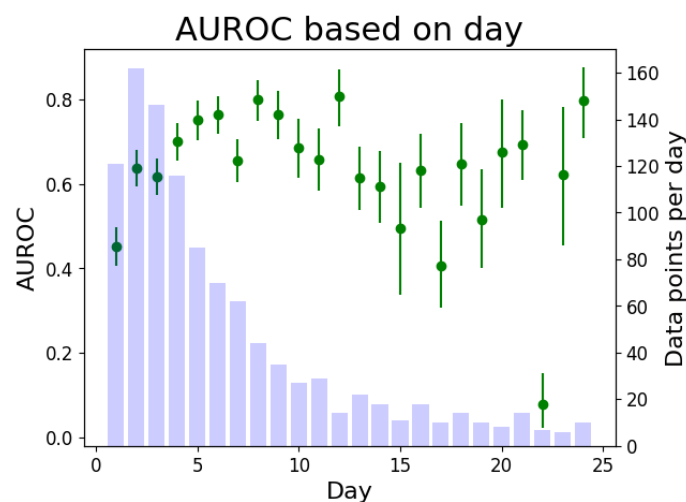
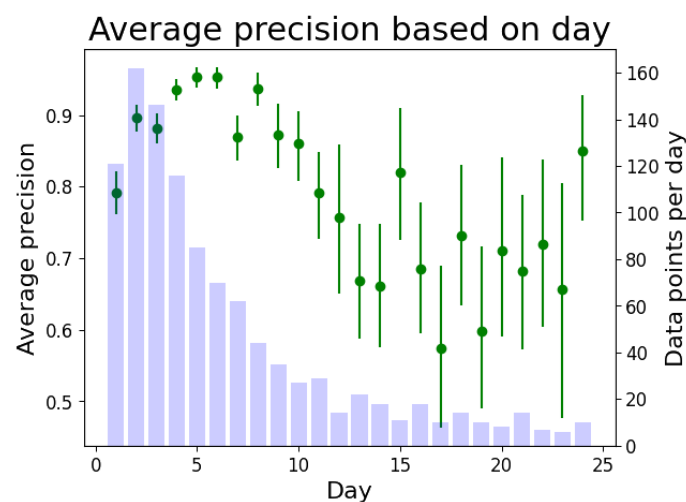


Figure 5. The average precision (AP; in green) computed from all samples within a 1-day time frame, for example, AP for samples collected between days 0 and 1 are shown on day 1. Purple bars indicate the number of data points occurring on that day (right y-axis).



Model Comparison on Real Data

As detailed in Table S2 in [Multimedia Appendix 1](#), ECMOnet, Predicting Death for Severe Ards on VV-ECMO, Respiratory ECMO Survival Prediction, Roch, Venovenous ecmo mortality

score, and Prediction of Survival on ECMO Therapy score rely on either logistic regression or recursive partitioning analysis to determine the patient grouping or scoring classification. To provide a fair comparison with the proposed model, the AUROC and AP calculations were repeated with a logistic regression

model and a decision tree. Both models were trained on the same training data and assessed on the same test data as the proposed model. Moreover, the 95% CIs were determined with the bootstrapping algorithm presented in algorithm S1 in [Multimedia Appendix 1](#). Furthermore, to provide a comparison with the study by Ayers et al [13], a dense neural network was included. Finally, a Naive Bayes model was included to demonstrate the necessity of including dependence between time points. Unlike the LSTM, Naive Bayes assumes conditional

independence between features, making previous models incapable of understanding the time series as anything beyond a bag of values. [Table 1](#) demonstrates that CEVVO is the most effective model for ECMO data, showing a significant improvement compared with other methods. Using a permutation test, CEVVO demonstrated a significantly higher AUROC (all P values $<.001$) and AP (all P values $<.04$) than all other methods.

Table 1. Comparison of Continuous Evaluation of Venovenous Extracorporeal Membrane Oxygenation Outcomes (CEVVO) with other models used previously.

Model name	Total AUROC ^a , 95% CI	Total AP ^b , 95% CI	P value compared with CEVVO (AUROC)	P value compared with CEVVO (AP)
CEVVO	0.6822-0.7055	0.8515-0.8682	— ^c	—
Logistic regression	0.6395-0.6626	0.8396-0.8566	$<.001$.04
Naive Bayes	0.5876-0.6081	0.8111-0.8255	$<.001$	$<.001$
Dense network	0.5673-0.5908	0.8148-0.8322	$<.001$	$<.001$
Decision tree	0.5419-0.5596	0.5273-0.5467	$<.001$	$<.001$

^aAUROC: area under the receiver operator characteristic.

^bAP: average precision.

^cNot applicable.

Risk Classification System

The calibration plot of the training data is shown in [Figure S1](#) in [Multimedia Appendix 1](#). The classic S-shaped misalignment indicated that Platt scaling would improve the calibration. Both the calibrated training and test sets are shown in [Figure 6](#).

Using the predictions as an indication of favorable or unfavorable outcomes, patients can be stratified into groups based on their prediction value using equation 5. The clinically relevant measures of performance are shown in [Figures 7](#) and

8. These charts were created by finding the nearest predicted value of each patient before either 72 or 96 hours, sorting them into groups according to equation 5, and then charting their decannulation result. Patients decannulated before 72 or 96 hours were excluded. In the 72-hour case, the groups had a successful decannulation rate of 58% (7/12) for group 0, 77% (17/22) for group 1, 88% (42/48) for group 2, and 92% (11/12) for group 3. In the 96-hour case, the groups had successful decannulation rates of 54% (6/11), 85% (17/20), 81% (42/50), and 100% (9/9), respectively.

Figure 6. The calibration plot for both the training and test set prediction. Each set of predictions has been scaled. The green line shows the theoretical perfect calibration, and the purple bars show the number of data points in each bin.

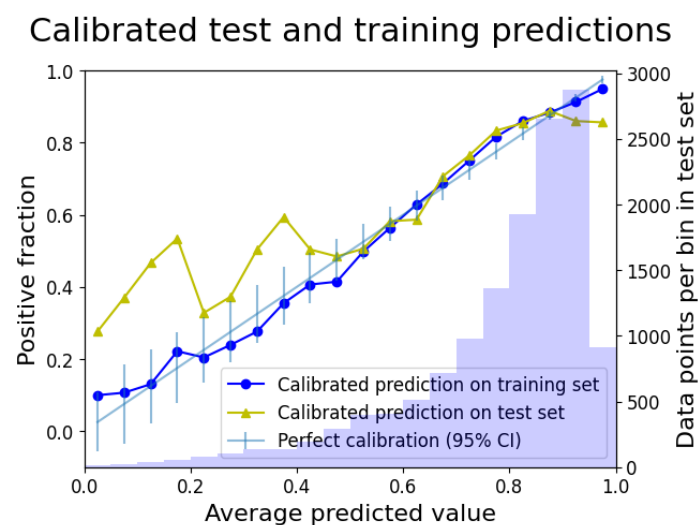


Figure 7. Patient result based on groupings at 72 hours. ECMO: extracorporeal membrane oxygenation.

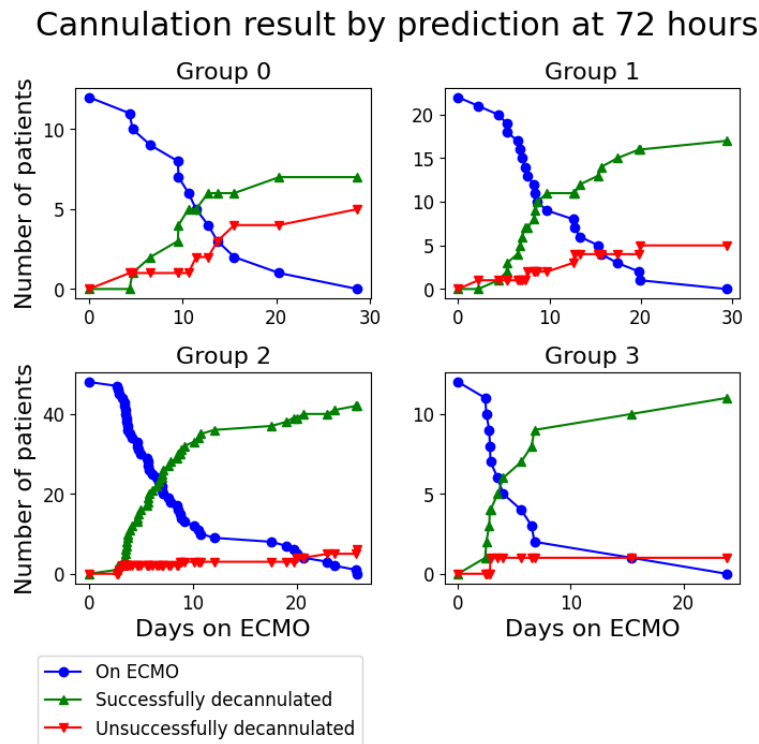
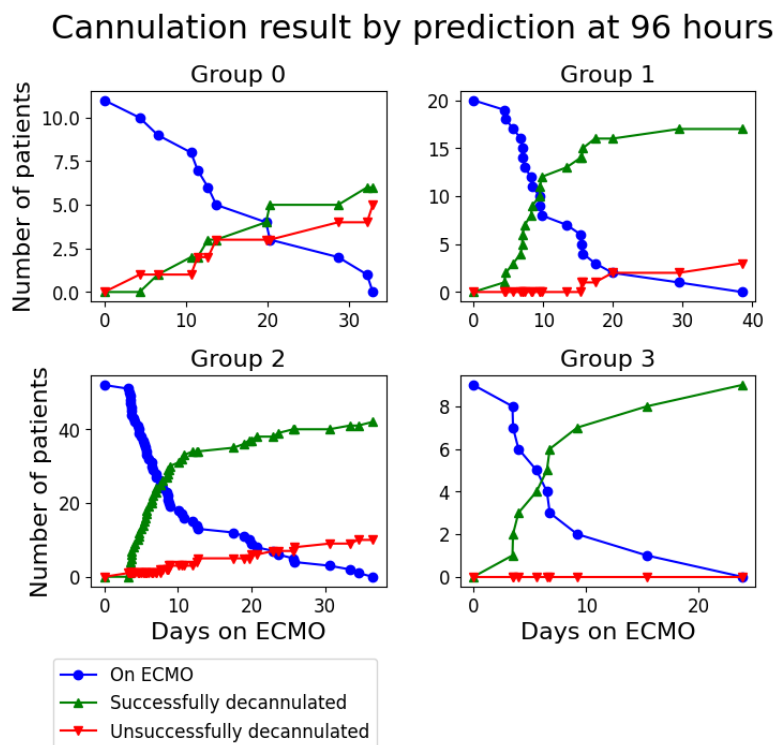


Figure 8. Patient result based on groupings at 96 hours. ECMO: extracorporeal membrane oxygenation.



A Boschloo exact test between groups 0 and 3 yielded P values of .04 for 72 hours and .01 for 96 hours.

Necessity of Time Dependencies

To ensure that each of the synthetic data sets were comparable with each other and the original, a t-distributed Stochastic Neighbor Embedding [25] was used (Figure 9). A more concrete example is shown in Figure 10, where a single synthetic input

was run through both the $L=1$ and $L=100$ GPRs and then compared with an original patient.

The procedure specified in the *Model Assessment* section was repeated for CEVVO, logistic regression, and dense network on the synthetic data set. The 95% CI for the AUROC is shown in Table 2. The expected result is observed where logistic regression and the dense network show no change in performance. CEVVO shows a significant drop in performance

despite having similar, nonlinear properties to the dense network.

Figure 9. 2D t-distributed Stochastic Neighbor Embedding (tSNE) of the 2 synthetic sequential data set and the original patient data. Each dot represents a synthetic patient; the red dots indicate data generated using a radial basis function (RBF) with $L=1$, the green dots indicate data generated using length $L=100$, and the blue dots indicate the original data. The significant overlap connotes similarity between the literal values.

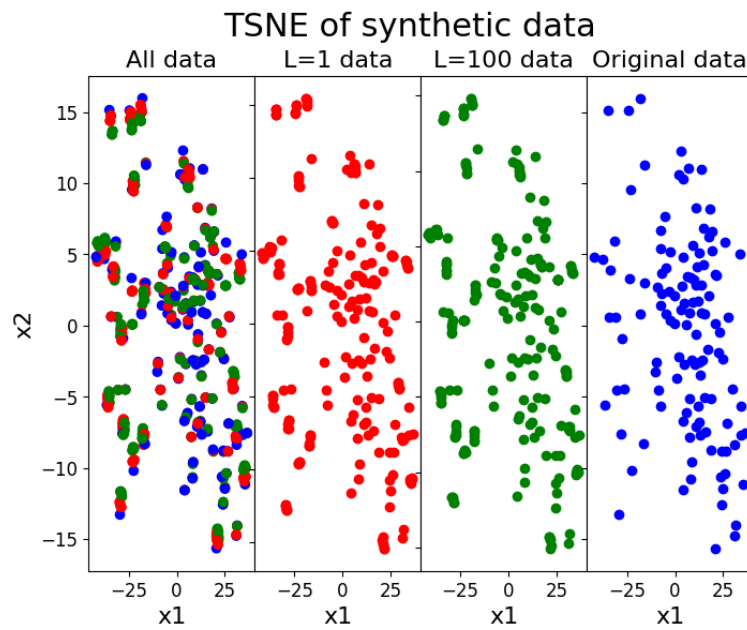


Figure 10. An example synthetic patient, shown in both the $L=1$ and $L=100$ data sets compared with a similar real patient (in blue). ECMO: extracorporeal membrane oxygenation; RPM: revolutions per minute.

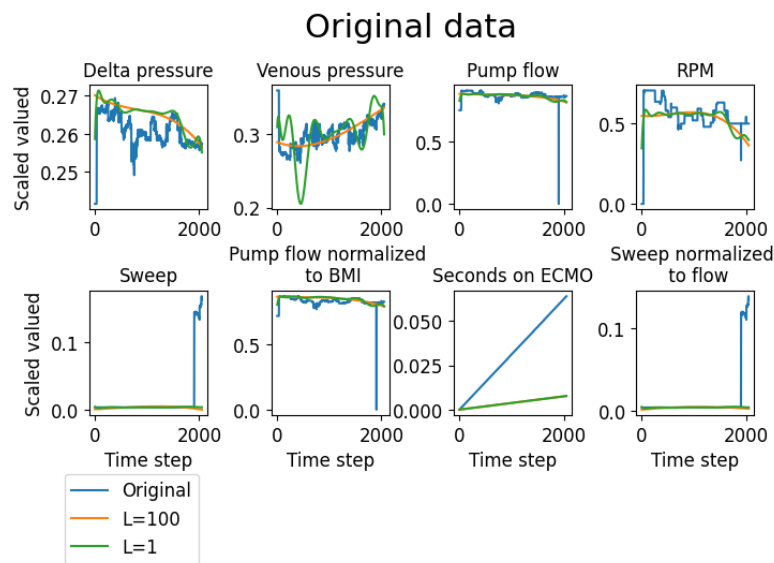


Table 2. Comparison of Continuous Evaluation of Venovenous Extracorporeal Membrane Oxygenation Outcomes (CEVVO) with top-performing models used previously on each synthetic data set.

Model name	Total AUROC ^a for $L=1$ synthetic data set, 95% CI	Total AUROC for $L=100$ synthetic data set, 95% CI	P value between $L=1$ and $L=100$
CEVVO	0.8223-0.8583	0.7424-0.7849	<.001
Logistic regression	0.7813-0.8213	0.7814-0.8190	.46
Dense network	0.7080-0.7513	0.6924-0.7352	.17

^aAUROC: area under the receiver operator characteristic.

Discussion

Principal Findings

VV-ECMO is an invasive and resource-intensive therapy used for patients with refractive respiratory failure. Decannulation from ECMO is generally performed through a weaning trial, in which the ECMO support, measured as flow through the circuit, is titrated down. Experienced clinician decision-making with careful consideration of patient hemodynamics, response to therapy, and pathophysiology informs the decision on when to perform the weaning trial. Our study investigates a novel approach to analyzing clinical information and perfusion hemodynamics in real time to assist clinicians with the decision of when to move forward with decannulation from VV-ECMO.

Although CEVVO was more accurate at predicting the success of decannulation than other models, the model should be considered as an additional data point to guide clinical management. Patients stratified to the high-risk group had a higher risk of therapy failure, with >50% of the patients in this group successfully decannulated in both the 72 and 96 hour cases. As expected, the calibration plot also showed that patients in the low-risk group were decannulated successfully more often. Using these data, clinicians may reference the model and elect to start weaning trials on patients stratified to the low-risk cohort sooner.

Comparison With Prior Work

To the best of our knowledge, CEVVO is the first to use ECMO perfusion data and a deep learning architecture to provide clinical decision support for the decannulation decision for VV-ECMO. By using a model that can successfully combine dynamic and static data, significantly improved performance on binary classification can be achieved when compared with other models. Using perfusion data and clinical information, CEVVO was trained to classify patients by decannulation outcome (successful or unsuccessful). The performance was evaluated using 3 criteria: AUROC, AP, and the clinical usefulness of predictions. Relative to other models noted in the literature, such as logistic regression and decision trees, the LSTM-based model showed significant improvement on the ECMO machine data set.

Performance

The AUROC and AP scores for the full data set had 95% CIs of 0.6822-0.7055 and 0.8515-0.8682, demonstrating a fair ability to predict exact outcomes. This was marginally improved to 95% CIs of 0.7048-0.7428 and 0.9074-0.9261 by limiting the data set to only consider data points collected 3 to 11 days after cannulation. However, these numbers only represented the average performance.

Synthetic Data

The use of GPR-created data sets further cemented the notion that the novelty of the architecture, understanding time dependence, is truly what is responsible for the performance edge over other models. The assumption of temporal dependence is inherent in the data as it is medically motivated. There is an expectation that the specific value of the perfusion data shares

much mutual information with the outcome. The $L=1$, $L=100$, and original data sets are very similar in their t-distributed Stochastic Neighbor Embedding projection, differing only slightly in the specific values. However, as shown in Figure 10, within the $L=100$ group, the local structure was obliterated, leading to a loss of information about how later time points affect the outcome probability. Logistic regression explicitly assumes that each time point is independent, and thus, it has highly similar AUROC distributions ($P=.46$) compared with the dense network ($P=.17$) and CEVVO ($P<.001$). The nonlinear nature of dense neural networks is able to approximate time dependence but is less efficient than the LSTM-based architecture.

Risk Classification

These initial measures of performance were then used to contextualize the clinical predictions: stratifying people into groups, based on associated risk, to predict recovery. The numerical value of each patient's prediction was divided into groups, and patients were followed to their decannulation result. For the grouping to be useful, there should be some difference in the success percentage that increases from the high-risk group to the low-risk group. This result was observed in this study. When measured at 72 hours, 58% (7/12) of the patients in the high-risk group had a successful decannulation, whereas 92% (11/12) of the patients in the low-risk group were successfully decannulated ($P=.04$). When measured at 96 hours, the successful decannulation percentage was similar: 54% (6/11) of the patients in the high-risk group and 100% (9/9) patients in the low-risk group were successfully decannulated ($P=.01$).

Limitations

Cohort studies using retrospective data collection are subject to inherent bias. We mitigated bias in this study by including all consecutive patients supported on VV-ECMO at our center.

Incomplete data recording from the ECMO devices may have contributed to this model. In the future, this could be mitigated by increasing the sample size and improving data capture methodology.

Clinically, patients with different indications for ECMO support vary in their hospital course, and the number of different disease etiologies may have been too few for the model to learn. Larger cohorts may help mitigate the issues related to an underpowered data set. Furthermore, model performance declined beyond a period of approximately 11 days, which may be attributed to a challenging hospital course with heterogeneous factors and an increased risk for complications. The effectiveness of ECMO as a long-term therapy remains unclear, and our data support this conclusion.

Future Direction

In the future, more information about each patient's hospital course, such as administration of vasopressors, ventilator settings, imaging studies, and other interventions may be used to develop an improved model. Indeed, more data and reducing unaccounted variables may improve model performance over longer periods. Extending this study to include patients on other forms of ECLS, such as VA-ECMO and cardiogenic shock,

may be helpful in guiding clinical management. We suggest that larger and more comprehensive repositories of health care data may improve the management of patients considered most critically ill.

Acknowledgments

The authors would like to thank members of the Columbia University Irving Institute for Cancer Dynamics, Data Science Institute, the Biomedical Engineering Department, and the Columbia University Irving Medical Center for helpful discussions and support.

Data Availability

The main clinical data set is unavailable for the public because it contains protected health information. However, 2 synthetic data sets generated based on the original data are available at the Aizilab GitHub data set, along with the corresponding Python code.

Authors' Contributions

JF played a key role in conceptualization and data curation and took the lead in formal analysis, investigation, methodology, software, visualization, and writing the original draft. AA led data curation and provided support in conceptualization, investigation, and writing the review. DM and JB contributed to data curation. PL led conceptualization, supported investigation, and coled supervision, contributing to the writing of the review. EA provided support in conceptualization, formal analysis, investigation, methodology, and software. In addition, EA coled supervision and contributed to writing the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic data for the patient population, along with the bootstrapping algorithm.

[[DOCX File, 79 KB - biomedeng_v9i1e48497_app1.docx](#)]

References

1. Ali J, Vuylsteke A. Extracorporeal membrane oxygenation: indications, technique and contemporary outcomes. *Heart* 2019 Sep 30;105(18):1437-1443. [doi: [10.1136/heartjnl-2017-311928](https://doi.org/10.1136/heartjnl-2017-311928)] [Medline: [31040171](https://pubmed.ncbi.nlm.nih.gov/31040171/)]
2. Akoumianaki E, Jonkman A, Sklar MC, Georgopoulos D, Brochard L. A rational approach on the use of extracorporeal membrane oxygenation in severe hypoxemia: advanced technology is not a panacea. *Ann Intensive Care* 2021 Jul 12;11(1):107 [FREE Full text] [doi: [10.1186/s13613-021-00897-3](https://doi.org/10.1186/s13613-021-00897-3)] [Medline: [34250563](https://pubmed.ncbi.nlm.nih.gov/34250563/)]
3. Tran A, Fernando SM, Rochweg B, Barbaro RP, Hodgson CL, Munshi L, et al. Prognostic factors associated with mortality among patients receiving venovenous extracorporeal membrane oxygenation for COVID-19: a systematic review and meta-analysis. *Lancet Respir Med* 2023 Mar;11(3):235-244. [doi: [10.1016/s2213-2600\(22\)00296-x](https://doi.org/10.1016/s2213-2600(22)00296-x)]
4. Kim JH, Pieri M, Landoni G, Scandroglio AM, Calabrò MG, Fominskiy E, et al. Venovenous ECMO treatment, outcomes, and complications in adults according to large case series: a systematic review. *Int J Artif Organs* 2021 Jul 01;44(7):481-488. [doi: [10.1177/0391398820975408](https://doi.org/10.1177/0391398820975408)] [Medline: [33259258](https://pubmed.ncbi.nlm.nih.gov/33259258/)]
5. Shah N, Said AS. Extracorporeal support prognostication-time to move the goal posts? *Membranes (Basel)* 2021 Jul 15;11(7):537 [FREE Full text] [doi: [10.3390/membranes11070537](https://doi.org/10.3390/membranes11070537)] [Medline: [34357187](https://pubmed.ncbi.nlm.nih.gov/34357187/)]
6. Pappalardo F, Pieri M, Greco T, Patroniti N, Pesenti A, Arcadipane A, et al. Predicting mortality risk in patients undergoing venovenous ECMO for ARDS due to influenza A (H1N1) pneumonia: the ECMOnet score. *Intensive Care Med* 2013 Feb 16;39(2):275-281 [FREE Full text] [doi: [10.1007/s00134-012-2747-1](https://doi.org/10.1007/s00134-012-2747-1)] [Medline: [23160769](https://pubmed.ncbi.nlm.nih.gov/23160769/)]
7. Schmidt M, Zogheib E, Rozé H, Repesse X, Lebreton G, Luyt CE, et al. The PRESERVE mortality risk score and analysis of long-term outcomes after extracorporeal membrane oxygenation for severe acute respiratory distress syndrome. *Intensive Care Med* 2013 Oct 2;39(10):1704-1713 [FREE Full text] [doi: [10.1007/s00134-013-3037-2](https://doi.org/10.1007/s00134-013-3037-2)] [Medline: [23907497](https://pubmed.ncbi.nlm.nih.gov/23907497/)]
8. Schmidt M, Bailey M, Sheldrake J, Hodgson C, Aubron C, Rycus PT, et al. Predicting survival after extracorporeal membrane oxygenation for severe acute respiratory failure. The respiratory extracorporeal membrane oxygenation survival prediction (RESP) score. *Am J Respir Crit Care Med* 2014 Jun 01;189(11):1374-1382. [doi: [10.1164/rccm.201311-2023oc](https://doi.org/10.1164/rccm.201311-2023oc)]
9. Roch A, Hraiech S, Masson E, Grisoli D, Forel JM, Boucekine M, et al. Outcome of acute respiratory distress syndrome patients treated with extracorporeal membrane oxygenation and brought to a referral center. *Intensive Care Med* 2014 Jan 30;40(1):74-83 [FREE Full text] [doi: [10.1007/s00134-013-3135-1](https://doi.org/10.1007/s00134-013-3135-1)] [Medline: [24170143](https://pubmed.ncbi.nlm.nih.gov/24170143/)]
10. Cheng YT, Wu MY, Chang YS, Huang CC, Lin PJ. Developing a simple preinterventional score to predict hospital mortality in adult venovenous extracorporeal membrane oxygenation: a pilot study. *Medicine (Baltimore)* 2016 Jul;95(30):e4380 [FREE Full text] [doi: [10.1097/MD.0000000000004380](https://doi.org/10.1097/MD.0000000000004380)] [Medline: [27472730](https://pubmed.ncbi.nlm.nih.gov/27472730/)]

11. Hilder M, Herbstreit F, Adamzik M, Beiderlinden M, Bürschen M, Peters J, et al. Comparison of mortality prediction models in acute respiratory distress syndrome undergoing extracorporeal membrane oxygenation and development of a novel prediction score: the PRediction of Survival on ECMO Therapy-Score (PRESET-Score). *Crit Care* 2017 Dec 12;21(1):301 [FREE Full text] [doi: [10.1186/s13054-017-1888-6](https://doi.org/10.1186/s13054-017-1888-6)] [Medline: [29233160](https://pubmed.ncbi.nlm.nih.gov/29233160/)]
12. Abbasi A, Karasu Y, Li C, Sodha NR, Eickhoff C, Ventetuolo CE. Machine learning to predict hemorrhage and thrombosis during extracorporeal membrane oxygenation. *Crit Care* 2020 Dec 10;24(1):689 [FREE Full text] [doi: [10.1186/s13054-020-03403-6](https://doi.org/10.1186/s13054-020-03403-6)] [Medline: [33302954](https://pubmed.ncbi.nlm.nih.gov/33302954/)]
13. Ayers B, Wood K, Gosev I, Prasad S. Predicting survival after extracorporeal membrane oxygenation by using machine learning. *Ann Thorac Surg* 2020 Oct;110(4):1193-1200. [doi: [10.1016/j.athoracsur.2020.03.128](https://doi.org/10.1016/j.athoracsur.2020.03.128)] [Medline: [32454016](https://pubmed.ncbi.nlm.nih.gov/32454016/)]
14. Loyaga-Rendon RY, Fermin DR, Grayburn RL, Matthew GH, Dickinson MG, Manandhar-Shrestha N, et al. Predicting short-term mortality in ECMO-supported patients secondary to decompensated heart failure and acute myocardial infarction using machine learning. *J Heart Lung Transplant* 2022 Apr;41(4):S470-S471. [doi: [10.1016/j.healun.2022.01.1189](https://doi.org/10.1016/j.healun.2022.01.1189)]
15. Fung K, Beck JR, Lopez HC2, Mongero LB. Case report: remote monitoring using Spectrum Medical Live Vue allows improved response time and improved quality of care for patients on cardiopulmonary support. *Perfusion* 2013 Nov 19;28(6):561-564. [doi: [10.1177/0267659113497498](https://doi.org/10.1177/0267659113497498)] [Medline: [23873484](https://pubmed.ncbi.nlm.nih.gov/23873484/)]
16. Perna D, Tagarelli A. Deep auscultation: predicting respiratory anomalies and diseases via recurrent neural networks. arXiv. Preprint posted online July 11, 2019 2024 [FREE Full text] [doi: [10.1109/cbms.2019.00020](https://doi.org/10.1109/cbms.2019.00020)]
17. Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS One* 2019;14(7):e0218942 [FREE Full text] [doi: [10.1371/journal.pone.0218942](https://doi.org/10.1371/journal.pone.0218942)] [Medline: [31283759](https://pubmed.ncbi.nlm.nih.gov/31283759/)]
18. Li K, Daniels J, Liu C, Herrero P, Georgiou P. Convolutional recurrent neural networks for glucose prediction. arXiv. Preprint posted online July 9, 2018 2024 [FREE Full text] [doi: [10.1109/jbhi.2019.2908488](https://doi.org/10.1109/jbhi.2019.2908488)]
19. Nguyen M, He T, An L, Alexander DC, Feng J, Yeo BT. Predicting Alzheimer's disease progression using deep recurrent neural networks. *Neuroimage* 2020 Nov 15;222:117203 [FREE Full text] [doi: [10.1016/j.neuroimage.2020.117203](https://doi.org/10.1016/j.neuroimage.2020.117203)] [Medline: [32763427](https://pubmed.ncbi.nlm.nih.gov/32763427/)]
20. Banerjee I, Ling Y, Chen MC, Hasan SA, Langlotz CP, Moradzadeh N, et al. Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. *Artif Intell Med* 2019 Jun;97:79-88 [FREE Full text] [doi: [10.1016/j.artmed.2018.11.004](https://doi.org/10.1016/j.artmed.2018.11.004)] [Medline: [30477892](https://pubmed.ncbi.nlm.nih.gov/30477892/)]
21. Yoon J, Jarrett D, van der Schaar M. Time-series generative adversarial networks. In: Proceedings of the 2019 Conference on Neural Information Processing Systems. 2019 Presented at: NeurIPS 2019; December 08-14, 2019; Vancouver, BC URL: <https://proceedings.neurips.cc/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf>
22. Williams C, Rasmussen C. Gaussian processes for regression. In: Proceedings of the 8th International Conference on Neural Information Processing Systems. 1995 Presented at: NIPS'95; November 27-December 2, 1995; Denver, CO URL: <https://doi.org/10.5555/2998828.2998901>
23. Keras: deep learning for humans. Keras. URL: <https://keras.io/> [accessed 2023-01-27]
24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12(85):2825-2830.
25. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9(86):2579-2605 [FREE Full text]

Abbreviations

- AUROC:** area under the receiver operating characteristic curve
- AP:** average precision
- CEVVO:** Continuous Evaluation of Venous Arterial Extracorporeal Membrane Oxygenation Outcomes
- ECLS:** extracorporeal life support
- ECMO:** extracorporeal membrane oxygenation
- GPR:** Gaussian process regression
- LSTM:** long short-term memory
- RBF:** radial basis function
- RNN:** recurrent neural network
- VA-ECMO:** venous arterial extracorporeal membrane oxygenation
- VV-ECMO:** venovenous extracorporeal membrane oxygenation

Edited by T Leung; submitted 26.04.23; peer-reviewed by C Zhao, J Chen; comments to author 29.09.23; revised version received 03.11.23; accepted 29.12.23; published 02.02.24.

Please cite as:

Fuller J, Abramov A, Mullin D, Beck J, Lemaitre P, Azizi E

A Deep Learning Framework for Predicting Patient Decannulation on Extracorporeal Membrane Oxygenation Devices: Development and Model Analysis Study

JMIR Biomed Eng 2024;9:e48497

URL: <https://biomedeng.jmir.org/2024/1/e48497>

doi: [10.2196/48497](https://doi.org/10.2196/48497)

PMID: [38875691](https://pubmed.ncbi.nlm.nih.gov/38875691/)

©Joshua Fuller, Alexey Abramov, Dana Mullin, James Beck, Philippe Lemaitre, Elham Azizi. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 02.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Enhancing Energy Efficiency in Telehealth Internet of Things Systems Through Fog and Cloud Computing Integration: Simulation Study

Yunyong Guo^{1*}, PhD; Sudhakar Ganti^{1*}, PhD; Yi Wu¹, BSc

Computer Science Department, University of Victoria, Victoria, BC, Canada

*these authors contributed equally

Corresponding Author:

Yunyong Guo, PhD

Computer Science Department, University of Victoria

3800 Finnerty Rd

Victoria, BC, V8P 5C2

Canada

Phone: 1 (250) 721 7211

Email: yunyong@uvic.ca

Abstract

Background: The increasing adoption of telehealth Internet of Things (IoT) devices in health care informatics has led to concerns about energy use and data processing efficiency.

Objective: This paper introduces an innovative model that integrates telehealth IoT devices with a fog and cloud computing-based platform, aiming to enhance energy efficiency in telehealth IoT systems.

Methods: The proposed model incorporates adaptive energy-saving strategies, localized fog nodes, and a hybrid cloud infrastructure. Simulation analyses were conducted to assess the model's effectiveness in reducing energy consumption and enhancing data processing efficiency.

Results: Simulation results demonstrated significant energy savings, with a 2% reduction in energy consumption achieved through adaptive energy-saving strategies. The sample size for the simulation was 10-40, providing statistical robustness to the findings.

Conclusions: The proposed model successfully addresses energy and data processing challenges in telehealth IoT scenarios. By integrating fog computing for local processing and a hybrid cloud infrastructure, substantial energy savings are achieved. Ongoing research will focus on refining the energy conservation model and exploring additional functional enhancements for broader applicability in health care and industrial contexts.

(*JMIR Biomed Eng* 2024;9:e50175) doi:[10.2196/50175](https://doi.org/10.2196/50175)

KEYWORDS

cloud computing; energy-efficient; fog computing; Internet of Things; IoT; telehealth

Introduction

Overview

Health care is a critical global industry, and the advent of the Internet of Things (IoT) and cloud computing has significantly transformed health care system management. The ever-increasing data volume generated by these systems demands efficient, energy-saving computing platforms. In response, we present a groundbreaking energy-efficient model that seamlessly integrates telehealth IoT devices with fog and cloud computing-based platforms, offering a unique solution

to address energy efficiency and data processing challenges. The rapid proliferation of IoT devices in health care has transformed approaches to patient care, diagnostics, and treatment. Telehealth, a key IoT health care application, has proven its potential to enhance care quality, reduce costs, and boost patient satisfaction. Despite these benefits, issues such as scalability, latency, and resource management persist, along with the significant challenge of energy consumption in smart devices within fog environments [1]. As a result, energy efficiency must be prioritized in the development of fog computing solutions, given its substantial impact on reducing carbon footprints and mitigating climate change effects. The

large-scale deployment of telehealth IoT devices also raises concerns about energy consumption and data processing efficiency in delivering quality health care services. Intelligent choices for telehealth IoT devices should consider factors such as device movement or relevant environmental conditions to optimize energy consumption and manage associated equipment effectively. Typically, cloud-based analytical assessments are conducted for these devices [2]. To tackle these challenges, we propose an energy-saving model that integrates telehealth IoT devices with a fog and public or private cloud computing-based platform. The aim of the study is to develop an energy-efficient model that optimally integrates telehealth IoT devices with fog and cloud computing platforms, addressing challenges related to energy consumption, scalability, and data processing efficiency in delivering quality medical and patient services.

Telehealth IoT devices refer to a wide range of interconnected medical devices and sensors that facilitate remote health care services. These devices enable the continuous monitoring of patient's vital signs, timely diagnostics, and personalized treatment plans, thereby improving the overall quality of health care. Some common examples of telehealth IoT devices include wearable health monitors, smart glucose meters, remote patient monitoring systems, and telemedicine platforms. The large-scale deployment of telehealth IoT devices presents several challenges [3], including energy consumption, data management, latency, security and privacy, scalability, and interoperability.

Related Work

Telehealth has emerged as a promising solution to address various challenges in health care, such as accessibility, cost, and quality of care [4]. IoT devices play a significant role in telehealth applications, enabling remote monitoring, diagnostics, and treatment [2]. Several studies have investigated the implementation and efficacy of telehealth IoT devices in various health care scenarios, highlighting their potential to improve patient outcomes and satisfaction [5,6]. Fog computing has been identified as a promising approach to address the challenges associated with large-scale IoT deployments in health care, such as latency, energy consumption, and data management [7,8]. Researchers have proposed several fog computing-based architectures and frameworks for health care applications, demonstrating the potential of fog computing to enhance the performance and efficiency of telehealth IoT devices [9-11]. Cloud computing has gained significant attention in health care due to its scalability, cost-effectiveness, and advanced data analytics capabilities [12,13]. Several studies have explored the integration of cloud computing with telehealth IoT devices, showing its potential to address the challenges related to data storage, processing, and security [14-16].

Energy efficiency is critical in large-scale IoT deployments, especially in health care applications where device longevity and reliability are essential [17]. Researchers have proposed various energy-saving models and strategies for IoT devices, including adaptive power management [18], energy-efficient routing protocols [19], and data compression techniques [20]. However, few studies have specifically focused on energy-saving models that integrate telehealth IoT devices with fog and cloud computing-based platforms. The integration of

fog and cloud computing has emerged as a promising approach to harness the benefits of both paradigms and address the challenges of large-scale IoT deployments [21,22]. Several studies have proposed models and frameworks that combine fog and cloud computing for various IoT applications [23-25], but few have specifically targeted energy-saving in telehealth IoT deployments.

In recent years, several simulation methods have been developed to study the integration of fog nodes in IoT devices and cloud computing. Gupta et al [26] introduced iFogSim, a toolkit for modeling and simulating resource management techniques in IoT, edge, and fog computing environments. Oueis et al [27] presented a simulation study on load distribution in small-cell cloud computing using fog computing and proposed a fog balancing technique to optimize resource allocation and reduce latency. Barcelo et al [28] explored IoT-cloud service optimization through simulation in smart environments, presenting a novel optimization framework that uses fog nodes to reduce latency and energy consumption. Zeng et al [29] conducted a comparative study of IoT cloud and fog computing simulations using iFogSim and Cooja, discussing the advantages and limitations of both simulators and providing insights into selecting an appropriate tool for specific scenarios. Lastly, Byers and Wetterwald [30] discussed the concept of fog computing and its importance in distributing data and intelligence for IoT resiliency and scalability, presenting various simulation models and techniques used to evaluate the performance of fog computing in IoT environments. Several studies have focused on the Yet Another Fog Simulator (YAFS) framework, a simulator designed for modeling and simulating fog computing environments in IoT scenarios. Bermejo et al [31] introduced YAFS, presenting the architecture, components, and use cases of the simulator, demonstrating its effectiveness in modeling and simulating fog computing deployments. García et al [32] showcased YAFS's ability to model and simulate fog computing scenarios and analyze the performance of different scheduling algorithms. In a comparative study, Rodríguez et al [33] analyzed the features, capabilities, and limitations of YAFS, iFogSim, and EdgeCloudSim simulators, providing insights into selecting the most suitable tool for specific fog computing scenarios.

Several studies have explored different aspects of telehealth simulations, fog nodes, IoT devices, and cloud computing for energy-saving purposes. Aazam and Huh [34] discussed a smart gateway-based communication approach using fog computing for energy-saving in the Cloud of Things, which can be applied to various IoT applications, including telehealth. Verma and Sood [35] presented a fog-assisted IoT framework for patient health monitoring in smart homes, focusing on energy efficiency and reduced latency through a decentralized fog computing architecture. Koubaâ et al [36] proposed a fog-based emergency and health care system for smart cities, which leverages fog nodes and IoT devices to optimize energy consumption and provide real-time health care services, thus addressing energy-saving concerns in telehealth scenarios. Sareen et al [37] introduced an energy-efficient context-aware framework for managing application execution in cloud-fog environments,

which can potentially improve energy efficiency in various IoT applications, including telehealth scenarios.

Methods

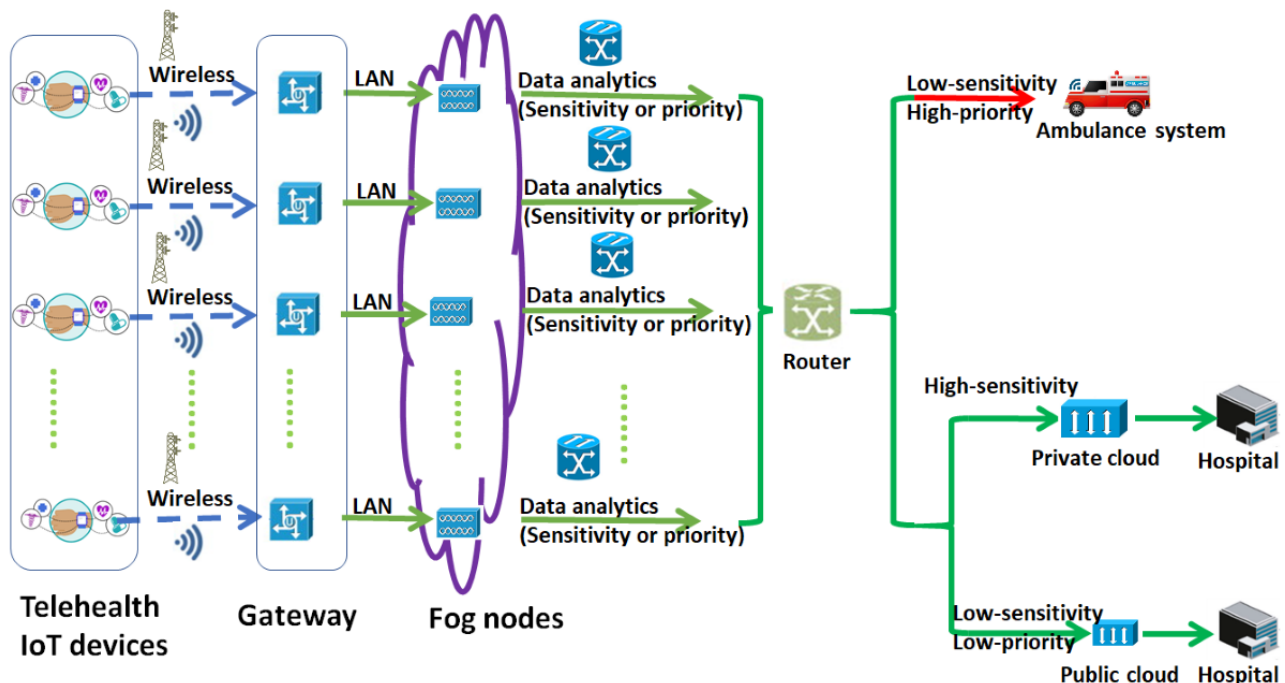
Model Overview

The proposed energy-saving model is designed to integrate telehealth IoT devices with a fog and cloud computing-based platform, leveraging the advantages of both paradigms to optimize energy consumption and ensure efficient data processing. The model comprises 3 main components: IoT devices, fog nodes, and public or private cloud servers, which are interconnected through a communication network.

The model architecture is shown in Figure 1 [38].

1. IoT devices: telehealth IoT devices, such as wearables, sensors, and remote monitoring systems, collect and transmit patient data in real time. These devices can dynamically adjust their power states (eg, active, idle, and sleep) based on their tasks, reducing energy consumption without compromising the quality of health care services.
2. Fog nodes: fog nodes, located near IoT devices, serve as intermediate processing units. They perform localized data processing, analytics, and storage, reducing the amount of data transmitted to the cloud servers.
3. Cloud servers: cloud servers provide a robust infrastructure for large-scale data storage, processing, and advanced analytics.
4. Communication network: a communication network connects IoT devices, fog nodes, and cloud servers, enabling seamless data transmission and task allocation.

Figure 1. Telehealth Internet of Things (IoT) devices integrated with fog nodes and a private or public cloud architecture model. LAN: local area network.



The telehealth IoT network depicted in the diagram is designed to ensure efficient and secure data transmission between the different network components. To ensure network security, firewalls are placed between IoT devices and fog nodes. This ensures that unauthorized access to the network is prevented, and sensitive health care data are kept confidential. To process the data requests, the fog nodes are equipped with data analytics functions that enable them to intelligently assign different types of requests to either fog nodes, a private cloud, or a public cloud. This intelligent decision-making process is more effective and efficient than the traditional “first-come, first-served” approach. The gateway and router are integral components in the network that enable seamless data transmission between the fog nodes and cloud instances. The gateway acts as the entry point for the network and connects the IoT devices to the local fog nodes. It is responsible for handling the data transmission and conversion between different protocols used by IoT devices and fog nodes. The router, on the other hand, is responsible for directing the data traffic between the fog nodes and cloud instances based

on various factors, such as the sensitivity, priority, and latency requirements of the data. It determines which data should be sent to the cloud and which data should be processed by the fog nodes, ensuring efficient use of network resources. The router also handles the communication between different fog nodes and cloud instances, enabling seamless data transmission across the network.

The proposed telehealth IoT system shown in Figure 2 intelligently manages data transmission based on the sensitivity and priority of the data. For high-sensitivity data, the system ensures privacy and security by sending it directly to the private cloud, which then transfers the data to authorized health facilities as needed. On the other hand, low-sensitivity but high-priority requests are routed to the fog nodes as they have the capability to process urgent requests in a timely manner, such as in life-threatening emergency situations. These requests are then transmitted to ambulance systems for immediate treatment. Lastly, data with low sensitivity and low priority are sent to the

public cloud as it has more space and scalability to store and process such data. The public cloud can also serve as a repository for future research or clinical purposes.

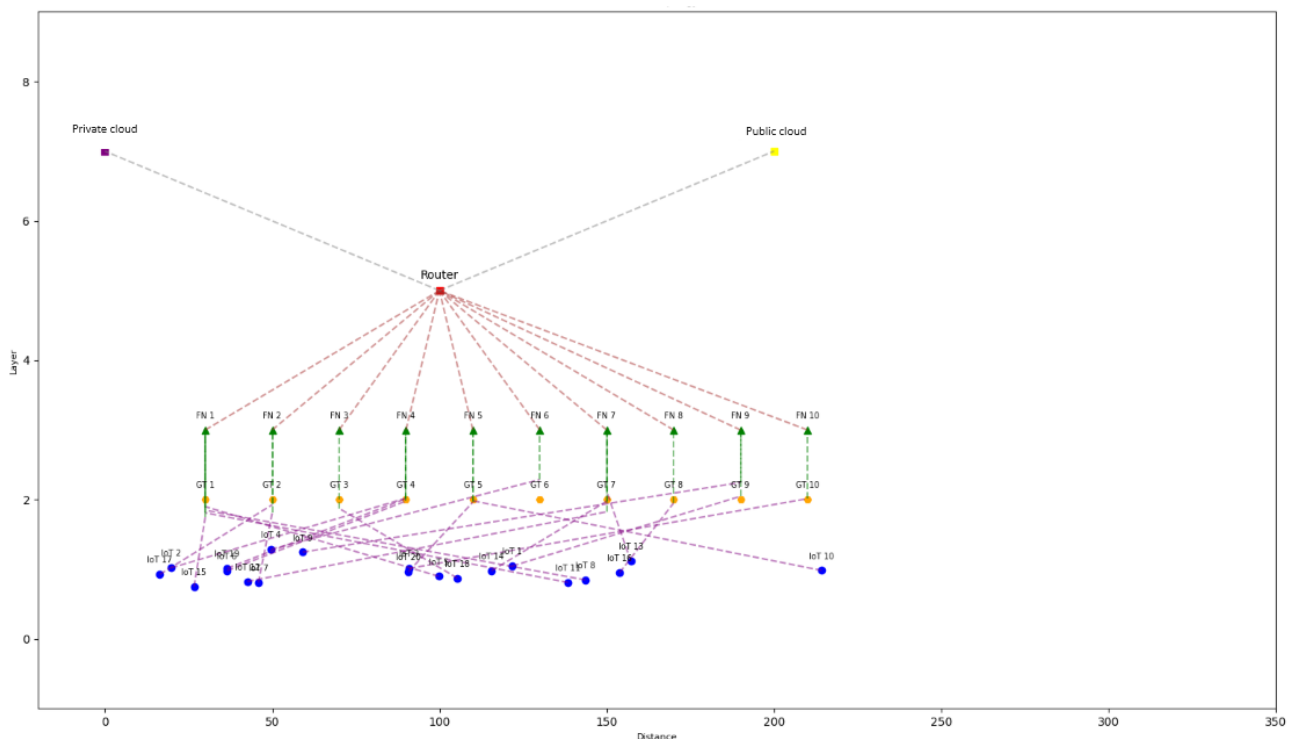
By allocating data transmission to the appropriate destination, the proposed system ensures efficient and effective data processing while maintaining privacy and security for sensitive health care data. This approach also optimizes energy consumption and reduces latency, ensuring a seamless experience for health care providers and patients. The categorization of high and low sensitivity and high and low priority data sent from telehealth IoT monitor devices can depend on various factors, including the specific use case, regulatory requirements, and patient needs. One possible approach could be to use threshold values based on vital signs such as pulse and heartbeat to categorize the data. For example, data related to vital signs that fall within normal ranges may be classified as low sensitivity and low priority, as they do not require immediate attention. Data related to vital signs that are outside the normal range but do not pose an immediate threat to the patient's health may be classified as low sensitivity but high priority. Data related to vital signs that indicate a life-threatening condition, such as cardiac arrest, may be classified as high sensitivity and high priority, requiring immediate attention from health care providers.

The exact vital sign thresholds for patient emergencies can vary depending on a range of factors, including the age and health condition of the patient, the specific symptoms, and other medical history [39]. In general, some common vital sign thresholds used to classify emergencies include the following:

- Heart rate: a heart rate above 100 bpm or below 60 bpm may be indicative of an emergency [40].
- Blood pressure: a systolic blood pressure (the top number) above 180 mm Hg or below 90 mm Hg, or a diastolic blood pressure (the bottom number) above 110 mm Hg or below 60 mm Hg may indicate an emergency [41].
- Respiratory rate: a respiratory rate above 30 breaths per minute or below 10 breaths per minute may be indicative of an emergency [42].
- Oxygen saturation: an oxygen saturation level below 90% may be indicative of an emergency [43].

However, it is important to note that this is just one possible approach, and the categorization of data should be customized based on the specific needs of the patient and health care provider. It is also important to comply with relevant regulations and ensure patient privacy and security while handling sensitive health care data.

Figure 2. Network topology for the proposed Internet of Things (IoT) devices integrated with fog nodes and cloud. A brief overview of the components in the network topology: (1) IoT devices (blue circles) represent individual IoT devices in the network, each associated with a specific fog node. (2) Gateways (GT; orange hexagons) are used to connect the IoT devices to the fog nodes. (3) Fog nodes (FN; green triangles) are intermediate computing resources that process and store data from IoT devices. (4) A router (red square) connects the fog nodes to the private cloud and public cloud. (5) A private cloud (purple square) and a public cloud (yellow square) are the 2 cloud resources in the network.



Key Components and Energy-Saving Strategies

The proposed energy-saving model incorporates several strategies to minimize energy consumption.

Task Allocation

The model intelligently allocates tasks between fog nodes and cloud servers based on factors such as computational capacity, proximity to IoT devices, and current workload. This ensures

efficient data processing and reduces energy consumption for data transmission.

Adaptive Power Management

IoT devices and fog nodes can dynamically adjust their power states (eg, active, idle, and sleep) based on their tasks and workload, ensuring optimal energy consumption without compromising the quality of health care services.

Data Compression and Aggregation

Data generated by IoT devices can be compressed and aggregated at the fog nodes before transmission to cloud servers, reducing the volume of data transmitted and, consequently, energy consumption.

Network Optimization

The communication network can be optimized to minimize energy consumption by using energy-efficient routing protocols and minimizing transmission distances.

Simulation Study

To assess the effectiveness of the proposed energy-efficient model, we developed a simulation model that emulates a real-world telehealth scenario focused on remote patient monitoring. Within this simulated scenario, numerous patients with chronic conditions are equipped with wearable IoT devices that continuously track vital signs such as heart rate, blood pressure, and blood glucose levels. The gathered data are processed and analyzed by the integrated fog and cloud computing-based platform, facilitating timely diagnostics and personalized treatment plans. [Textbox 1](#) contains the pseudocode for the provided code.

In short, this code is devised to emulate an IoT network, scrutinizing the influence of fog nodes on energy consumption while providing a graphical representation of the network architecture to elucidate the connections among IoT devices, fog nodes, and cloud services. IoT devices transmit data to their corresponding destinations, such as fog nodes, private clouds, or public clouds, contingent upon their sensitivity and priority attributes. The energy expenditure for data transmission to these target locations differs; hence, the code performs a simulation to determine the residual energy for each device under 2 distinct scenarios (ie, with and without fog nodes). Subsequently, the code generates a bar chart to depict the energy consumption patterns of IoT devices in both cases, and it stores the energy usage outcomes in 2 separate Microsoft Excel (Microsoft Corporation) files, enabling in-depth examination and assessment of the results.

The algorithm of the code can be analyzed in the following steps:

1. Initialization: create IoT devices, fog nodes, and cloud instances with their respective properties.
2. Connection: connect IoT devices to fog nodes and then fog nodes, and determine which data are transferred to cloud instances (private and public). Each device is connected to a corresponding fog node.
3. Data transmission simulation: simulate data transmission from IoT devices to their respective fog nodes, and then

fog nodes assign the requests to a private cloud or a public cloud based on their priority and sensitivity. If the sensitivity of the device is “high,” data are sent to the private cloud. If the sensitivity is “low” and the priority is “high,” there is a chance (defined by *self.fog_node[chance]*) that data are sent to the fog node. If this condition is not met, the device does not send data. If the sensitivity is “low” and the priority is “low,” data are sent to the public cloud.

4. Energy consumption calculation: calculate the energy consumed by each IoT device during data transmission, considering the parameter of latency. Different energy costs are associated with sending data to different destinations (fog nodes, private cloud, or public cloud).
5. Comparison: compare the energy consumption of IoT devices when using fog nodes and when not using fog nodes. (1) Run the simulation with fog nodes connected and store the remaining energy for each device. (2) Reset the energy of the devices, disconnect them from fog nodes, and run the simulation without fog nodes, storing the remaining energy for each device again.
6. Export the energy usage results to Excel files for both cases (with and without fog nodes).
7. Visualize the network topology with devices, fog nodes, and clouds using the *show_topology* function.

In this enhanced task allocation algorithm, we incorporate additional factors such as device distance, data sensitivity, request priority, energy consumption, and latency to provide a more sophisticated and adaptable solution for large-scale telehealth IoT deployments. The algorithm starts by defining parameters such as latency, distance, energy consumption, and sensitivity thresholds. The task queues for each fog node and cloud server are initialized. For each task type, average processing times, energy consumption, sensitivity, and priority are calculated for each fog node and cloud server according to some random data sent from each IoT device. The algorithm then assesses the latency, priority, sensitivity, and energy consumption for transmitting data from each device to each fog node and then to the private and public cloud server. Based on these factors, the algorithm selects the optimal fog node and cloud server for each device, ensuring that the chosen nodes meet the specified thresholds for latency, sensitivity, and energy consumption. Tasks are allocated to fog nodes and cloud servers based on data sensitivity, priority, and energy consumption, ensuring that the selected nodes do not exceed the energy consumption threshold. If no suitable nodes are found, alternative energy-saving strategies may be considered, or the energy consumption threshold may be adjusted. Finally, the tasks are processed in fog nodes and cloud servers based on their queues. By considering these additional factors, the enhanced algorithm can provide better energy-saving performance and adaptability to various telehealth scenarios, ensuring that the large-scale deployment of telehealth IoT devices on a fog and cloud computing-based platform remains efficient and effective.

[Textbox 2](#) contains a task allocation algorithm for telehealth IoT devices integrated with a fog and cloud computing-based platform.

Textbox 1. The pseudocode for the provided code.

1. Define *IoTDevice* class

- Initialize with attributes: *id*, *distance*, *priority*, *sensitivity*, *fog_node*, *private_cloud*, *public_cloud*, *energy*, *transmit_power*, *idle_power*, and *transmit_time*
- Define *send_data* method
 - Check if the device has energy left
 - Send high-sensitivity data to private cloud if sensitivity is high
 - Send low-sensitivity, high-priority data to fog node if priority is high and *fog_node* exists
 - Send low-sensitivity, low-priority data to public cloud otherwise
- Define *idle* method to reduce energy based on idle power and time

2. Define *FogNode* class

- Initialize with attributes: *id*, *public_cloud*, *energy*, *latency*, *devices*, *fog_energy_cost*, *cloud_energy_cost*, *chance*, *process_power*, *idle_power*, and *process_time*
- Define *connect_device* method to connect a device to the fog node
- Define *store_data* method to store data from a device with given sensitivity and priority
- Define *idle* method to reduce energy based on idle power and time
- Define *send_data* method to send data from connected devices based on their sensitivity and priority

3. Define *PublicCloud* class

- Initialize with attributes: *id*, *energy*, *latency*, and *cloud_energy_cost*
- Define *store_data* method to store data from a device

4. Define *simulate* function

- Create Internet of Things (IoT) devices with random priority and sensitivity
- Create fog nodes connected to a public cloud
- Connect IoT devices to fog nodes
- Connect IoT devices to private and public clouds
- Initialize lists to store energy usage results
- Simulate data transmission with fog nodes, store energy usage results
- Store energy usage with fog nodes
- Reset device energy
- Disconnect devices from fog nodes
- Simulate data transmission without fog nodes, store energy usage results
- Store energy usage without fog nodes
- Create energy usage bar plot and save as an image
- Save energy usage results to Microsoft Excel files (with and without fog nodes)

Textbox 2. Task allocation algorithm for telehealth Internet of Things devices integrated with a fog and cloud computing–based platform.

1. Define parameters

- Internet of Things (IoT) devices: $D = \{d1, d2, \dots, dn\}$
- Fog nodes: $F = \{f1, f2, \dots, fm\}$
- Cloud servers: $C = \{c1, c2\}$
- Task types: $T = \{t1, t2, \dots, tq\}$
- Data sensitivity threshold: S_t
- Data priority threshold: Pr_t
- Latency threshold: L_t
- Energy consumption threshold: E_t

2. Initialize task queues for each IoT device, fog node, and cloud server

- $Q_D [i] = \{\}$ for all i in D
- $Q_F [j] = \{\}$ for all j in F
- $Q_C [l] = \{\}$ for all l in C

3. For each task type t in T

- Calculate the average processing time P_t and energy consumption E_t for each IoT device i in D and fog node j in F .
- Calculate average energy consumption E_t , sensitivity S_t , and priority Pr_t for each IoT device i in D and fog node j in F .

4. For each device d in D and task type t in T

- Calculate the latency L_{dt} for transmitting data from device d to each fog node i in F and cloud server j in C .
- Calculate the priority Pr_{dt} , sensitivity S_{dt} , energy consumption E_{dt} for device d , and each fog node i in F and cloud server j in C .
- Find the fog node j^* and cloud server l^* with the minimum latency for device i^* , considering Pr_t, S_t , and E_t :
 - $j^* = \text{argmin}_j(L_{dt})$ for j in F , such that $L_{dt} \leq L_t, Pr_{dt} \leq Pr_t$ and $S_{dt} \leq S_t$
 - $l^* = \text{argmin}_l(L_{dt})$ for l in C , such that $L_{dt} \leq L_t, Pr_{dt} \leq Pr_t$ and $S_{dt} \leq S_t$

5. Allocate tasks from devices to fog nodes and cloud servers: for each device d in D and task type t in T

- If $S_{dt} [j^*] \leq S_t$, then allocate task t to cloud server l^* and add it to the queue: $Q_C [l^*].\text{append}((d, t))$
- If $Pr_{dt} [j^*] \leq Pr_t$, then allocate task t to fog node j^* and add it to the queue: $Q_F [j^*].\text{append}((d, t))$
- Else if $Pr_{dt} [l^*] \leq Pr_t$, then allocate task t to cloud server l^* and add it to the queue: $Q_C [l^*].\text{append}((d, t))$
- Otherwise, consider alternative energy-saving strategies or adjust the energy consumption threshold E_t .

6. Process tasks in fog nodes and cloud servers based on their queues

- For each fog node j in F , process tasks in $Q_F [j]$
- For each cloud server l in C , process tasks in $Q_C [l]$

This algorithm aims to balance the load between fog nodes and cloud servers while considering latency, sensitivity, request priority, and energy consumption constraints. It can be further optimized by incorporating additional factors, such as device mobility. It is mainly focused on simulating data transmission from IoT devices to different destinations based on their priority and sensitivity, as well as comparing the energy consumption given the various latency when using fog nodes versus not using them. The objective is to demonstrate the potential benefits of using fog nodes in terms of energy efficiency for IoT devices.

Results

Parameters in Results

Based on the simulation results, we can analyze the impact of different parameters on the energy efficiency and performance of the proposed telehealth model with and without fog computing. The parameters in the results are given below.

Snapshot Interval

The *snapshot interval* parameter represents the frequency at which the IoT devices send their data to the fog nodes or cloud servers. As the *snapshot interval* increases, the frequency of

data transmission decreases. With a *snapshot interval* of 1, the IoT devices are sending data continuously. As the number of devices increases, the energy consumption of both with fog and without fog scenarios increases slightly, but the *with fog mean* remains consistently higher than the *without fog mean*. With a *snapshot interval* of 5, the IoT devices are sending data less frequently, which results in reduced energy consumption. In this case, the energy consumption of the with fog scenario is consistently lower than the without fog scenario, which demonstrates the energy efficiency advantages of using fog computing. With a *snapshot interval* of 10, the IoT devices send data even less frequently, and the difference in energy consumption between the with fog and without fog scenarios becomes more pronounced. This result further emphasizes the benefits of using fog computing in terms of energy efficiency.

Number of Devices

The *number of devices* parameter refers to the number of telehealth IoT devices in the network. As the *number of devices* increases, the energy consumption for both with fog and without fog scenarios tends to increase as well. This is expected, as more devices lead to higher data transmission and processing loads. However, the increase in energy consumption is consistently smaller in the with fog scenario compared to the without fog scenario across all snapshot intervals. This shows that the proposed fog-based model is more scalable and can better handle the energy requirements of a growing number of devices.

With Fog Mean and Without Fog Mean

The *with fog mean* and *without fog mean* parameters represent the average energy consumption in the scenarios with and without fog computing, respectively. Across all snapshot intervals and several devices, the *with fog mean* is generally

lower than the *without fog mean*, indicating that the fog-based model is more energy-efficient than the cloud-only model.

With Fog SD and Without Fog SD

The *with fog SD* and *without fog SD* parameters represent the SD of the energy consumption in the scenarios with and without fog computing, respectively. In general, the SD values are lower in the with fog scenario compared to the without fog scenario. This suggests that the energy consumption is more consistent and less variable in the fog-based model, which could lead to more predictable and stable system performance.

With Fog 95% CI and Without Fog 95% CI

The CI in the simulation code is a range within which a certain percentage of the population parameter is expected to lie, with a specified level of confidence. In the context of the provided simulation results, the 95% CIs represent the range within which the true mean performance of the system (either with or without fog computing) is likely to fall, with a certain level of confidence, typically 95%.

A 95% CI is calculated using the sample mean, sample SD, and sample size. The formula for a 95% CI is:

$$CI = \text{sample mean} \pm (1.96 \times [\text{sample SD}/\sqrt{\text{sample size}}])$$

The 95% CI helps to quantify the uncertainty in the estimation of the true mean performance. A narrower 95% CI indicates a more precise estimate, while a wider interval suggests more uncertainty.

Analysis of Results

Table 1 contains the summary of statistical results.

Table 1. Summary of statistical results.

Snapshot interval	Number of devices	With fog, mean (SD)	With fog, 95% CI	Without fog, mean (SD)	Without fog, 95% CI
1	10	90.43 (0.45)	90.11-90.76	89.74 (0.05)	89.69-89.79
1	20	90.53 (0.33)	90.30-90.77	89.74 (0.06)	89.69-89.79
1	30	90.61 (0.23)	90.45-90.78	89.74 (0.04)	89.71-89.78
1	40	90.55 (0.24)	90.38-90.72	89.76 (0.05)	89.71-89.90
5	10	87.39 (0.70)	86.89-87.89	86.04 (0.13)	85.94-86.13
5	20	86.59 (0.21)	86.44-86.73	85.91 (0.06)	85.86-85.95
5	30	87.02 (0.46)	86.70-87.34	86.01 (0.09)	85.95-86.08
5	40	87.30 (0.27)	87.12-87.50	86.00 (0.07)	85.95-86.05
10	10	82.85 (0.73)	82.34-83.38	81.36 (0.11)	81.27-81.44
10	20	83.28 (0.63)	82.83-83.72	81.42 (0.11)	81.33-81.50
10	30	82.62 (0.59)	82.80-83.03	81.33 (0.12)	81.24-81.43
10	40	82.7 (0.37)	82.43-82.95	81.36 (0.07)	81.31-81.41

Here is a step-by-step analysis of the results (Table 1):

1. Observe the “With fog, mean (SD)” and “Without fog, mean (SD)” columns for each combination of “Snapshot interval” and “Number of devices.” In all cases, the *with fog mean* is higher than the *without fog mean*, indicating that, on average, the remaining energy is higher when using fog computing.
2. Look at the 95% CIs for both “with fog” and “without fog” scenarios. If the 95% CIs do not overlap, it suggests that the difference in energy remaining between the 2 scenarios is statistically significant. For example, in the first row

(snapshot interval: 1, number of devices: 10), the “with fog, 95% CI” is 87.98-89.45, and the “without fog, 95% CI” is 84.90-87.47. Since these intervals do not overlap, there is strong evidence that using fog computing leads to significantly higher energy remaining for this specific combination of parameters.

3. Compare the width of the 95% CIs for each scenario. A narrower 95% CI indicates a more precise estimate of the true population mean. For most 95% CI values, the “with fog, 95% CI” is narrower than the “without fog, 95% CI,” suggesting that the “with fog” scenario has a more precise estimate.
4. Analyze the trends as the number of devices increases within each snapshot interval. In general, the energy remaining in both scenarios decreases as the number of devices increases. However, the rate of decrease seems to be lower when using fog computing.
5. Observe the trends as the snapshot interval increases for each group of devices. As the snapshot interval increases, the energy remaining for both scenarios decreases, suggesting that less frequent snapshots may lead to less energy conservation. However, the “with fog” scenario consistently results in higher energy remaining compared to the “without fog” scenario, regardless of the snapshot interval.

In conclusion, based on the analysis of the means and 95% CIs, it appears that using fog computing is beneficial for conserving energy, especially when the number of devices and the snapshot intervals increase. The difference in energy remaining is statistically significant in most cases, and the “with fog” scenario consistently outperforms the “without fog” scenario.

Therefore, the simulation results demonstrate that the proposed fog-based telehealth model provides improved energy efficiency and scalability compared to a cloud-only model, especially when the IoT devices send data less frequently. The lower energy consumption and SD values in the with fog scenario indicate that fog computing is a viable solution for managing energy requirements and maintaining consistent performance in telehealth IoT networks. Furthermore, we conducted the sensitivity simulation analysis to systematically investigate the impact of variations in model parameters on the simulation outcomes. Sensitivity analysis helps in understanding how different input parameters influence the system’s behavior and performance and identifies critical factors that have a significant effect on the results. According to the simulation code running, the sensitivity analysis was performed for various parameters such as *transmit_power*, *idle_power*, *latency*, and *energy_cost*. By varying these parameters across a range of values, the impact on the energy remaining in IoT devices with and without fog nodes can be evaluated.

Table 2 compares the mean energy remaining for IoT devices with and without fog nodes for each energy cost value. The “Mean difference” column shows the difference in mean energy remaining, with positive values indicating that devices with fog nodes have higher energy remaining compared to those without fog nodes. In the with fog scenario, the mean energy remaining for devices with fog nodes stays relatively stable, ranging from a minimum of 93 to a maximum of 95 across different energy costs. In the without fog scenario, the mean energy remaining for devices without fog nodes also remains relatively stable, ranging from a minimum of 91 to a maximum of 92 across different energy costs.

Table 2. Sensitivity analysis with energy cost.

Energy cost	With fog, mean (SD)	Without fog, mean (SD)	Mean difference
0.20	94 (5)	91 (4)	1.72
0.26	93 (6)	92 (5)	0.33
0.32	94 (4)	92 (7)	1.72
0.38	93 (5)	91 (2)	1.72
0.44	94 (3)	91 (5)	2.75
0.5	93 (2)	91 (4)	1.71
0.56	95 (2)	92 (2)	1.72
0.62	94 (3)	92 (4)	1.37
0.68	93 (2)	92 (2)	0.68
0.74	94 (4)	92 (3)	1.02
0.80	95 (6)	92 (3)	2.06

Based on the sensitivity analysis of energy cost, the mean energy remaining for IoT devices with fog nodes is consistently higher than that of devices without fog nodes across all energy cost values. This indicates that IoT devices with fog nodes perform better in terms of energy consumption as compared to devices without fog nodes.

Table 3 compares the mean energy remaining for IoT devices with and without fog nodes for each latency parameter value.

The “Mean difference” column shows the difference in mean energy remaining, with positive values indicating that devices with fog nodes have higher energy remaining compared to those without fog nodes. In the with fog scenario, the mean energy remaining for devices with fog nodes stays relatively stable, ranging from a minimum of 94 to a maximum of 95 across different latency values. In the without fog scenario, the mean energy remaining for devices without fog nodes also remains

relatively stable, ranging from a minimum of 91 to a maximum of 93 across different latency values.

Based on the sensitivity analysis of latency, the mean energy remaining for IoT devices with fog nodes is consistently higher than that of devices without fog nodes across all latency parameter values. This indicates that IoT devices with fog nodes perform better in terms of energy consumption as compared to devices without fog nodes.

Table 4 compares the mean energy remaining for IoT devices with and without fog nodes for each idle power value. The

“Mean difference” column shows the difference in mean energy remaining, with positive values indicating that devices with fog nodes have higher energy remaining compared to those without fog nodes. In the with fog scenario, the mean energy remaining for devices with fog nodes stays relatively stable, ranging from a minimum of 93 to a maximum of 95 across different idle power values. In the without fog scenario, the mean energy remaining for devices without fog nodes also remains relatively stable, ranging from a minimum of 90 to a maximum of 92 across different idle power values.

Table 3. Sensitivity analysis with latency.

Latency parameter	With fog, mean (SD)	Without fog, mean (SD)	Mean difference
0.20	94 (4)	92 (4)	2.06
0.26	93 (3)	91 (4)	2.06
0.32	94 (2)	93 (1)	1.72
0.38	95 (2)	92 (2)	2.06
0.44	94 (2)	92 (2)	1.72
0.5	94 (3)	91 (3)	2.4
0.56	94 (4)	92 (5)	2.06
0.62	94 (2)	92 (2)	2.06
0.68	94 (3)	92 (3)	1.71
0.74	94 (5)	93 (3)	0.68
0.80	94 (2)	92 (4)	0.68

Table 4. Sensitivity analysis with idle power.

Idle power	With fog, mean (SD)	Without fog, mean (SD)	Mean difference
0.5	95 (2)	92 (2)	2.06
0.6	94 (2)	92 (4)	1.71
0.7	95 (3)	92 (4)	2.41
0.8	93 (2)	90 (3)	1.72
0.9	94 (4)	92 (4)	1.03
1.0	93 (5)	91 (4)	1.02
1.1	95 (4)	92 (6)	1.71
1.2	94 (4)	91 (5)	2.06
1.3	95 (2)	91 (2)	2.75
1.4	94 (1)	91 (4)	2.06
1.5	94 (2)	92 (4)	1.72

Based on the sensitivity analysis of idle power, the mean energy remaining for IoT devices with fog nodes is consistently higher than that of devices without fog nodes across all idle power values. This indicates that IoT devices with fog nodes perform better in terms of energy consumption as compared to devices without fog nodes.

Table 5 compares the mean energy remaining for IoT devices with and without fog nodes for each transmit power value. The “Mean difference” column shows the difference in mean energy

remaining, with positive values indicating that devices with fog nodes have higher energy remaining compared to those without fog nodes. In the with fog scenario, the mean energy remaining for devices with fog nodes stays relatively stable, ranging from a minimum of 94 to a maximum of 96 across different transmit power values. In the without fog scenario, the mean energy remaining for devices without fog nodes also remains relatively stable, ranging from a minimum of 91 to a maximum of 92 across different transmit power values.

Table 5. Sensitivity analysis with transmit power.

Transmit power	With fog, mean (SD)	Without fog, mean (SD)	Mean difference
0.5	94 (3)	91 (2)	1.37
0.6	95 (4)	92 (3)	2.41
0.7	94 (2)	92 (4)	1.37
0.8	94 (3)	91 (5)	1.71
0.9	94 (2)	92 (4)	1.37
1.0	94 (5)	91 (2)	2.06
1.1	94 (6)	92 (3)	1.37
1.2	94 (6)	92 (3)	1.02
1.3	95 (4)	92 (3)	2.40
1.4	96 (2)	91 (2)	3.79
1.5	95 (2)	91 (1)	3.44

Based on the sensitivity analysis of transmit power, the mean energy remaining for IoT devices with fog nodes is consistently higher than that of devices without fog nodes across all transmit power values. This indicates that IoT devices with fog nodes perform better in terms of energy consumption as compared to devices without fog nodes.

Ethical Considerations

The study did not apply for any ethical approval, as the research did not involve any human participants or animals [44].

Discussion

Overview

The simulation study results indicate that the proposed energy-saving model could be effective in reducing energy consumption in real-world telehealth scenarios. Key findings include the following:

1. **Scalability:** the model demonstrates the ability to accommodate an increasing number of IoT devices without compromising performance, energy efficiency, or quality of health care services.
2. **Task allocation algorithm:** the proposed task allocation algorithm outperforms other algorithms in terms of energy efficiency and data processing efficiency, indicating its effectiveness in balancing the workload between fog nodes and cloud servers.
3. **Energy consumption metrics:** the overall energy consumption is reduced across all levels, demonstrating the success of the model's energy-saving strategies, such as adaptive power management, data compression, and network optimization.

The code and methodology described aim to simulate an IoT network with different components (IoT devices, fog nodes, and cloud servers) and analyze the impact of fog nodes on energy consumption. The code creates and connects these components and simulates data transmission, storage, and energy consumption for IoT devices, fog nodes, and cloud servers. The simulation results are analyzed to understand the network

behavior and demonstrate the potential benefits of using fog nodes for energy efficiency.

Our novel energy-efficient model integrates fog and cloud computing paradigms to optimize data processing for telehealth IoT devices without compromising real-time health care services. This stands out from previous works by enabling localized data processing through the incorporation of fog computing. This intermediary layer, situated between IoT devices and cloud servers, effectively reduces latency and data transfer overhead. The concurrent use of public and private cloud computing further fortifies the system's infrastructure, allowing for the handling of large data volumes and resource-intensive computations. The model enables localized data processing by incorporating fog computing as an intermediary layer between IoT devices and public or private cloud servers, effectively reducing latency and data transfer overhead. Simultaneously, public and private cloud computing provides a robust infrastructure for handling large data volumes and performing resource-intensive computations. The primary goal of this model is to minimize energy consumption through intelligent task allocation between fog nodes and cloud servers, by considering their computational capacity and proximity to IoT devices. This task allocation process also considers various sensitivity and priority levels within the health care context, ensuring prompt responses to critical and high-sensitivity requests. Our innovative model strategically integrates fog and cloud computing, aiming to establish an energy-efficient telehealth IoT system capable of adeptly managing data processing and delivering real-time health care services, accommodating various levels of sensitivity and priorities. While these aspirations suggest promising opportunities for further optimization and diverse applications within health care contexts, it is crucial to note that the subsequent simulation method serves to objectively assess the model's effectiveness and efficiency. The empirical evidence derived from the simulation provides a foundation for a more nuanced understanding of the model's capabilities and potential benefits. This is because exploring diverse large-scale network topologies is rarely feasible in the real world. Although the requirements for such a simulator are straightforward—providing a detailed,

accurate, and granular model of all components—implementing corresponding simulators demands considerable effort.

The primary strength of our model lies in its holistic approach toward minimizing energy consumption. The intelligent task allocation mechanism, considering computational capacity and proximity to IoT devices, ensures a fine balance. Furthermore, the incorporation of sensitivity and priority levels within the health care context enhances the model's responsiveness to critical requests. The synergistic integration of fog and cloud computing contributes to the creation of an energy-efficient telehealth IoT system capable of real-time data processing in accordance with varying sensitivity levels and priorities.

Despite the positive outcomes, several limitations should be acknowledged. (1) Simulation environment realism: the simulation, while essential for its controlled environment, may not perfectly mirror real-world complexities. Variations in network behaviors and external factors may influence results differently in practical implementations. (2) Sensitivity analysis scope: the sensitivity analysis, while comprehensive, focused on specific parameters such as energy cost, latency, idle power, and transmit power. Additional parameters and their potential interactions may provide a more nuanced understanding of the model's behavior. (3) Simplifications in simulation: certain simplifications, inherent in simulation models, may oversimplify the intricacies of a live telehealth IoT deployment. Real-world complexities such as device failures, communication errors, or dynamic changes in the environment are challenging to fully capture.

To address these limitations and advance the research, the following suggestions should be considered. (1) Future studies should aim for more realistic simulation environments, incorporating dynamic factors and diverse network topologies to enhance the model's external validity. (2) Expanding the scope of sensitivity analysis to include a broader range of parameters and exploring their interactions could provide a more comprehensive understanding of the model's performance under diverse conditions. (3) The development of more sophisticated simulators, despite their challenges, remains crucial. Detailed, accurate, and granular models of all components can better simulate the intricacies of large-scale IoT-fog-cloud systems.

While our model exhibits significant promise in reducing energy consumption and enhancing data processing efficiency in telehealth IoT scenarios, ongoing refinement and exploration of diverse scenarios will contribute to its continued evolution and real-world applicability.

Conclusion

This paper provides a compelling model for the use of fog and cloud computing-based platforms in telehealth IoT deployments to reduce energy consumption, improve data processing efficiency, and maintain high-quality health care services. The model leverages the strengths of both fog and cloud computing paradigms to address the challenges associated with large-scale telehealth IoT deployments, such as energy consumption, data processing efficiency, latency, security, and privacy. The simulation results show that the proposed fog-based model significantly reduces energy consumption compared to the cloud-only model while maintaining high-quality data processing and transmission. Moreover, the methodology described in this paper provides a comprehensive approach to analyzing network performance and energy consumption, which includes examining the impact of various parameters, such as the number of devices, fog node deployment, task allocation algorithm, energy consumption metrics, and performance metrics. Sensitivity analyses were conducted with respect to energy cost, latency, idle power, and transmit power, consistently showing that IoT devices with fog nodes had higher mean energy remaining compared to devices without fog nodes. This approach allows for a more detailed understanding of the network behavior and potential bottlenecks and provides insights into how to optimize the model to be more resilient and efficient. The simulation results and methodology demonstrate the effectiveness of the proposed model and provide a roadmap for future research in this area. We demonstrated the effectiveness of the proposed model in reducing energy consumption while, more importantly, ensuring efficient data processing and maintaining the quality of health care services. The proposed model can help health care providers and stakeholders improve patient care and outcomes while reducing costs and energy consumption.

Acknowledgments

This work was supported by the National Natural Science Foundation of Canada and the Department of Computer Science, University of Victoria.

Conflicts of Interest

None declared.

References

1. Atlam HF, Walters RJ, Wills GB. Fog computing and the internet of things: a review. *Big Data Cogn Comput* 2018;2(2):10 [FREE Full text] [doi: [10.3390/bdcc2020010](https://doi.org/10.3390/bdcc2020010)]
2. Motlagh NH, Mohammadrezaei M, Hunt J, Zakeri B. Internet of things (IoT) and the energy sector. *Energies* 2020;13(2):494 [FREE Full text] [doi: [10.3390/en13020494](https://doi.org/10.3390/en13020494)]
3. Wootton R. Telemedicine. *Br J Hosp Med* 2012;73(9):504-507. [Medline: [22968584](https://pubmed.ncbi.nlm.nih.gov/22968584/)]

4. Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak KS. The internet of things for health care: a comprehensive survey. *IEEE Access* 2015;3:678-708 [[FREE Full text](#)] [doi: [10.1109/ACCESS.2015.2437951](https://doi.org/10.1109/ACCESS.2015.2437951)]
5. Kumar P, Patil K, Lee JH, Lee HJ. IoT-based remote patient monitoring: a survey on the capabilities, challenges, and future directions. *Electronics* 2020;9(10):1702. [doi: [10.3390/electronics9101702](https://doi.org/10.3390/electronics9101702)]
6. Silva BMC, Rodrigues JJPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. *J Biomed Inform* 2015;56:265-272 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2015.06.003](https://doi.org/10.1016/j.jbi.2015.06.003)] [Medline: [26071682](https://pubmed.ncbi.nlm.nih.gov/26071682/)]
7. Bonomi F, Milito R, Zhu J, Addepalli S. Fog computing and its role in the internet of things. 2012 Presented at: MCC '12: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing; August 17, 2012; Helsinki, Finland p. 13-16. [doi: [10.1145/2342509.2342513](https://doi.org/10.1145/2342509.2342513)]
8. Yi S, Li C, Li Q. A survey of fog computing: concepts, applications and issues. 2015 Presented at: Mobidata '15: Proceedings of the 2015 Workshop on Mobile Big Data; June 21, 2015; Hangzhou, China p. 37-42. [doi: [10.1145/2757384.2757397](https://doi.org/10.1145/2757384.2757397)]
9. Aazam M, Khan I. Cloud of things: integrating internet of things and cloud computing and the issues involved. 2014 Presented at: Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST); January 14-18, 2014; Islamabad, Pakistan p. 414-419. [doi: [10.1109/ibcast.2014.6778179](https://doi.org/10.1109/ibcast.2014.6778179)]
10. Dubey H, Yang J, Constant N, Mankodiya K. Fog data: enhancing telehealth big data through fog computing. 2017 Presented at: ASE BD&SI '15: Proceedings of the ASE BigData & SocialInformatics 2015; October 7-9, 2015; Kaohsiung, Taiwan p. 1-6. [doi: [10.1145/2818869.2818889](https://doi.org/10.1145/2818869.2818889)]
11. Skarlat O, Schulte S, Borkowski M, Leitner P. Resource provisioning for IoT services in the fog. 2017 Presented at: IEEE 5th International Conference on Future Internet of Things and Cloud; August 21-23, 2017; Prague p. 138-144. [doi: [10.1109/soca.2016.10](https://doi.org/10.1109/soca.2016.10)]
12. Kuo AMH. Opportunities and challenges of cloud computing to improve health care services. *J Med Internet Res* 2011;13(3):e67 [[FREE Full text](#)] [doi: [10.2196/jmir.1867](https://doi.org/10.2196/jmir.1867)] [Medline: [21937354](https://pubmed.ncbi.nlm.nih.gov/21937354/)]
13. Rodrigues JJPC, de la Torre I, Fernández G, López-Coronado M. Analysis of the security and privacy requirements of cloud-based electronic health records systems. *J Med Internet Res* 2013;15(8):e186 [[FREE Full text](#)] [doi: [10.2196/jmir.2494](https://doi.org/10.2196/jmir.2494)] [Medline: [23965254](https://pubmed.ncbi.nlm.nih.gov/23965254/)]
14. Hussain M, Chen F, Ali A. Cloud-based telehealth system for integrated services in smart cities. *IEEE Commun Mag* 2015;53(12):68-73. [doi: [10.1109/MCOM.2015.7321979](https://doi.org/10.1109/MCOM.2015.7321979)]
15. Jalali F, Hinton K, Ayre R, Alpcan T, Tucker RS. Fog computing may help to save energy in cloud computing. *IEEE J Select Areas Commun* 2016;34(5):1728-1739 [[FREE Full text](#)] [doi: [10.1109/JSAC.2016.2545559](https://doi.org/10.1109/JSAC.2016.2545559)]
16. Tuli S, Basumatary N, Gill SS, Kahani M, Arya RC, Wander GS, et al. HealthFog: an ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments. *Future Gener Comput Syst* 2020;104:187-200. [doi: [10.1016/j.future.2019.10.043](https://doi.org/10.1016/j.future.2019.10.043)]
17. Orsini G, Bade D, Lamersdorf W. Computing in the fog: a real-world IoT architecture for smart cities. 2016 Presented at: 2016 IEEE Global Communications Conference (GLOBECOM); December 4-8, 2016; Washington, DC p. 1-6.
18. Makhdoom I, Abolhasan M, Lipman J, Liu RP, Ni W. Anatomy of threats to the internet of things. *IEEE Commun Surv Tutor* 2018;20(4):2964-2996. [Medline: [29744278](https://pubmed.ncbi.nlm.nih.gov/29744278/)]
19. Khan Z, Anjum A, Soomro K, Tahir MA. Towards cloud-based smart cities data security and privacy management. *Future Gener Comput Syst* 2017;77:437-448. [Medline: [29201963](https://pubmed.ncbi.nlm.nih.gov/29201963/)]
20. Gormus U, Kansal A, Srivastava MB. Energy-aware lossless data compression. *ACM Trans Sens Netw* 2011;7(3):1-25. [Medline: [22081221](https://pubmed.ncbi.nlm.nih.gov/22081221/)]
21. Chiang M, Zhang T. Fog and IoT: an overview of research opportunities. *IEEE Internet Things J* 2016;3(6):854-864. [doi: [10.1109/JIOT.2016.2584538](https://doi.org/10.1109/JIOT.2016.2584538)]
22. Sarkar S, Chatterjee S, Misra S. Assessment of the suitability of fog computing in the context of internet of things. *IEEE Trans Cloud Comput* 2018;6(1):46-59. [doi: [10.1109/TCC.2015.2485206](https://doi.org/10.1109/TCC.2015.2485206)]
23. Abdel-Basset M, Chang V, Hawash H. A fusion of cloud-fog based smart city model for efficient performance. *Comput Mater Continua* 2019;57(3):337-355.
24. Morabito R. Virtualization on internet of things edge devices with container technologies: a performance evaluation. *IEEE Access* 2017;5:8835-8850 [[FREE Full text](#)] [doi: [10.1109/ACCESS.2017.2704444](https://doi.org/10.1109/ACCESS.2017.2704444)]
25. Tuli S, Gill SS, Arya RC, Shojafar M, Buyya R. FogBus: a lightweight and QoS-aware framework for the internet of things. *IEEE Internet Things J* 2020;7(5):4493-4510.
26. Gupta H, Nath AR, Chakraborty S, Ghosh SK. iFogSim: a toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments. *Software Pract Exper* 2016;47(9):1275-1296.
27. Oueis J, Strinati EC, Barbarossa S. The fog balancing: load distribution for small cell cloud computing. 2015 Presented at: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring); May 11-14, 2015; Glasgow, UK p. 1-6. [doi: [10.1109/VTCspring.2015.7146129](https://doi.org/10.1109/VTCspring.2015.7146129)]
28. Barcelo M, Correa A, Llorca J, Tulino AM, Vicario JL, Morell A. IoT-cloud service optimization in next generation smart environments. *IEEE J Select Areas Commun* 2016;34(12):4077-4090. [doi: [10.1109/JSAC.2016.2621398](https://doi.org/10.1109/JSAC.2016.2621398)]

29. Zeng X, Garg S, Strazdins P. A comparative study of IoT cloud and fog computing simulations using iFogSim and Cooja. 2017 Presented at: 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC); May 16-18, 2017; Calabria, Italy p. 108-113.
30. Byers CC, Wetterwald P. Fog computing distributing data and intelligence for resiliency and scale necessary for IoT: the Internet of Things (Ubiquity symposium). Ubiquity 2015;2015(November):1-12 [[FREE Full text](#)] [doi: [10.1145/2836953.2836960](https://doi.org/10.1145/2836953.2836960)]
31. Bermejo PJ, Rodríguez S, Valladares DR, Boubeta-Puig J. YAFS: a simulator for IoT scenarios in fog computing. IEEE Access 2020;8:111908-111922.
32. García AM, Pérez JP, Bellido OJ. YAFS: a simulator for IoT scenarios in fog computing. 2018 Presented at: 2018 IEEE International Conference on Smart Computing (SMARTCOMP); June 18-20, 2018; Taormina, Sicily, Italy p. 215-222.
33. Rodríguez S, Bermejo PJ, Boubeta-Puig J. A comparative study of IoT application deployment on YAFS, iFogSim, and EdgeCloudSim simulators. 2019 Presented at: 2019 IEEE International Conference on Edge Computing (EDGE); July 8-13, 2019; Milan, Italy p. 55-62.
34. Aazam M, Huh EN. Fog computing and smart gateway based communication for cloud of things. 2015 Presented at: 2015 International Conference on Future Internet of Things and Cloud (FiCloud); August 24-26, 2015; Rome, Italy p. 464-470.
35. Verma P, Sood SK. Fog assisted-IoT enabled patient health monitoring in smart homes. IEEE Internet Things J 2017;5(3):1789-1796.
36. Koubaâ A, Qureshi B, Sriti MF, Allouch A, Javed Y. A fog-based emergency and healthcare system for smart cities. 2019 Presented at: 2019 IEEE Global Communications Conference (GLOBECOM); December 9-13, 2019; Waikoloa, HI, USA p. 1-6.
37. Sareen S, Sood SK, Gupta SK. Energy-efficient context-aware framework for managing application execution in cloud-fog environment. J Commun Netw Distrib Syst 2015;15(2-3):137-154.
38. Alam MM, Reaz MBI, Ashraf MA, Hossain MS, Islam MT, Otoum S. Telehealth IoT devices integrated with fog nodes and private/public cloud architecture model. IEEE Access 2019;7:108636-108647.
39. Understanding blood pressure readings. American Heart Association. 2017. URL: <https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings> [accessed 2024-02-17]
40. Target heart rates chart. American Heart Association. 2017. URL: <https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates> [accessed 2024-02-17]
41. Respiratory rate. American Lung Association. 2018. URL: <https://www.lung.org/lung-health-diseases/lung-procedures-and-tests/respiratory-rate> [accessed 2024-02-17]
42. Vital signs. Centers for Disease Control and Prevention. 2021. URL: <https://www.cdc.gov/vitalsigns/index.html> [accessed 2024-02-17]
43. Oxygen saturation. MedlinePlus. 2021. URL: <https://medlineplus.gov/oxygenlevels.html> [accessed 2024-02-17]
44. Get ethics approval. University of Victoria Office of Research Services. URL: <https://www.uvic.ca/research-services/how-do-i/get-ethics-approval/index.php>

Abbreviations

IoT: Internet of Things

YAFS: Yet Another Fog Simulator

Edited by T Leung; submitted 21.06.23; peer-reviewed by A Al-Muqarm, R Bidkar, A Anastasiou, A Nayyar, R Gore; comments to author 22.01.24; revised version received 23.01.24; accepted 15.02.24; published 06.03.24.

Please cite as:

Guo Y, Ganti S, Wu Y

Enhancing Energy Efficiency in Telehealth Internet of Things Systems Through Fog and Cloud Computing Integration: Simulation Study

JMIR Biomed Eng 2024;9:e50175

URL: <https://biomedeng.jmir.org/2024/1/e50175>

doi: [10.2196/50175](https://doi.org/10.2196/50175)

PMID: [38875671](https://pubmed.ncbi.nlm.nih.gov/38875671/)

©Yunyong Guo, Sudhakar Ganti, Yi Wu. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 06.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic

information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

An Engineering Alternative to Lockdown During COVID-19 and Other Airborne Infectious Disease Pandemics: Feasibility Study

Yusaku Fujii¹, PhD

School of Science and Technology, Gunma University, Kiryu, Japan

Corresponding Author:

Yusaku Fujii, PhD

School of Science and Technology

Gunma University

1-5-1 Tenjin-cho

Kiryu, 3768515

Japan

Phone: 81 8035505585

Fax: 81 277301757

Email: fujii@gunma-u.ac.jp

Abstract

Background: Now and in the future, airborne diseases such as COVID-19 could become uncontrollable and lead the world into lockdowns. Finding alternatives to lockdowns, which limit individual freedoms and cause enormous economic losses, is critical.

Objective: The purpose of this study was to assess the feasibility of achieving a society or a nation that does not require lockdown during a pandemic due to airborne infectious diseases through the mass production and distribution of high-performance, low-cost, and comfortable powered air purifying respirators (PAPRs).

Methods: The feasibility of a social system using PAPR as an alternative to lockdown was examined from the following perspectives: first, what PAPRs can do as an alternative to lockdown; second, how to operate a social system utilizing PAPR; third, directions of improvement of PAPR as an alternative to lockdown; and finally, balancing between efficiency of infection control and personal freedom through the use of Internet of Things (IoT).

Results: PAPR was shown to be a possible alternative to lockdown through the reduction of airborne and droplet transmissions and through a temporary reduction of infection probability per contact. A social system in which individual constraints imposed by lockdown are replaced by PAPRs was proposed, and an example of its operation is presented in this paper. For example, the government determines the type and intensity of the lockdown and activates it. At that time, the government will also indicate how PAPR can be substituted for the different activity and movement restrictions imposed during a lockdown, for example, a curfew order may be replaced with the permission to go outside if wearing a PAPR. The following 7 points were raised as directions for improvement of PAPR as an alternative method to lockdown: flow optimization, precise differential pressure control, design improvement, maintenance method, variation development such as booth type, information terminal function, and performance evaluation method. In order to achieve the effectiveness and efficiency in controlling the spread of infection and the individual freedom at a high level in a social system that uses PAPRs as an alternative to lockdown, it was considered effective to develop a PAPR wearing rate network management system utilizing IoT.

Conclusions: This study shows that using PAPR with infection control ability and with less economic and social damage as an alternative to nationwide lockdown is possible during a pandemic due to airborne infectious diseases. Further, the efficiency of the government's infection control and each citizen's freedom can be balanced by using the PAPR wearing rate network management system utilizing an IoT system.

(*JMIR Biomed Eng* 2024;9:e54666) doi:[10.2196/54666](https://doi.org/10.2196/54666)

KEYWORDS

COVID-19; airborne infectious diseases; lockdown; powered air purifying respirator (PAPR); infectious dose; airborne transmission; emergency evacuation; herd immunity; pandemic; aerosol; air; quality; infection control; infectious; respiratory; purifier; purifiers; purifying; respirator; respirators; device; devices; airborne

Introduction

For more than 3 years, herd immunity has been pursued worldwide through vaccination as a countermeasure against COVID-19, in addition to new lifestyle measures (social distancing, wearing of face masks, washing of hands, etc) [1,2]. However, due to the emergence of new variants, the spread of infection has sometimes been uncontrollable in different parts of the world. Each time, lockdown has been implemented to temporarily buy time, causing great economic loss and restriction of freedom for individuals, businesses, and society [3-7]. In this study, lockdown is defined as restricting the actions and activities of people and businesses to temporarily slow the spread of infection and buy time for other measures such as herd immunity through vaccination. A strong lockdown includes orders with criminal penalties that prohibit going out, working, and doing business, and a weak lockdown includes voluntary requests to refrain from going out, working, and doing business. The type and strength of lockdowns are determined by the government on a case-by-case basis, considering the impact on infection control and harm to society, depending on the situation of the outbreak of infection at the time [8,9].

Even in the current situation, the rapid development and delivery of effective vaccines against new variants of the coronavirus (SARS-CoV-2) and new airborne viruses that may emerge in succession is not well assured. In the future, it is likely that we will continue to be in a situation where we do not know when a lockdown will again be necessary around the world [10,11]. In light of this, alternatives to lockdown that can cause less economic damage, avoid restrictions on freedom of action, and other disadvantages to individuals, companies, and societies would be beneficial. Among the modes of COVID-19 transmission, contact and oral infections are relatively easy to prevent by environmental hygiene, including handwashing and food hygiene management. Droplet infection (particle size $\geq 100 \mu\text{m}$) is thought to be preventable by social distancing (droplets fall by gravity) and by wearing a mask. Currently, airborne transmission via aerosols (particle size $< 100 \mu\text{m}$) is thought to be the main route of infection [12,13]. Although the infectious dose of COVID-19, that is, the number of ingested viruses required for infection, is not well known [14], it is estimated to be in the range of 300 to 2000 virions [15]. It is believed that the possibility of viral infection can be effectively reduced by shielding aerosols that may contain viruses. This paper discusses alternative means to lockdown, assuming that contact and oral infections are prevented by environmental hygiene and food hygiene management and that only airborne and droplet infections remain as infection routes.

Methods

The feasibility of a social system utilizing powered air purifying respirators (PAPRs) as an alternative to lockdown was examined from the following perspectives: (1) what PAPR can do as an alternative to lockdown, (2) how to operate a social system utilizing PAPR, (3) direction of improvement of PAPR as an alternative to lockdown, and (4) balance between efficiency of

infection control and personal freedom through the use of Internet of Things (IoT).

What PAPR Can Do as an Alternative to Lockdown

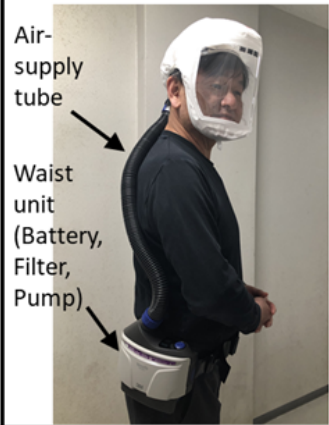
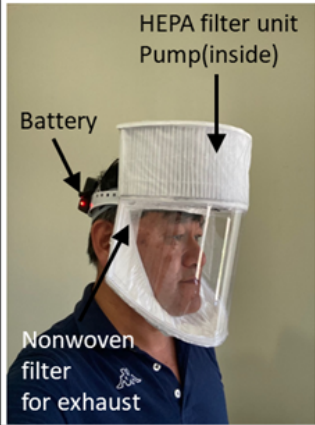
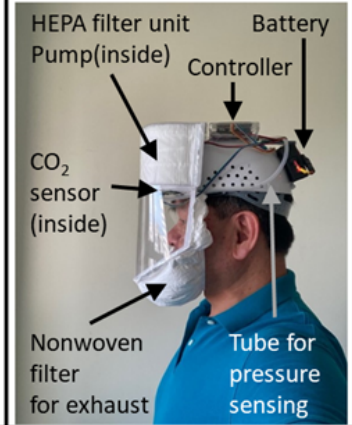
The PAPR for practical medical use is a device that drastically reduces the number of viruses inhaled by the wearer and effectively reduces the risk of infection. Medical PAPRs are used by medical personnel working in high-risk environments [16,17]. The assigned protection factor (APF), as defined by the National Institute for Occupational Safety and Health in the United States, is widely used as an indicator of the shielding performance of respiratory protection devices, including PAPRs [18]. APF is defined as the external concentration/internal concentration of the target particles (aerosols). For a medical face mask, APF=10 is given when a person who has been trained to wear it wears it completely with no gaps between the mask and the facial surface. However, 3M PAPR is rated at APF=1000 and is considered to have excellent protective performance [19]. In other words, the concentration of virus-containing aerosols can be reduced to 1/10 or less with a full-face mask when worn without gaps, while it is reduced to 1/1000 or less with 3M PAPR. Therefore, high-performance PAPRs can be used as an alternative to the movement restrictions and activity restrictions imposed by lockdowns used as a countermeasure against COVID-19 and other airborne infectious diseases in future pandemics. The necessary conditions for an alternative to lockdown are that it should have the same deterrent effect on the spread of infection as lockdown, and the economic damage, activity restrictions, and other disadvantages to individuals and society should be smaller compared to those during lockdown.

The effective reproduction number R_t is the average number of secondary cases per infectious case in a population of both susceptible and nonsusceptible individuals. In this population, $R_t < 1$ means converging, $R_t = 1$ means stationary, and $R_t > 1$ means expanding [20]. The effective reproduction number R_t can be expressed schematically by the following equation.

$R_t = \beta \times k \times D$, where β =probability of infection being transmitted during a contact, k (contact/day)=contact rate in the host population, and D (day)=duration of infectiousness.

To control the infection of the whole society, it is sufficient to set the effective reproduction number (R_t) to < 1 . To do this, we should reduce the above β , k , and D . Vaccines are expected to reduce β in the long term. Lockdown is expected to play a role in temporarily reducing k . In this study, PAPR is expected to play a role in temporarily reducing β by reducing the airborne and droplet transmission. During COVID-19, the maximum R_t value reported worldwide was around 5 [21,22]. If lockdown is used as a means of correcting $R_t=5$ to $R_t=1$, it is sufficient if the lockdown makes k 1/5 of its current value. Alternatively, if PAPRs are used as an alternative to lockdown, it would be sufficient if citizens could wear PAPRs to reduce β to 1/5 of its current value. It is worth noting here that the degree of reduction in k and β may be about 1/5 rather than 1/100 or 1/1000. The performance of the existing PAPRs was evaluated. Figure 1 shows the specifications of 3 existing PAPRs.

Figure 1. Specifications of the 3 existing powered air purifying respirators. HEPA: high efficiency particulate air.

Model	Versaflo TR-301N+ (3M)	Simple model (prototype)	Controller model (prototype)
Photo			
Filter	Nonwoven filter	HEPA filter	HEPA filter
Internal Pressure	positive	positive	positive (controllable)
Flow rate (L/min)	180 (Low) or 200 (High)	400	300-400
Computer	not applicable	not applicable	Board computer
Sensor	not applicable	not applicable	2 CO ₂ & 1 pressure sensors
Cost	Price: US \$1000	Parts cost: US \$40	Parts cost: US \$300

Currently Existing PAPRs

Medical PAPR: Versaflo TR-301N+

This PAPR is marketed as a medical PAPR. The waist unit (model TR-301N+; 3M Corp) contains a nonwoven fabric filter (model TR-3712N), pump, and battery (model TR-332). Purified air is pumped from the waist unit into the hood (model S-133L) through a flexible hose (model PSD-0225). The outline of this PAPR is as follows. Only air purified by a high-performance nonwoven filter is introduced into the hood by a pump. Since positive pressure is naturally maintained inside the hood, outside air is prevented from entering even if there is a gap between the hood seal and the face and the head. Thus, the air, which the wearer breathes, is only the air purified by the nonwoven filter. As for the face mask, since the inside of the mask becomes negative pressure during inhalation, if there is a gap between the face and the mask, the outside air will enter directly through the gap during the wearer’s inhalation [23,24]. Therefore, PAPR is structurally capable of shielding most of the aerosols present in the ambient air. This is supported by the fact that this PAPR has a high value of APF=1000. The flow rate can be selected in 2 stages, that is, high and low, and the specified flow rates are approximately 6.5 cubic feet per minute (180 L/min) and 7.2 cubic feet per minute (200 L/min), respectively [25].

PAPR Prototype With a Simple Structure

This PAPR is a prototype developed as a low-cost PAPR that has a simple structure similar to the abovementioned commercially available medical PAPR. The specifications of the air purification characteristics in this prototype are as follows [26].

1. Only air purified by a high-efficiency nonwoven filter is pumped into the hood.

2. Because positive pressure is naturally maintained inside the hood, outside air is prevented from entering even if there is a gap between the hood seal and the face.

3. The exhaust is natural exhaust from a thin nonwoven fabric filter due to the positive pressure inside the hood. Even if the wearer is infected and emits droplets or aerosols containing the virus, it is possible to prevent some of the external emissions.

The first 2 characteristics mentioned above are the same as those for the medical PAPR mentioned above (TR-301N+). In the air supply side, a nonwoven fabric filter—high-efficiency particulate air filter—which can filter 99.97% or more of aerosols down to 0.3 μm or larger is used on the air supply side. For aerosols containing viruses, it is considered sufficient to target aerosols with a particle size of 0.3 μm or larger [27,28]. A high-performance PAPR can be assembled at a total parts cost of approximately US \$40 with a simple configuration of only a high performance filter, battery, and pump.

PAPR Prototype With a Controller

This PAPR is a prototype developed as a high-performance PAPR equipped with a controller/computer for measurement and control [29]. The specifications for the air purification characteristics are the same as those in the PAPR prototype with a simple structure. In addition to the characteristics of the PAPR prototype with a simple structure, a controller (on-board computer) and sensors (2 CO₂ concentration sensors and 1 differential pressure sensor) are added, and the pump is controlled by means of the pulse width modulation control. In this PAPR, the pump output is adjusted and controlled according to the output of the differential pressure sensor so as to suppress the internal pressure fluctuations due to breathing, that is, higher pressure during exhalation and lower pressure during inspiration. Operating parameters can be set and monitored using a

smartphone. It is also possible to connect to the internet via the smartphone. As a result, it is possible to connect to the PAPR wearing rate network management system. In addition, by installing a pump and filter on the exhaust side and making it a differential type, it is possible to set the internal pressure to either positive pressure or negative pressure. Thus, it is possible to set the positive pressure setting to protect the wearer from the outside and to set the negative pressure setting to protect the outside from the wearer. It is possible to manufacture this PAPR with a total parts cost of approximately US \$300; this PAPR allows pump control based on sensor signals and settings and monitoring of the operating parameters by using a smartphone.

How to Operate a Social System Utilizing PAPR

A case in which a fixed percentage of the population wears PAPRs is considered. Quantitative evaluation of the reduction in the aforementioned β (infection probability per contact) by wearing PAPR and quantitative evaluation of the reduction in R_t by wearing PAPR are considered to require empirical and social experiments, as described below, because they involve humans. However, as shown below, quantitative evaluation is possible under limited conditions. Assuming that the aerosols and droplet shielding rate by PAPRs is the same for all particle sizes of aerosols and droplets, the shielding rate can be expressed as follows: (1) shielding rate for aerosols and droplets in the air supply ($R_{S,in}$) and (2) shielding rate for aerosols and droplets in the exhaust air ($R_{S,ex}$). Assume that the reduction rate of the probability of the wearer himself/herself becoming infected and the reduction rate of the probability of the wearer infecting others by wearing PAPR are as follows: (3) reduction rate of the probability that the wearer will be infected by aerosols and droplets in the air supply ($R_{I,in}$) and (4) reduction rate of the probability of infecting others by aerosols and droplets in exhaust air ($R_{I,ex}$). It is difficult to quantitatively determine the relationship between (1) and (3) and between (2) and (4) above. However, the following limited arrangement can be made. For the air supply side, the following can be said.

1. If the aerosol and droplet shielding ratio of PAPR is perfect ($R_{S,in}=1$), wearing PAPR will reduce the probability of infection by means of aerosols and droplets in the supply air by 100% ($R_{I,i}=1$).
2. If the aerosol and droplet shielding ratio of PAPR is nothing ($R_{S,in}=0$), wearing PAPR will reduce the probability of infection by means of aerosols and droplets in the supply air by nothing ($R_{I,in}=0$).
3. In the intervals of $0 < R_{S,in} < 1$ and $0 < R_{I,in} < 1$, there is a positive correlation between $R_{S,in}$ and $R_{I,in}$.

For the exhaust side, the following can be said.

4. If the aerosol and droplet shielding ratio of PAPR is perfect ($R_{S,ex}=1$), wearing PAPR will reduce the probability of infecting

others by means of the aerosols and droplets in the exhaust air by 100% ($R_{I,ex}=1$).

5. If the aerosol and droplet shielding ratio of PAPR is nothing ($R_{S,ex}=0$), wearing PAPR will reduce the probability of infecting others by means of the aerosols and droplets in the exhaust air by nothing ($R_{I,ex}=0$).
6. In the intervals of $0 < R_{S,ex} < 1$ and $0 < R_{I,ex} < 1$, there is a positive correlation between $R_{S,ex}$ and $R_{I,ex}$.

A social group was assumed to be completely free from contact and oral infections, and airborne and droplet infections were the only routes of infection. It is assumed that a certain percentage of the population in the social group always wears a PAPR. The performance of PAPR is assumed to be as follows.

1. The shielding rate of aerosols and droplets in the air supply side is 100% ($R_{S,in}=1$). As a result, the reduction rate of the probability of infection by aerosol and droplets in the air supply is 100% ($R_{I,in}=1$).
2. The shielding rate of aerosols and droplets in the exhaust side is the same as that of the face mask used in the population at that time.
3. When the above PAPR is worn by a percentage of people in W_R , the following relationship is established between the effective reproduction number R_t immediately before the start of wearing and the modified effective reproduction number R_{tm} immediately after the start of wearing.

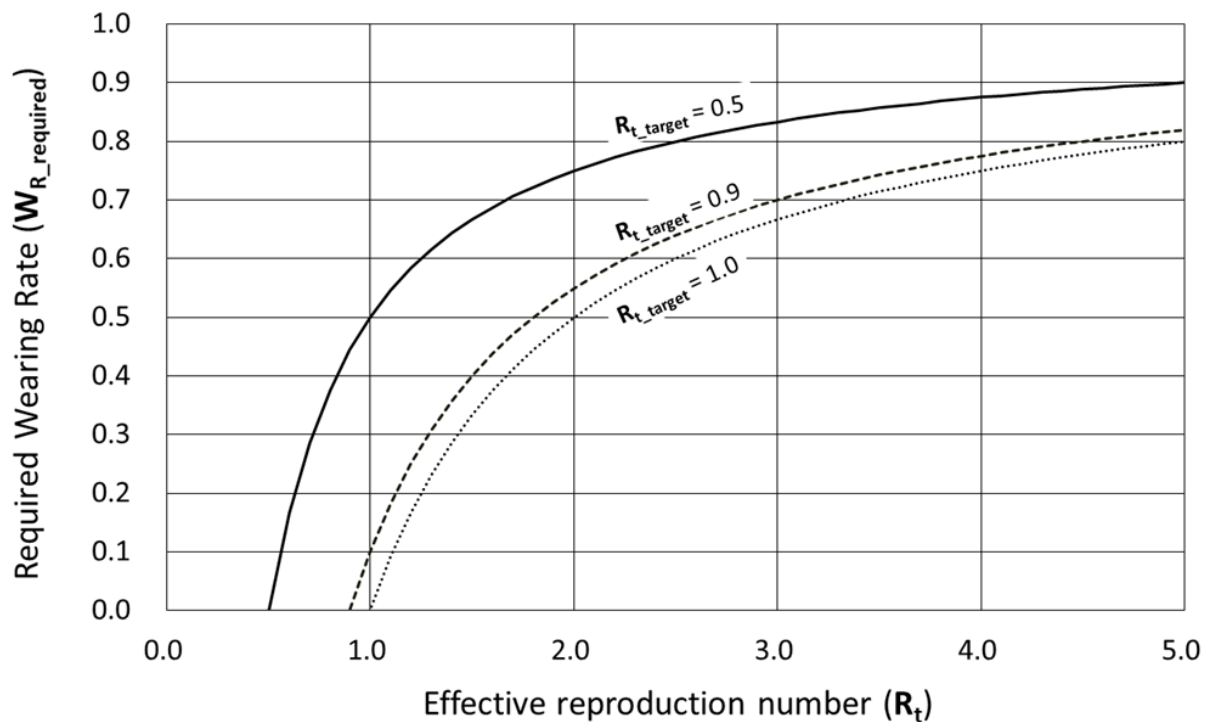
$$R_{tm} = [0.0 W_R + 1.0 (1 - W_R)] R_t = (1 - W_R) R_t$$

The expression for W_R is as follows.

$$W_R = 1 - R_{tm} / R_t$$

Figure 2 shows the relationship between the effective reproduction number R_t at the time in question and the required wearing rate $W_{R_required}$, which is required to achieve the target effective reproduction number R_{tm} of 0.5, 0.9 and 1.0. For example, consider an event in which a certain percentage of the population is wearing PAPR at all the time. As an example of a situation of severe infection spread, consider the case where $R_t=2$ immediately before the start of the event. In this case, to achieve the target effective reproduction number R_{t_target} of 1.0, 0.9, and 0.5, 50%, 55%, and 75% of the population should wear PAPRs at all times, respectively. The above simulation targets a social group in which airborne and droplet infections are the only routes of infection and an extreme setting in which PAPR is worn at all times. Future studies should consider more realistic settings that suit the conditions of daily life. For example, the situations in which the effects of not wearing PAPR should be considered, including contact with family members in the home, eating, drinking, washing, and bathing.

Figure 2. Relationship between the effective reproduction number and the required wearing rate, which is required to achieve the target effective reproduction number of 0.5, 0.9, and 1.0. R_{t_target} : target effective reproduction number.



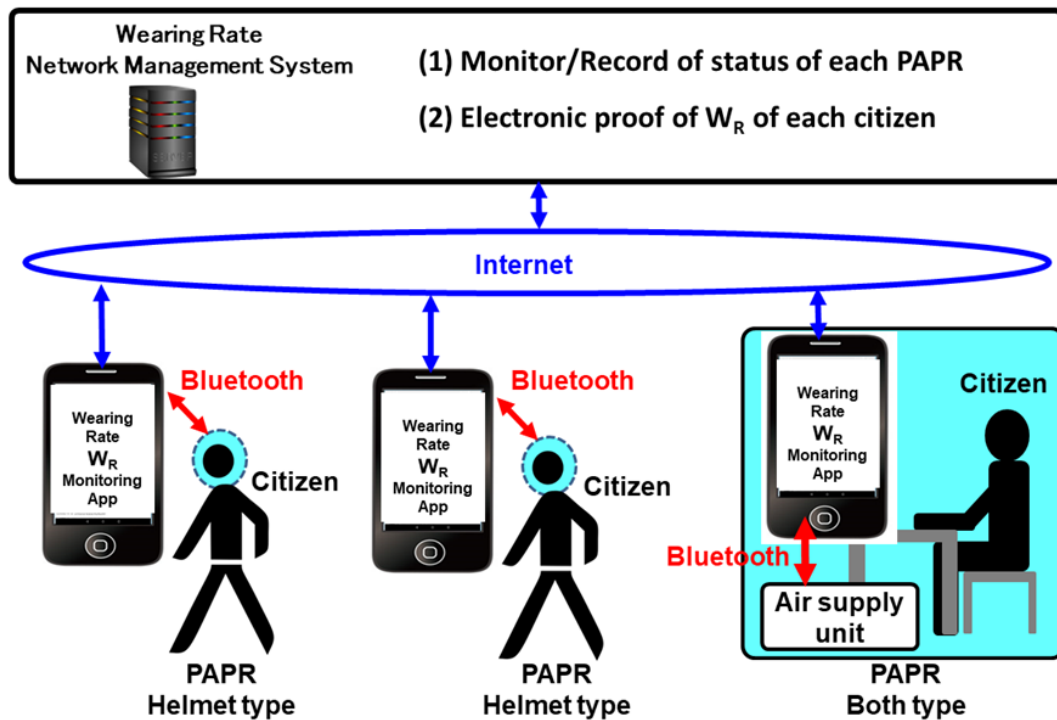
Direction of Improvement of PAPR as an Alternative to Lockdown

As the directions of improvements of PAPR as an alternative to lockdown, the following 7 points are proposed and discussed: (1) flow path optimization, (2) precise pressure control by fluid modeling, (3) improved design, (4) maintenance method, (5) variations suitable for different places of use and activity contents, (6) PAPR with information terminal function, and (7) evaluation indicators and evaluation methods.

Balance Between Efficiency of Infection Control and Personal Freedom Through the Use of IoT

In order to achieve both (1) effectiveness and efficiency in controlling the spread of infection and (2) individual freedom (limiting the obligation to wear PAPRs to the minimum necessary) at a high level in a social system that uses PAPRs as an alternative to lockdown, it is considered effective to develop a PAPR wearing rate network management system as shown in Figure 3.

Figure 3. The powered air purifying respirator wearing rate network management system. PAPR: powered air purifying respirator; W_R : wearing rate.



1. PAPR (helmet type, booth type, etc) is connected to the wearer's smartphone via Bluetooth.
2. The smartphone is connected to the internet and connected to the PAPR wearing rate network management system server operated by the government.
3. The government will be able to monitor, record, and manage each citizen's PAPR wearing rate along with smartphone location information by using the system.

Wearers (citizens) can display the electronic proof of their wearing rate on their smartphones provided by the system. Various parameters can be considered for the PAPR wearing rate (W_R). As simple examples, the following definitions of wearing rate (W_R) can be considered.

W_R =time spent outside with PAPR/time spent outside

Instead of the PAPR wearing rate W_R , time spent outside without PAPR T_{WT} can be considered.

T_{WT} =time spent outside without PAPR.

Instead of the PAPR wearing rate W_R , the number of viruses inhaled during an outing I_V (virions), which is considered to have a direct correlation with infection, could be used as a parameter for evaluation. If the estimated viral concentration d (virions/ m^3) in the activity range is available, the following definition can be adopted.

I_V =number of viruses inhaled during outings (virions)

$$= \int d (1-R_{S,in}) Q_{breath,in} dt$$

where, d (virions/ m^3)=estimated viral concentration at the location, $R_{S,in}$ =aerosol shielding ratio for the air supply of PAPR,

and $Q_{breath,in}$ (m^3/s)=estimated amount of inhaled air of the wearer at the time (exhaled air is not counted). As a request from the government to each citizen, it is assumed that keeping the above W_R , T_{WT} , or I_V at a certain level or better will be requested.

Ethical Considerations

This study is based on known facts and the author's own thinking, and no new experiments were performed. Therefore, the author has not applied to Gunma University, to which the author belongs, for ethics approval. Consent for publication has been granted from the identifiable individual (author YF) in Figure 1 in this paper.

Results

What PAPR Can Do as an Alternative to Lockdown

PAPR was shown to be a possible alternative to lockdown through the reduction of airborne and droplet transmissions and through the temporary reduction of β . The existing medical PAPRs appear to have sufficiently high virus shielding performance and appear to have already reached a level that should be experimentally tested as an alternative to lockdown. The current medical PAPR shown in Figure 1 is expensive and does not have measurement control functions. However, the prototypes shown in Figure 1 indicate that cost reduction and high functionality are possible. In addition, a variety of PAPRs are commercially available for nonmedical use, some of which are inexpensive. If an inexpensive PAPR is supplied to everyone, using PAPRs during a pandemic instead of issuing a countrywide lockdown will become a reality.

How to Operate a Social System Utilizing PAPR

A realistic process is shown below for quantitatively evaluating the effect of the aerosol shielding performance of PAPRs (for air intake side and exhaust side), PAPR wearing rate and wearing condition for reducing the β , and the effective reproduction number R_t for realizing this proposal.

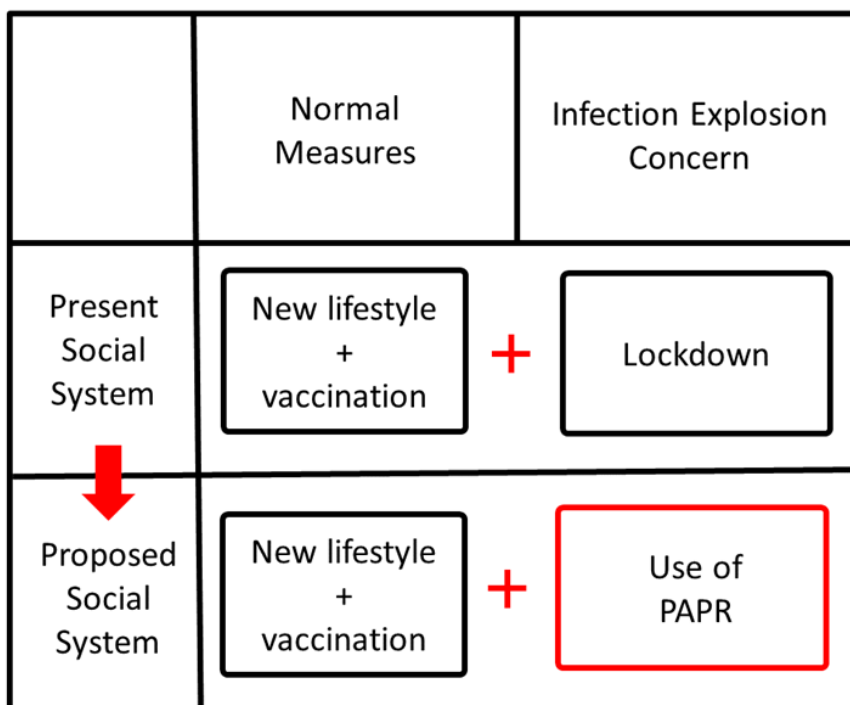
1. Select and prepare special experimental zones for social experiments in the next pandemic.
2. In the experimental zone, when lockdown is applied to the surrounding area, PAPR can be substituted for the various activity restrictions during lockdown.
3. Compare the spread of infection between the special experimental zone and other areas, and change the aerosol shielding performance of PAPR (air supply side and exhaust side), PAPR wearing rate and condition, and other operational

conditions within the special experimental zone. The obtained results can be used to quantitatively evaluate the effects of the aerosol shielding performance of PAPR (air supply side and exhaust side), PAPR wearing rate and wearing condition for reducing the β , and the effective reproduction number R_t .

4. When the effectiveness of PAPR as an alternative to lockdown is confirmed and the problems are sufficiently resolved, PAPR as an alternative to lockdown can be applied to other regions.

Examples of operations in special experimental zones include the following: (1) people can go out freely if they wear PAPRs, even in circumstances where going out is restricted in other surrounding areas and (2) factories can be operated freely if its employees wear PAPRs, even in circumstances where factories are prohibited to operate in other surrounding areas. Figure 4 shows the proposed social system where PAPRs are used as an alternative to lockdown.

Figure 4. Proposed social system where powered air purifying respirators are used as an alternative to lockdown. PAPR: powered air purifying respirator.



In the proposal in this research, PAPR will be utilized under the leadership of the government as described below:

1. The government distributes PAPRs (helmet/hood type) to all citizens as emergency equipment.
2. If the estimated effective reproduction number R_t is high and there is concern about an outbreak of infection, the government will determine the type and intensity of the lockdown and decide how to replace each constraint during lockdown with PAPR. Examples include prohibition on going outside (going outside is possible if wearing a PAPR), prohibition of factory operation (operation is possible if all employees wear PAPRs), prohibition of restaurant operation (operation is possible if all employees wear PAPRs and all customers wear PAPRs suitable for eating and drinking), and overseas entry prohibition (entry is allowed if visitors agree to wear a negative pressure PAPR for a specified

period of time and accept government remote monitoring of wearing conditions). This will make it possible to open the door to foreigners and returnees who wish to enter the country, although they would be subject to the same level of inconvenience as ordinary citizens suspected of being infected.

In the initial implementation of the proposed social system, as described above, special experimental zones will be established in various regions, various trials will be conducted based on various assumptions, and data will be collected. Based on the data obtained, qualitative and quantitative evaluations of the benefits (reduction of infection probability) and burdens borne by individuals and the benefits (reduction of infection spread) and burdens for the society as a whole will be attempted. The proposed social system should be compared and verified with the lockdown in each of the different situations, and the best way to be found as an alternative to the various restrictions

imposed by the lockdown should be identified. Ultimately, a PAPR-utilizing social system will be constructed that effectively functions as an alternative to lockdown.

Directions for Improvement of PAPRs as an Alternative to Lockdown

The following 7 points can be considered as directions for improvements of PAPRs as a lockdown alternative.

Flow Path Optimization

A hood shape and part configuration should be developed that provides a smooth flow of the exhaled air out of the hood. The concentration of carbon dioxide is approximately 500 ppm (0.05%) in ambient air and approximately 50,000 ppm (5%) in exhaled air [30]. The oxygen concentration in the exhaled air is expected to decrease from the oxygen concentration in the ambient air (approximately 21%) by an amount equal to the increase in the carbon dioxide concentration in the exhaled air (approximately 5%). In the commercially available PAPR and the developed PAPRs shown in Figure 1, a large flow rate (approximately 200-400 L/min) is delivered compared to the resting respiratory flow rate (approximately 6-10 L/min) [31] in order to suppress the carbon dioxide concentration in the hood [25,26,29]. Efficient expiration of exhaled air to the outside allows for a significant reduction in the air supply flow rate, resulting in a significant reduction in the size and weight of pumps, batteries, and filters, as well as design diversification. By minimizing the volume inside the mask, it is also possible to minimize the retention of the exhaled air from the nose and mouth inside the mask. As an extreme example, consider a configuration in which the nose is used for inhalation, the mouth is used for exhalation, and the air supply to the nose and the exhaust from the mouth are mechanically separated. As a result, the flow rate of the air filtered through the nonwoven filter and delivered to the nose becomes the same as the flow rate inhaled from the nose, and this dramatic reduction in flow rate results in a drastic reduction in the pump and battery capacity.

Precise Pressure Control by Fluid Modeling

Fluid modeling of PAPRs should be considered. For the PAPR prototypes (simple PAPR and controller PAPR) shown in Figure 1, the air supply through a nonwoven fabric filter is realized by a pump, and the exhaust through a nonwoven filter is created through the positive internal pressure. A simple modeling for these PAPRs is as follows.

Air Supply Flow Rate

The flow rate $Q_{in}(\Delta P, V)$ through the filter is determined by the pressure difference ΔP_f before and after the filter. The flow rate through the pump is determined by the pressure difference ΔP_p before and after the pump and the applied voltage (V) of the pump. When the differential pressure $\Delta P (= \Delta P_f + \Delta P_p)$ inside and outside the PAPR and the pump applied voltage (V) are determined, the air supply flow rate Q_{in} is determined.

Exhaust Flow Rate

The flow rate $Q_{out}(\Delta P)$ through the filter is determined by the pressure difference ΔP_f before and after the filter.

Respiratory Flow

The flow difference Q_{diff} between the air supply flow rate Q_{in} and the exhaust flow rate Q_{out} can be expressed as follows.

$$Q_{diff} = Q_{in}(\Delta P, V) - Q_{out} = Q_{breath} + Q_{leak} + Q_{volume}$$

Here, Q_{breath} =respiratory flow rate of the wearer of PAPR, positive with inspiration; Q_{leak} =leak flow rate, positive for leakage from the inside to the outside; and Q_{volume} =volume change inside PAPR, positive with volume increase.

Since the time averages of respiratory flow Q_{breath} and volume change Q_{volume} are zero, the time average of Q_{diff} is the time average of Q_{leak} . In addition, Q_{leak} is expressed as a function $Q_{leak}(\Delta P)$ of the differential pressure ΔP , assuming that the shape of the gap between the face and the mask is constant. Furthermore, the volume change Q_{volume} is considered to be expressed as a function $Q_{volume}(\Delta P)$ of the differential pressure ΔP .

$$Q_{breath} = Q_{diff} - Q_{leak}(\Delta P) - Q_{volume}(\Delta P)$$

In this case, $Q_{breath} < 0$ is judged as expiration, and $Q_{breath} > 0$ is judged as inspiration. In this way, the exhalation and inhalation movements of the wearer can be detected in real time. For example, based on this detection result, the following control can be considered.

1. If an exhalation movement is detected, the minimum positive pressure setting (eg, 10 Pa) is set to minimize the resistance to exhalation movement while preventing leakage from the gap.
2. If an inhalation movement is detected, a strong positive pressure setting (eg, 100 Pa) is used to positively assist the inhalation movement.

In addition to the forced air supply by the pump and filter, the introduction of forced exhaust by the pump and filter enables the following differential pressure control.

1. When an exhalation movement is detected, a strong negative pressure setting (eg, -100 Pa) is used to actively assist the exhalation movement.
2. When an inhalation movement is detected, a strong positive pressure setting (eg, 100 Pa) is used to actively assist the inhalation movement.

In light of this, a PAPR facilitates the easy movement of the wearer's exhalation and inhalation. In this case, the direction of leak flow at the possible gap is opposite to normal—from outside to inside during exhalation and from inside to outside during inhalation. It will be possible to detect coughing from the measurement results of the differential pressure ΔP . It will also be possible to estimate the possibility of infection of the wearer together with other measurement results such as body temperature. The ability to efficiently identify infected persons will enable efficient isolation and treatment of infected persons and will have a significant effect in reducing the spread of infection throughout the society.

Improved Design

All the 3 types of PAPRs shown in [Figure 1](#) have bulky and exaggerated designs. As a lockdown alternative, the design may not be very important; however, it is better to have an excellent design. If the above flow path optimization achieves a dramatic reduction in the air supply flow rate, then a dramatic reduction in pump and battery size and various designs will become possible. Once PAPRs are widely accepted as a lockdown alternative, many people will be dissatisfied with the bare-bones PAPRs provided by the government; it is conceivable that companies of various genres will develop models with different characteristics.

Maintenance Method

It is necessary for every citizen to be able to easily perform maintenance such as cleaning and disinfecting the PAPR and replacing the nonwoven filter unit at home. However, when PAPRs are used as an alternative to lockdown, it is expected that the virus concentration in the external environment will be extremely low compared to the environment assumed in medical PAPRs due to the following reasons.

1. Infected persons would wear PAPRs.
2. PAPR has the ability to not only stop the entry of droplets and aerosols containing viruses but also prevent their release to the outside.
3. PAPR purifies indoor air in the same way as an air purifier.

Therefore, in terms of maintenance standards, it may be possible to set relatively lenient standards for nonmedical use PAPRs compared to those set for medical PAPRs, which are assumed to be used in environments with high virus concentrations such as hospital wards where infected people are congregated.

Variations Suitable for Different Places of Use and Activity Contents

The following variations should be developed, which are fine-tuned to suit various places of use and activities, as well as to suit societies and populations at different stages of social, economic, and cultural development: (1) a model that pursues comfort for everyday use; (2) models suitable for specific activities such as sports, eating, and drinking, for example, a model for eating and drinking with a face shield opening and closing mechanism and an air shower function or a model for jogging with a structure that mechanically separates the nose (exhalation) and mouth (inhalation); (3) a booth type model that wraps around a desk and a chair in an office, vehicle, restaurant, etc; (4) a model compatible with the standard unit of the ceiling-mounted air conditioner; and (5) a very inexpensive model suitable for low-income countries and regions.

PAPR With Information Terminal Function

The PAPR prototype (controller type) shown in [Figure 1](#) is an all-in-one type PAPR with a computer and a power supply on the wearer's head. Therefore, it is easy to make the PAPR an advanced information terminal by means of installing a computer equivalent to that of a high-end smartphone and adding various devices as follows: (1) equipped with smartphone function and virtual reality screen, (2) equipped with a

noncontact input system using eye gaze and brain waves, and (3) equipped with a physical condition measurement and management system using body temperature sensors, cough sensors (pressure sensors), electroencephalogram sensors, etc. If the PAPR is comfortable to breathe, comfortable to wear, and has advanced information terminal functions, it is expected that some people will not be able to part with it. In particular, if a physical condition measurement and management system is installed to accurately estimate the presence or absence of infection, it will be easier to isolate, examine, and treat those who are deemed to have a high probability of being infected. In many cases, the various behavioral and activity restrictions during lockdown are uniformly applied to all persons under conditions where it is not known who is infected. If it becomes possible to know with a high degree of certainty who is infected, the use of PAPR as an alternative to lockdown can be changed to a more targeted approach.

Evaluation Indicators and Evaluation Methods

As an evaluation indicator of PAPR as a lockdown alternative, it is desirable to be able to quantitatively evaluate the effect of reducing the aforementioned β by means of wearing a PAPR. However, in order to estimate the β reduction rate with high accuracy, it is necessary to conduct elemental experiments and social experiments under various conditions.

The most important evaluation indicators of the basic performance of PAPR that should be obtained from elementary experiments are as follows: (1) reduction rate of virus-containing aerosols and droplets inhaled by potentially infectious persons wearing PAPR and (2) reduction rate of virus-containing aerosols and droplets exhaled by infected persons wearing PAPR. Current standards (eg, APF) usually refer only to the reduction rate of virus-containing aerosols and droplets inhaled by potentially infectious persons wearing PAPR. However, both (1) and (2) are considered to be equally important when requiring the uniform wearing of PAPRs by the general public in cases where presence or absence of infection is unclear for the purpose of reducing the effective reproduction number R_t . In the 3 PAPR models shown in [Figure 1](#), positive pressure was used to prevent outside air from entering directly through gaps, with an emphasis on protecting the inside (wearer) from the outside. If the above (1) and (2) are equally important, then it is equally important to protect the wearer from the outside environment and to protect the people from the wearer, thereby indicating that it is not necessary to make the internal pressure positive.

Impact of the PAPR Internal Environment on the Mind and Body

If PAPR is considered as an alternative to lockdown measures, the impact of the PAPR internal environment on the mind and body of the wearer will become important. It is necessary to comprehensively investigate the relationship between the following 2 types of parameters from the viewpoint of the influence of the PAPR internal environment on the mind and body of the wearer.

Physical Parameters Related to the PAPR Internal Environment

The physical parameters to be considered are concentrations of particulate pollutants (droplets, aerosols, pollen, particulate matter 2.5, mite corpses, dust, etc), gaseous pollutants, gas composition (carbon dioxide concentration, oxygen concentration, etc), differential pressure, temperature, humidity, acoustic characteristics (sound transfer characteristics, noise, etc), vibration, and airflow.

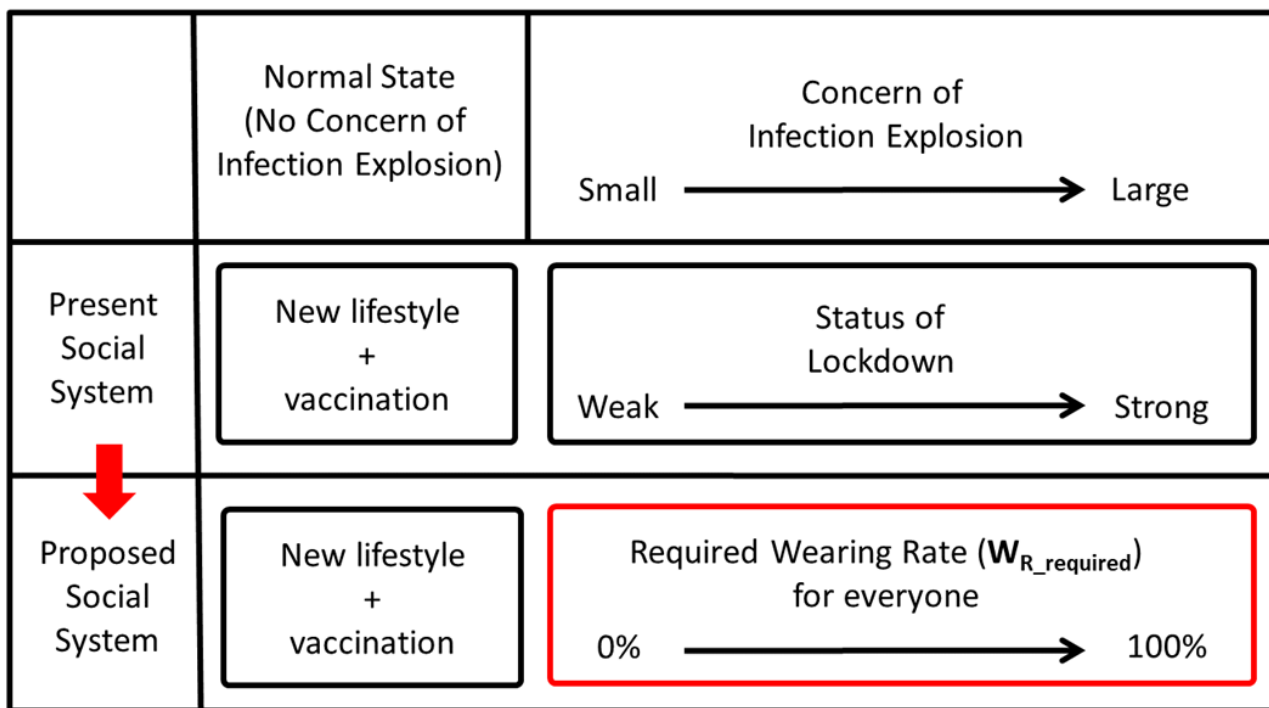
Biological and Psychological Parameters of PAPR Wearers

The biological and psychological parameters are respiratory status, electroencephalogram, body temperature, pulse rate, comfort, safety (physical danger, probability of infection), and degree of relaxation.

Balance Between Infection Control Efficiency and Personal Freedom Through the Use of IoT

As shown in Figure 5, the operation of the PAPR wearing rate network management system led by the government will be performed as follows.

Figure 5. The social system with the powered air purifying respirator wearing rate network management system.



1. The government will distribute PAPRs (with smartphone connectivity) with sufficient aerosol shielding performance to all citizens as emergency equipment. Each citizen installs an app with a wearing rate proof function on his/her own smartphone.

2. If the estimated effective reproduction number R_t is high and there is concern about an outbreak of infection, the government will (1) set the target effective reproduction number R_{t_target} , (2) solve the formula based on appropriate assumptions to calculate the required wearing rate $W_{R_required}$ required to achieve R_{t_target} , (3) show $W_{R_required}$ and require all citizens to comply with it—each citizen can spend their time without PAPR at any place (party venue, restaurant, pub, etc) and any time by showing proof of wearing rate W_R within the scope of fulfilling their obligations, and (4) pay close attention to changes in the effective reproduction number R_t and raise $W_{R_required}$ if the goal of controlling the spread of infection is in jeopardy. Conversely, if it exceeds the target, lower $W_{R_required}$ and increase the degree of freedom in citizen life.

If the PAPR lockdown alternatives are strong enough, the government can quickly contain the spread of infection by setting $W_{R_required}$ to 1.0 (100%) even when the government makes a big mistake in estimating $W_{R_required}$ and falls into the worst situation. In that case, the government can conduct various trials and countermeasures on various assumptions and hypotheses with a leeway. The government is freed from constraints that limit them to overly conservative measures. In addition, throughout the entire process, the government will be able to improve the accuracy of the above formula based on appropriate assumptions by using the big data collected on the relationships between “changes in aerosol shielding performance of PAPR (air intake side, exhaust side), PAPR wearing rate W_R , and wearing condition, etc” and “changes in infection spread status and the effective reproduction number R_t .”

Discussion

Principal Findings

In this study, the feasibility of the following 2 ideas was examined. First, the construction of a social system using PAPR

with similar infection control ability as lockdown measures and with less economic and social damage as an alternative to lockdown is possible. Second, balancing the efficiency of the government's infection control and each citizen's personal freedom is possible by means of an IoT system.

Extended Functionality and Privacy Protection in the PAPR Wearing Rate Network Management System

By utilizing PAPR with several sensors (thermometer, cough sensor, etc), the government can make this system much more powerful than conventional apps for measuring contact with infected persons. For example, the system may be able to improve the accuracy of infection detection based on big data concerning changes in body temperature, cough (condition and frequency), and the presence or absence of severe disease after infection until the onset of illness. In the case of PAPR equipped with both air supply and exhaust pumps, the wearer can switch the internal pressure between positive pressure when not infected and negative pressure when infected, thus prioritizing the prevention of the spread of infection in the society as a whole. From the viewpoint of privacy protection, social discussion is necessary for the following matters: (1) how much of the information from PAPRs should be passed on to the government server? and (2) how should the government's use of personal information be curbed? Especially for (2), it is considered necessary to develop and construct a technical and social mechanism to realize a brake. In order to prevent misuse of personal information, it is conceivable to apply the proposal for a street camera system's perfect recording of usage history by a reliable third party as the first step [30]. Methods to substitute each restriction in lockdown with PAPR utilization need to be considered in various social systems. Those who wish to use PAPR as a substitute for the constraints imposed by lockdown need to prepare to obtain the PAPR before the pandemic. Therefore, it is also important for the society as a whole to ensure and disseminate information on how to obtain PAPRs.

Different countries have different governance systems. In some countries, it might not be easy to make the public understand that PAPR can be used an alternative to lockdown measures—they may make it an option and not a mandate. This paper discusses how PAPR can substitute the primary constraints imposed by lockdown. Even in cases of other alternatives such as combination of lockdown and free mobility of low-risk populations during COVID-19 [32], PAPRs may be used for

controlling the infection rates. The proposed PAPR wearing rate network management system utilizes IoT technology, which is currently being widely pursued by various societies and companies. In order to build a social system that makes the government's control of the spread of infection more efficient and that respects the freedom of individuals to the maximum extent, social experiments should be first conducted under various conditions to identify the challenges and improve the effectiveness of PAPRs.

Society of People Breathing Purified Air

Further, although this is a discussion that is far from the main point of this paper, if truly high-performance, comfortable, and low-cost PAPRs are successfully developed through this research and subsequent research and developments, it is possible that many people will desire PAPR-purified air instead of the air around them. This is similar to the situation of drinking water, that is, just as how populations consume purified water through water-treatment and water-purification technology rather than water from ponds and rivers, people would prefer purified air to breathe. It can be expected that many citizens will wear PAPRs when they go out, regardless of whether the government asks them to do so. As more people breathe purified air, there may be concerns about the public's immune system being weakened against airborne diseases and pollen allergies. However, there is no dispute that water-borne infectious diseases have become controllable because many people drink only purified water, and it would not be advisable to drink water without purification, as was the case in primitive times. A society in which the majority of the population breathes purified air will be resilient to all airborne diseases. The construction of such a society has the potential to be an opportunity for a historic change in the human race, which has been plagued by airborne diseases.

Conclusions

This study examines the feasibility of 2 ideas. First, this study shows that it is possible to construct a social system using PAPR with similar infection control effects as lockdown measures and with less economic and social damage as a means of temporarily reducing the effective reproduction number R_t . Second, the PAPR wearing rate network management system balances the achievement of the efficiency of the government's infection control and each citizen's personal right to choose the time and opportunity not to wear PAPR during a pandemic.

Acknowledgments

The author would like to thank Prof Naoya Ohta, Prof Noriaki Yoshiura, Prof Akihiro Takita, Prof Seiji Hashimoto, Prof Takao Yamaguchi, Prof Kenji Amagai, Prof Haruo Kobayashi, Prof Shu Dong Wei, Prof Osamu Takaki, Prof Edwin Carcasona, and Prof Ronald Galindo for their fruitful discussions. This research was funded by Japan Society for the Promotion of Science, Promotion of Joint International Research, KAKENHI (grants-in-aid for Scientific Research for FY2021, Issue 21KK0080).

Data Availability

This study was conducted using deductive thinking based on publicly available information. No confidential data were used in this study.

Authors' Contributions

YF is solely responsible for all matters related to this study.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19. WHO. URL: <https://www.who.int/news-room/questions-and-answers/item/herd-immunity-lockdowns-and-covid-19> [accessed 2024-04-28]
2. Barro RJ. Vaccination rates and COVID outcomes across U.S. states. *Econ Hum Biol* 2022 Dec;47:101201 [FREE Full text] [doi: [10.1016/j.ehb.2022.101201](https://doi.org/10.1016/j.ehb.2022.101201)] [Medline: [36434953](https://pubmed.ncbi.nlm.nih.gov/36434953/)]
3. Coccia M. The relation between length of lockdown, numbers of infected people and deaths of Covid-19, and economic growth of countries: Lessons learned to cope with future pandemics similar to COVID-19 and to constrain the deterioration of economic system. *Science of The Total Environment* 2021 Jun;775:145801 [FREE Full text] [doi: [10.1016/j.scitotenv.2021.145801](https://doi.org/10.1016/j.scitotenv.2021.145801)]
4. Zhang H, Li P, Zhang Z, Li W, Chen J, Song X, et al. Epidemic versus economic performances of the COVID-19 lockdown: A big data driven analysis. *Cities* 2022 Jan;120:103502 [FREE Full text] [doi: [10.1016/j.cities.2021.103502](https://doi.org/10.1016/j.cities.2021.103502)] [Medline: [34703071](https://pubmed.ncbi.nlm.nih.gov/34703071/)]
5. Wu J, Zhan X, Xu H, Ma C. The economic impacts of COVID-19 and city lockdown: Early evidence from China. *Struct Chang Econ Dyn* 2023 Jun;65:151-165 [FREE Full text] [doi: [10.1016/j.strueco.2023.02.018](https://doi.org/10.1016/j.strueco.2023.02.018)] [Medline: [36876039](https://pubmed.ncbi.nlm.nih.gov/36876039/)]
6. Romanyukha AA, Novikov KA, Avilov KK, Nestik TA, Sannikova TE. The trade-off between COVID-19 and mental diseases burden during a lockdown: Mathematical modeling of control measures. *Infect Dis Model* 2023 Jun;8(2):403-414 [FREE Full text] [doi: [10.1016/j.idm.2023.04.003](https://doi.org/10.1016/j.idm.2023.04.003)] [Medline: [37064013](https://pubmed.ncbi.nlm.nih.gov/37064013/)]
7. Caro JC, Clark AE, D'Ambrosio C, Vögele C. The impact of COVID-19 lockdown stringency on loneliness in five European countries. *Soc Sci Med* 2022 Dec;314:115492 [FREE Full text] [doi: [10.1016/j.socscimed.2022.115492](https://doi.org/10.1016/j.socscimed.2022.115492)] [Medline: [36343461](https://pubmed.ncbi.nlm.nih.gov/36343461/)]
8. Oraby T, Tyshenko MG, Maldonado JC, Vatcheva K, Elsaadany S, Alali WQ, et al. Modeling the effect of lockdown timing as a COVID-19 control measure in countries with differing social contacts. *Sci Rep* 2021 Feb 08;11(1):3354 [FREE Full text] [doi: [10.1038/s41598-021-82873-2](https://doi.org/10.1038/s41598-021-82873-2)] [Medline: [33558571](https://pubmed.ncbi.nlm.nih.gov/33558571/)]
9. Lewis D. What scientists have learnt from COVID lockdowns. *Nature* 2022 Sep;609(7926):236-239. [doi: [10.1038/d41586-022-02823-4](https://doi.org/10.1038/d41586-022-02823-4)] [Medline: [36071184](https://pubmed.ncbi.nlm.nih.gov/36071184/)]
10. Hodson R. Preparing the world for the next pandemic. *Nature* 2022 Oct;610(7933):S33. [doi: [10.1038/d41586-022-03353-9](https://doi.org/10.1038/d41586-022-03353-9)] [Medline: [36289375](https://pubmed.ncbi.nlm.nih.gov/36289375/)]
11. Jones S. How to eradicate the next pandemic disease. *Nature* 2022 Oct;610(7933):S48-S49. [doi: [10.1038/d41586-022-03361-9](https://doi.org/10.1038/d41586-022-03361-9)] [Medline: [36289379](https://pubmed.ncbi.nlm.nih.gov/36289379/)]
12. Lewis D. COVID-19 rarely spreads through surfaces. So why are we still deep cleaning? *Nature* 2021 Feb;590(7844):26-28. [doi: [10.1038/d41586-021-00251-4](https://doi.org/10.1038/d41586-021-00251-4)] [Medline: [33514939](https://pubmed.ncbi.nlm.nih.gov/33514939/)]
13. Prather K, Marr LC, Schooley RT, McDiarmid MA, Wilson ME, Milton DK. Airborne transmission of SARS-CoV-2. *Science* 2020 Oct 16;370(6514):303-304. [doi: [10.1126/science.abf0521](https://doi.org/10.1126/science.abf0521)] [Medline: [33020250](https://pubmed.ncbi.nlm.nih.gov/33020250/)]
14. Karimzadeh S, Bhopal R, Nguyen Tien H. Review of infective dose, routes of transmission and outcome of COVID-19 caused by the SARS-COV-2: comparison with other respiratory viruses– CORRIGENDUM. *Epidemiol. Infect* 2021 May 14;149:1-8 [FREE Full text] [doi: [10.1017/s0950268821001084](https://doi.org/10.1017/s0950268821001084)]
15. Prentiss M, Chu A, Berggren KK. Finding the infectious dose for COVID-19 by applying an airborne-transmission model to superspreader events. *PLoS One* 2022;17(6):e0265816 [FREE Full text] [doi: [10.1371/journal.pone.0265816](https://doi.org/10.1371/journal.pone.0265816)] [Medline: [35679278](https://pubmed.ncbi.nlm.nih.gov/35679278/)]
16. Mattiuzzi C, Lippi G, Nocini R. Highly efficient respirators are needed for the Omicron variant of SARS-CoV-2. *Public Health* 2022 May;206:e2 [FREE Full text] [doi: [10.1016/j.puhe.2022.03.003](https://doi.org/10.1016/j.puhe.2022.03.003)] [Medline: [35450730](https://pubmed.ncbi.nlm.nih.gov/35450730/)]
17. Considerations for optimizing the supply of powered air-purifying respirators (PAPRs): for healthcare practitioners (HCP). CDC 2020. URL: <https://stacks.cdc.gov/view/cdc/96858> [accessed 2024-04-28]
18. Assigned protection factors: for the revised respiratory protection standard. OSHA 2009. URL: <https://www.osha.gov/sites/default/files/publications/3352-APF-respirators.pdf> [accessed 2024-04-28]
19. Assigned protection factors (APF) for 3M hoods and helmets. 3M 2021. URL: <https://tinyurl.com/4m33pn4h> [accessed 2024-04-28]
20. HealthKnowledge 2018. URL: <https://tinyurl.com/55vk8jsy> [accessed 2024-04-28]
21. Kevin L, Mathias P, Ellen K. The reproduction number of COVID-19 and its correlation with public health interventions. *Computational Mechanics* 2020;66:1035-1050 [FREE Full text] [doi: [10.1007/s00466-020-01880-8](https://doi.org/10.1007/s00466-020-01880-8)]
22. Inglesby T. Public health measures and the reproduction number of SARS-CoV-2. *JAMA* 2020 Jun 02;323(21):2186-2187. [doi: [10.1001/jama.2020.7878](https://doi.org/10.1001/jama.2020.7878)] [Medline: [32356869](https://pubmed.ncbi.nlm.nih.gov/32356869/)]

23. Verma S, Dhanak M, Frankenfield J. Visualizing the effectiveness of face masks in obstructing respiratory jets. *Phys Fluids* (1994) 2020 Jun 01;32(6):061708 [FREE Full text] [doi: [10.1063/5.0016018](https://doi.org/10.1063/5.0016018)] [Medline: [32624649](https://pubmed.ncbi.nlm.nih.gov/32624649/)]
24. Ju JTT, Boisvert LN, Zuo YY. Face masks against COVID-19: Standards, efficacy, testing and decontamination methods. *Adv Colloid Interface Sci* 2021 Jun;292:102435 [FREE Full text] [doi: [10.1016/j.cis.2021.102435](https://doi.org/10.1016/j.cis.2021.102435)] [Medline: [33971389](https://pubmed.ncbi.nlm.nih.gov/33971389/)]
25. 3M replacement blower TR-301N+. 3M (2023). URL: https://www.3m.com/3M/en_US/p/d/v100559001/ [accessed 2024-04-28]
26. Carcasona E, Galindo RM, Takita A, Magalang E, et al. Very-low-cost powered air-purifying respirator (PAPR) “distancing-free mask industry (DFM-I) Prototype No.1” and proposal for a lockdown-free industry. *Journal of Technology and Social Science*. 2022. URL: https://jtss.e-jikei.org/issue/archives/v06n02/JTSS_v06n02a001.pdf [accessed 2024-04-30]
27. Ueki H, Ujie M, Komori Y, Kato T, Imai M, Kawaoka Y. Effectiveness of HEPA filters at removing infectious SARS-CoV-2 from the air. *mSphere* 2022 Aug 31;7(4):e0008622 [FREE Full text] [doi: [10.1128/msphere.00086-22](https://doi.org/10.1128/msphere.00086-22)] [Medline: [35947419](https://pubmed.ncbi.nlm.nih.gov/35947419/)]
28. Byung Uk L. Minimum sizes of respiratory particles carrying SARS-CoV-2 and the possibility of aerosol generation. *Int J Environ Res Public Health* 2020;17(19):6960 [FREE Full text] [doi: [10.3390/ijerph17196960](https://doi.org/10.3390/ijerph17196960)]
29. Galindo RM, Takita A, Carcasona E, Magalang E, et al. Low-cost powered air-purifying respirator (PAPR) “distancing-free mask frontline (DFM-F) prototype no. 1” for the operational tests in hospitals in Cebu City, Philippines. *Journal of Mechanical and Electrical Intelligent System*. 2022. URL: http://jmeis.e-jikei.org/ARCHIVES/v05n02/JMEIS_v05n02a001.pdf [accessed 2024-04-29]
30. Fujii Y. Will every streetlight have network cameras in the near future? *Science*. 2016 Oct 21. URL: http://www.e-jikei.org/Conf/ICTSI2018/proceedings/materials/proc_files/KeyNote/KL-03/ICTSI2018_KL_Fujii_ejikei.pdf [accessed 2024-04-28]
31. Joachim D P, M Ariel Geer W, Michael D D, Christopher M M. The physics of human breathing: flow, timing, volume, and pressure parameters for normal, on-demand, and ventilator respiration. *Journal of Breath Research* 2021;15(4):042002 [FREE Full text] [doi: [10.1088/1752-7163/ac2589](https://doi.org/10.1088/1752-7163/ac2589)]
32. Yao L, Aleya L, Goldman E, Graff JC, Gu W. An alternative approach-combination of lockdown and open in fighting COVID-19 pandemics. *Environ Sci Pollut Res Int* 2022 Nov;29(54):82611-82614 [FREE Full text] [doi: [10.1007/s11356-022-23438-2](https://doi.org/10.1007/s11356-022-23438-2)] [Medline: [36229730](https://pubmed.ncbi.nlm.nih.gov/36229730/)]

Abbreviations

APF: assigned protection factor

IoT: Internet of Things

PAPR: powered air purifying respirator

Edited by T Leung; submitted 17.11.23; peer-reviewed by D Shu, W Gu; comments to author 12.03.24; revised version received 14.03.24; accepted 09.04.24; published 14.05.24.

Please cite as:

Fujii Y

An Engineering Alternative to Lockdown During COVID-19 and Other Airborne Infectious Disease Pandemics: Feasibility Study

JMIR Biomed Eng 2024;9:e54666

URL: <https://biomedeng.jmir.org/2024/1/e54666>

doi: [10.2196/54666](https://doi.org/10.2196/54666)

PMID: [38875692](https://pubmed.ncbi.nlm.nih.gov/38875692/)

©Yusaku Fujii. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 14.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Stroke Survivors' Interaction With Hand Rehabilitation Devices: Observational Study

Chioma Obinuchi Wodu^{1,2}, BTECH, MSc; Gillian Sweeney¹, PhD; Milena Slachetka¹, BSc; Andrew Kerr¹, PhD

¹Department of Biomedical Engineering, University of Strathclyde, Glasgow, United Kingdom

²Department of Biomedical Technology, University of Port Harcourt, Port Harcourt, Nigeria

Corresponding Author:

Chioma Obinuchi Wodu, BTECH, MSc

Department of Biomedical Engineering

University of Strathclyde

106 Rottenrow

G4 0NW

Glasgow,

United Kingdom

Phone: 44 79 3058 4076

Email: chiomawodu@gmail.com

Abstract

Background: The hand is crucial for carrying out activities of daily living as well as social interaction. Functional use of the upper limb is affected in up to 55% to 75% of stroke survivors 3 to 6 months after stroke. Rehabilitation can help restore function, and several rehabilitation devices have been designed to improve hand function. However, access to these devices is compromised in people with more severe loss of function.

Objective: In this study, we aimed to observe stroke survivors with poor hand function interacting with a range of commonly used hand rehabilitation devices.

Methods: Participants were engaged in an 8-week rehabilitation intervention at a technology-enriched rehabilitation gym. The participants spent 50-60 minutes of the 2-hour session in the upper limb section at least twice a week. Each participant communicated their rehabilitation goals, and an Action Research Arm Test (ARAT) was used to measure and categorize hand function as poor (scores of 0-9), moderate (scores of 10-56), or good (score of 57). Participants were observed during their interactions with 3 hand-based rehabilitation devices that focused on hand rehabilitation: the GripAble, NeuroBall, and Semi-Circular Peg Board. Observations of device interactions were recorded for each session.

Results: A total of 29 participants were included in this study, of whom 10 (34%) had poor hand function, 17 (59%) had moderate hand function, and 2 (7%) had good hand function. There were no differences in the age and years after stroke among participants with poor hand function and those with moderate ($P=.06$ and $P=.09$, respectively) and good ($P=.37$ and $P=.99$, respectively) hand function. Regarding the ability of the 10 participants with poor hand function to interact with the 3 hand-based rehabilitation devices, 2 (20%) participants with an ARAT score greater than 0 were able to interact with the devices, whereas the other 8 (80%) who had an ARAT score of 0 could not. Their inability to interact with these devices was clinically examined, and the reason was determined to be a result of either the presence of (1) muscle tone or stiffness or (2) muscle weakness.

Conclusions: Not all stroke survivors with impairments in their hands can make use of currently available rehabilitation technologies. Those with an ARAT score of 0 cannot actively interact with hand rehabilitation devices, as they cannot carry out the hand movement necessary for such interaction. The design of devices for hand rehabilitation should consider the accessibility needs of those with poor hand function.

(*JMIR Biomed Eng* 2024;9:e54159) doi:[10.2196/54159](https://doi.org/10.2196/54159)

KEYWORDS

stroke; rehabilitation; hand rehabilitation devices; accessibility; stroke survivors; rehabilitation technologies

Introduction

Stroke is a major cause of disability in the world [1]. Globally, about 17 million people have a stroke each year [2]. In the United Kingdom, the prevalence of stroke is projected to rise from 950,200 to 2,119,400 cases between 2015 and 2035 [3]. This projected rise in the prevalence of stroke has been associated with improvements in medical advances that have led to a decline in the number of deaths due to acute stroke, among other reasons [4]. Nevertheless, stroke survivors are faced with considerable long-term periods of enduring physical impairments, the likelihood of reoccurrence of strokes, transient ischemic attacks, or even death within 1 year of having a stroke [5]. Motor impairment (muscle weakness and the loss of movement control) is the most common consequence of stroke, impacting several aspects of life and reducing the ability of stroke survivors to lead an independent life [6]. About 55% to 75% of those who survive a stroke experience motor impairment in the upper limb 3 to 6 months after stroke [7].

The hand is crucial for carrying out activities of daily living such as eating, dressing, bathing, and communicating [8]. Besides, the hand is a defining feature of human beings and is vital for human daily interaction [9]. Due to this importance, impairments such as spasticity and weakness, which are common sequelae of stroke [10] and manifest in a fixed flexed position of the wrist and fingers, affect the function of the hand and impact the quality of life [10].

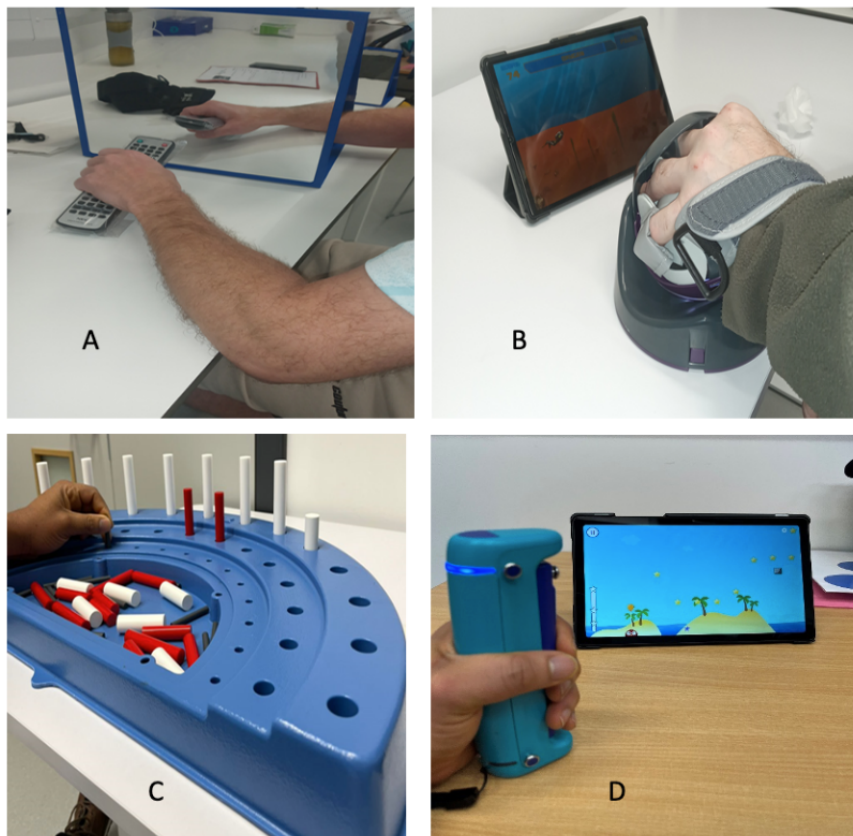
Rehabilitation can have a positive impact on the recovery of functions in persons with stroke [11] as well as in enhancing their quality of life [12], and movement restoration is a key goal in the rehabilitation of persons with neurological disorders [13].

The relearning of movement ability during rehabilitation is based on factors such as the repetitiveness, intensity, and regularity of task-specific movements [14]. It has been suggested that the rehabilitation of hand mobility and strength be prioritized once the general physical situation of stroke survivors has been stabilized owing to the importance of the hand [15].

Several new rehabilitation technologies that target the upper limb to improve motor functions are currently in use; these include the use of robotic-assisted technologies, virtual reality, and telerehabilitation [16]. Some others that are used in this study are gaming devices such as the GripAble (Gripable), NeuroBall (Neurofenix), and Semi-Circular Peg Board (Rolyan). The NeuroBall is an interactive device that connects wirelessly with a tablet app to carry out activities that can also be objectively measured [17]. The GripAble is a similar lightweight electronic handgrip [18] that also interacts wirelessly with a computer tablet, enabling users to interact with therapy games tailored to improve the upper limb and hand function in a way that can be objectively assessed [18,19]. The Rolyan Semi-Circular Peg Board consists of 3 colored pegs (red, white, and blue) of different diameters that the users are expected to pick up and place in their different peg holes (based on their diameter; see [Figure 1](#) below). The ability of stroke survivors with poor hand function to access these devices is a major concern, as according to a report [20], only hemiplegic stroke survivors who are mildly disabled are likely to access hand or arm training apps that are available on mobile devices.

This study aims to observe stroke survivors' interaction with hand rehabilitation devices and to understand how the different categories of hand function (Action Research Arm Test [ARAT] scores) influence the stroke survivors' rehabilitation goals.

Figure 1. Upper limb rehabilitation technologies and tools used: (A) mirror (mirror therapy), (B) NeuroBall device, (C) Semi-Circular Peg Board, and (D) GripAble device.



Methods

Participants

Participants were recruited from cohorts of stroke survivors attending a rehabilitation intervention at a cocreation center for accessible rehabilitation technology [21] between September 2021 and April 2023. The inclusion criteria for this study have been described in detail previously [21]; briefly, participants had to have had a stroke within the last 12 months that resulted in mobility problems, be aged over 18 years, be well enough to engage in light to moderate exercise, and be able to attend the rehabilitation program at least twice a week. A range of outcome measures were taken before and after the program, including the ARAT. An overview of the full rehabilitation program is available in our previously published report [21].

Out of a total of 36 participants who agreed to take part in the intervention, 7 (19%) were excluded from this study. Of the 7 excluded persons, 5 (71%) withdrew from the intervention (2/5, 40% withdrew before the commencement and 3/5, 60% withdrew due to ill health or unwillingness to continue), and the other 2 (29%) of the 7 were excluded as a result of incomplete data.

The Upper Limb Rehabilitation Intervention

The upper limb intervention involved activities designed to improve the upper limb functions of participants, delivered completely through the use of technology and therapy devices that either stimulated or promoted repetitive and intensive movement training. The upper limb and hand rehabilitation technologies available to the participants in this study are shown in Table 1. The participants spent at least 50-60 minutes of each of the 2-hour sessions engaging with these devices.

Table 1. Upper limb rehabilitation technologies used.

Technology or device	Manufacturer	Function
GripAble	Gripable	It connects wirelessly with an app on a computer tablet [19] to interact with specifically designed therapy games [22], to train 4 different types of upper limb movements, such as grip and release, pronation and supination, wrist flexion and extension, and radius and ulnar deviations.
NeuroBall	Neurofenix	It connects wirelessly with a tablet app and interacts with therapy games specifically designed to exercise the upper limb of stroke survivors [17]. It trains upper limb movements such as finger grip; hand grip; right, left, upward, and downward tilt; and elbow and shoulder movements.
Mirror box	Saebo	It is a form of mental practice that excites the primary motor cortex, thereby evoking the movement of the affected limb, as the participants move the unaffected side while looking into the mirror [23].
Sensory TENS ^a	Med-Fit	It is a noninvasive nerve stimulator used to relieve pain [24], stimulate the muscles, and relieve muscle stiffness [25].
Semi-Circular Peg Board	Rolyan	It is a therapy tool designed to improve upper limb strength, movement coordination, endurance, and range of motion. It aims to improve hand dexterity.
Armeo Spring	Hocoma	It provides arm weight support while encouraging users to carry out self-initiated arm movements in the shoulder, elbow, and wrist joints and trains different upper limb movements [26].
Vibrating or hot compress massage ball	Dongguan Kooeej	It stimulates the hand using the vibrations delivered at different intensities.
VR ^b headset	Oculus Quest with In-cisiv software	It immerses the user into a virtual environment, thereby encouraging them to use their affected limb to interact with functional tasks [27,28].

^aTENS: transcutaneous electrical nerve stimulation.

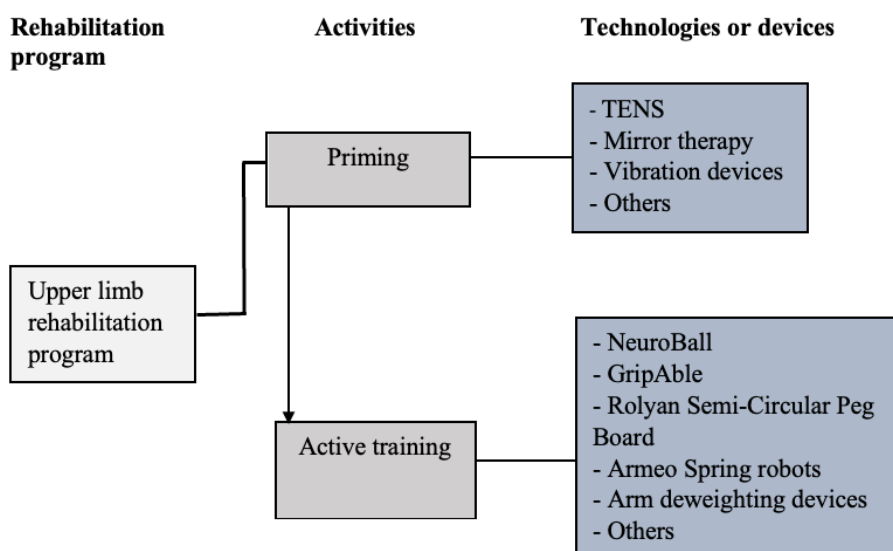
^bVR: virtual reality.

Overview of the Upper Limb Rehabilitation Program

Figure 2 is a representation of the upper limb rehabilitation program used in the rehabilitation gym. The activities were divided into 2 categories. The first part aimed at priming the brain to prepare it for plastic response [29]. Priming focused on sensory stimulation including mirror therapy and electrical, thermal, and vibrational stimulation. These priming activities

comprised the first 15-20 minutes of each rehabilitation session. This second part, that is, the “active training,” aimed to engage the participants in high-intensity motor tasks such as object grip and release, object manipulation, and reach to grasp, designed to improve range of motion, strength, and control. The participants were not limited in terms of the number of devices they could use.

Figure 2. Upper limb rehabilitation program model for stroke survivors. TENS: transcutaneous electrical nerve stimulation.



Categorizing Participants Into Different Hand Function Groups

Participants were given a 1-day initial appointment with a therapist at the rehabilitation gym before the commencement of the 8-week rehabilitation intervention. During this

appointment, demographic data including stroke history were collected, along with a range of baseline assessments for mobility, communication, and cognition, including the ARAT [30]. The ARAT was used to categorize the participants into 3 different hand function groups: poor (scores of 0-9), moderate (scores of 10-56), and good (score of 57) [30].

Understanding the Rehabilitation Goals of Those With Different Categories of Hand Function

During the preintervention visit, participants were allowed to communicate their rehabilitation goals and interact with the upper limb devices to understand how they are set up and operated. The rehabilitation goals of the participants were summarized based on their different hand functions to help understand the needs of stroke survivors who fall under each of the different hand functions, particularly the hand rehabilitation goals of those with poor hand function.

Observing the Interaction of Those With Poor Hand Function and the Hand Rehabilitation Devices

Following the goal setting and initial interaction with the devices, a rehabilitation program was drawn up. The rehabilitation program was individually tailored by a physiotherapist using the rehabilitation goals of the participants. The program however only acted as a guide, as participants had the freedom to interact with any of the devices. The ability of the participants to use each rehabilitation device was observed and recorded. At the end of the intervention, all the observations from participants with poor hand function were gathered and studied to see how they interacted with the hand-based rehabilitation devices. Three of the upper limb devices—the GripAble, NeuroBall, and Semi-Circular Peg Board (see Figure 1)—were selected for observation in this study. The reason for selecting these devices is because these 3 devices were the only devices listed under the “active training” category (see Figure 2) at the time of the study that were used to primarily train motor activities in the hand (involving the wrist and fingers) in addition to training other parts of the upper limb.

Data Organization and Analysis

The simple percentage method was used to estimate the percentage of stroke survivors who fall into each category of hand function. A 1-way ANOVA was carried out using Minitab

statistical software (Minitab LLC), with the Dunnett multiple comparison method used to compare the ages of the group with poor hand function to those with moderate and good hand function.

Ethical Considerations

This study was approved by the University of Strathclyde ethics committee (approval UEC 20/08). The participants provided written informed consent before the study, and their participation was voluntary (no compensation was provided). All identifiable data were pseudoanonymized and replaced with a code.

Results

Categorizing Participants Into Different Hand Function Groups

Observations from 29 participants were included in this study. Their average age was 59.10 (SD 13.62) years with an average of 3.140 (SD 2.31) years after stroke. Of the 29 participants, 17 (59%) were hemiplegic on the left side of their body, whereas the remaining 12 (41%) were hemiplegic on the right side of their body (Table 2).

Of the 29 participants, 10 (34%) scored between 0 and 9 on the ARAT and were grouped as having poor hand function, 17 (59%) scored between 10 and 56 on the ARAT and were grouped as having moderate hand function, and 2 (7%) scored 57 on the ARAT and were grouped as having a good hand function. There was no statistical difference in age between the poor hand function group and both the moderate hand function ($P=.06$) and the good hand function ($P=.37$) groups. Similarly, there was equally no difference in the years after a stroke between the poor hand function group and both the moderate hand function ($P=.09$), and the good hand function ($P=.99$) groups. There was also no observed difference in the hemiplegic side of those with poor hand function (left: 5/10, 50%; right: 5/10, 50%).

Table 2. Characteristics of participants and the 3 subgroups.

Group	Participants (n=29), n	Hand function	Age (years), mean (SD)	Years after stroke, mean (SD)	Hemiplegic side, n (%)		ARAT ^a score, mean (SD)
					Left	Right	
All	29 (100)	— ^b	59.10 (13.62)	3.14 (2.31)	17 (59) ^c	12 (41) ^c	26.63 (21.51)
1	10 (34)	Poor	64.70 (8.83)	2.10 (1.45)	5 (50) ^d	5 (50) ^d	2.00 (3.74)
2	17 (59)	Moderate	53.76 (13.89)	3.88 (2.57)	11 (65) ^e	6 (35) ^e	34.65 (16.09)
3	2 (7)	Good	76.50 (0.707)	2.00 (1.42)	1 (50) ^f	1 (50) ^f	57.00 (0.00)

^aARAT: Action Research Arm Test.

^bNot applicable.

^cn=29.

^dn=10.

^en=17.

^fn=2.

Understanding the Rehabilitation Goals of Those With Different Categories of Hand Function

Table 3 shows a summary of the rehabilitation goals of stroke survivors based on their different hand functions. Participants with poor hand function stated goals that were more toward gaining movements in different parts of their upper limb, as well as improving the ability to carry out active movements that

will enable them to grasp and release objects. However, stroke survivors with moderate and good hand function had goals that were focused on how to improve grip strength, fine motor movements, release time, as well as purposeful movement of the upper limb (see Table 3). Those with poor hand function who recorded a score greater than 0 on the ARAT equally communicated the need to improve grip strength.

Table 3. Upper limb and hand rehabilitation goals of participants separated into the 3 functional categories.

Group	Hand function	Rehabilitation goals as stated by the participants
1	Poor	<ul style="list-style-type: none"> • Gain the ability to hold objects (eg, paper) • Gain some shoulder movement • Gain arm movement • Recovery of any movement, primarily in the shoulder • Improve the grasp and release of objects • Improve active movements • Grip strength^a
2	Moderate	<ul style="list-style-type: none"> • Improve dexterity • Improve grip • Improve the range of upper limb movement • Improve upper limb strength • Improve supination or pronation range • Improve the grasp and release of objects • Improve release time • Gain the ability for small object manipulation • Gain the ability to move objects • Gain the ability for purposeful movement of the upper limb
3	Good	<ul style="list-style-type: none"> • Increase grip • Improve wrist extension

^aFor those who recorded a score >0 on the Action Research Arm Test (ARAT).

Interaction With Hand Rehabilitation Technologies by the Poor Hand Function Group

Table 4 shows that 8 (80%) of the 10 participants with poor hand function could not interact with any of the 3 aforementioned devices to carry out active training. This value

represents 28% (8/29) of the total population in this study. Only 2 (20%) of the 10 participants with poor hand function were able to engage with these devices; the ARAT score shows that these 2 participants had ARAT scores of 7 and 9, compared to the score of 0 that was recorded by the other 8 who were not able to engage with these devices.

Table 4. Interaction of stroke survivors who had poor hand function with the hand rehabilitation devices.

Participant ID	ARAT ^a score	Upper limb rehabilitation goal	Use of devices for active hand training			Comments on the participants' ability to use the devices
			GripAble	NeuroBall	Semi-Circular Peg Board	
1	0	General upper limb function	X ^b	X	X	Tightness in the hand and other parts of the upper limb did not allow the fitting of the devices into the hand
2	0	Improve active movements	X	X	X	Weakness of the upper limb and hand; not able to carry out the active movement necessary for device usage
3	0	Hold objects (eg, paper), gain some shoulder movement	X	X	X	Could not make use of any of the devices
4	0	Improve the grasp and release of object	X	X	X	Difficult to initiate movement on the GripAble and NeuroBall; could also not use the Semi-Circular Peg Board as a result of weakness in the hand
5	7	Grip strength, range of shoulder or elbow active movement	✓ ^c	✓	✓	Fought to maintain grip due to the presence of tightness; the participant noted that "Botox [had] not helped a lot" with hand function. However, they were able to make use of the devices
6	0	Gain arm movement	X	X	X	Upper limb and hand stiffness affected the ability to access the devices
7	0	Would like to get some movement	X	X	X	Had very limited movements
8	0	Recovery of any movement, primarily in the shoulder	X	X	X	Weakness of the upper limb and hand; not able to carry out active movement necessary for device usage
9	0	— ^d	X	X	X	Attempted the GripAble and NeuroBall once but was not able to make use of them
10	9	Grip strength	✓	✓	✓	—

^aARAT: Action Research Arm Test.

^bX: unable.

^c✓: able.

^dNot applicable.

Discussions

Principal Findings

This study was carried out to observe how stroke survivors with poor hand function interacted with hand rehabilitation devices such as the GripAble, NeuroBall, and Semi-Circular Peg Board. The findings show that stroke survivors whose poor hand function leads to an ARAT score of 0 cannot actively interact with hand rehabilitation devices.

Comparison to Prior Work

About two-thirds (55%-75%) of persons who had a stroke sustain upper limb impairments [7]. The extent of the impairments varies from person to person (see Table 2). In some, it results in poor hand function, whereas others present moderate or good hand function. The level of hand function present after stroke subsequently influences the upper limb rehabilitation goals of the stroke survivor (see Table 3). Stroke survivors with moderate to good hand function, who are likely to possess some range of motion in the hand, can grip, grasp, or pinch [30,31] hand rehabilitation devices and so have upper limb rehabilitation goals aimed at strengthening the existing

motor ability. These goals may be related to improving grip strength and endurance, the ability to release objects or release time, the existing range of upper limb movements, and finger dexterity and regaining the ability to manipulate small objects (see Table 3). However, those with poor hand function, especially those with an ARAT score of 0 who cannot grasp, grip, or pinch objects irrespective of the sizes [31], have upper limb rehabilitation goals that focus on recovering some movement in the joints (shoulder, elbow, wrist, and/or fingers; see Tables 3 and 4).

Muscle weakness and the appearance of muscle stiffness, tightness, or tone (evident by the presence of a clenched hand) were clinically examined as being responsible for the poor hand function of the participants in this study (see Figure 3). The appearance of clenched hands has been reported as a clinical feature of spasticity [32]; moreover, the presence of muscle stiffness, tightness, and tone have all been connected with spasticity [33,34]. Previous studies have reported both spasticity and muscle weakness as the 2 major motor impairments following a stroke [35,36]. The severity of these impairments led to difficulty in hand immobility in 80% of those with poor hand function (with an ARAT score of 0), and according to an

earlier report [36], spasticity and muscle weakness can result in immobility.

Figure 3. Participants with poor hand function taking part in the 8-week rehabilitation exercise.



Strengths

The *UK National Clinical Guideline for Stroke* stipulates that stroke survivors should be considered for rehabilitation at any point after the stroke to potentially gain benefits [37]. However, an earlier study [38] that measured the accuracy of physical therapists' early prediction of upper limb function reported that stroke survivors with ARAT scores more than 10 are those principally qualified to undergo rehabilitation exercises; this potentially excludes stroke survivors with poor hand function from taking part in hand rehabilitation. This study shows that not all stroke survivors with poor hand function should be considered ineligible to make use of hand rehabilitation devices, as those with some range of motion in their hand, as seen in participants with ARAT scores of 7 and 9 (see [Table 4](#)), can still benefit from hand rehabilitation devices and thus active hand rehabilitation.

Limitations

Only participants who exhibited poor hand function with an ARAT score of 0 were not able to benefit from active hand rehabilitation using devices. Those in this category whose poor hand function was due to muscle weakness were unable to carry out any intended active movement on the hand rehabilitation devices (see [Table 4](#)), even when supported to place their hand on them. In contrast, those whose poor hand function was due to hand stiffness or tightness, in addition to their inability to carry out intended active movement, were also faced with the problem of accessibility, which made it difficult for them to fit the device.

A limitation of this study was the inability to assess these conditions (muscle weakness and muscle tone or tightness)—examined to be responsible for the poor hand function—using the relevant outcome measures, such as motricity index, grip strength or pinch strength (for muscle

weakness), or the Modified Ashworth Scale (for spasticity) [39], to quantify their severity. However, their severity was such that the hand was not useful in carrying out any of the ARAT tasks [31], as indicated by an ARAT score of 0.

Future Direction

Improvement in technological advancement has led to the development of devices such as rehabilitation gloves (smart or robotic gloves) that can be useful in stretching the hands of stroke survivors with poor hand function without requiring their active participation [40,41]. However, only stroke survivors with low spasticity (who possess some range of active motion in the hand [42]) may be able to make use of these rehabilitation gloves [40]. This means those with considerable muscle stiffness resulting in difficulty in passive motion [42] are still unlikely to freely access these devices; thus, future design of rehabilitation devices for hand rehabilitation should consider the problem of device accessibility in people with poor hand function due to considerable muscle stiffness or tightness.

Conclusions

It is therefore concluded that not all stroke survivors with impairments in their hands can interact with the available hand rehabilitation technologies, as those with an ARAT score of 0 cannot actively interact with any hand rehabilitation device. Thus, the selection of devices for hand rehabilitation should first consider the hand function of the affected stroke survivor. Since muscle stiffness or tightness in the hand results in poor hand function that can impede access to hand rehabilitation devices, future design of devices for hand rehabilitation should consider the accessibility needs of those with poor hand function as a result of hand stiffness or tightness. A similar observational study involving more stroke survivors will help ascertain the percentage of stroke survivors who fall into the category of having poor hand function and is therefore recommended.

Acknowledgments

This study would not have been possible without the support of the Sir Jules Thorn Centre for the Co-creation of Rehabilitation Technology, University of Strathclyde.

Data Availability

The data sets generated during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors contributed to the study's methodology, investigation, and administration. Specifically, COW was involved with the conceptualizing, original draft writing, formal analysis, and visualization of the work. GS and MS were involved with editing and review of the draft, and AK was involved with the supervision of the project.

Conflicts of Interest

None declared.

References

1. Sarikaya H, Ferro J, Arnold M. Stroke prevention--medical and lifestyle measures. *Eur Neurol* 2015 Jan 6;73(3-4):150-157 [FREE Full text] [doi: [10.1159/000367652](https://doi.org/10.1159/000367652)] [Medline: [25573327](https://pubmed.ncbi.nlm.nih.gov/25573327/)]
2. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet* 2014 Jan 18;383(9913):245-254 [FREE Full text] [doi: [10.1016/s0140-6736\(13\)61953-4](https://doi.org/10.1016/s0140-6736(13)61953-4)] [Medline: [24449944](https://pubmed.ncbi.nlm.nih.gov/24449944/)]
3. King D, Wittenberg R, Patel A, Quayyum Z, Berdunov V, Knapp M. The future incidence, prevalence and costs of stroke in the UK. *Age Ageing* 2020 Feb 27;49(2):277-282 [FREE Full text] [doi: [10.1093/ageing/afz163](https://doi.org/10.1093/ageing/afz163)] [Medline: [31957781](https://pubmed.ncbi.nlm.nih.gov/31957781/)]
4. Rucker V, Wiedmann S, O'Flaherty M, Busch MA, Heuschmann PU. Decline in regional trends in mortality of stroke subtypes in Germany from 1998 to 2015. *Stroke* 2018 Nov;49(11):2577-2583. [doi: [10.1161/STROKEAHA.118.023193](https://doi.org/10.1161/STROKEAHA.118.023193)] [Medline: [30355214](https://pubmed.ncbi.nlm.nih.gov/30355214/)]
5. Patel A, Berdunov V, Quayyum Z, King D, Knapp M, Wittenberg R. Estimated societal costs of stroke in the UK based on a discrete event simulation. *Age Ageing* 2020 Feb 27;49(2):270-276 [FREE Full text] [doi: [10.1093/ageing/afz162](https://doi.org/10.1093/ageing/afz162)] [Medline: [31846500](https://pubmed.ncbi.nlm.nih.gov/31846500/)]
6. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. *Lancet Neurol* 2009 Aug;8(8):741-754. [doi: [10.1016/S1474-4422\(09\)70150-4](https://doi.org/10.1016/S1474-4422(09)70150-4)] [Medline: [19608100](https://pubmed.ncbi.nlm.nih.gov/19608100/)]
7. Lai SM, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the Stroke Impact Scale. *Stroke* 2002 Jul;33(7):1840-1844. [doi: [10.1161/01.str.0000019289.15440.f2](https://doi.org/10.1161/01.str.0000019289.15440.f2)] [Medline: [12105363](https://pubmed.ncbi.nlm.nih.gov/12105363/)]
8. Buccino G, Solodkin A, Small SL. Functions of the mirror neuron system: implications for neurorehabilitation. *Cogn Behav Neurol* 2006 Mar;19(1):55-63. [doi: [10.1097/00146965-200603000-00007](https://doi.org/10.1097/00146965-200603000-00007)] [Medline: [16633020](https://pubmed.ncbi.nlm.nih.gov/16633020/)]
9. Borghese NA, Essenziale J, Mainetti R, Mancon E, Pagliaro R, Pajardi G. Hand rehabilitation and telemonitoring through smart toys. *Sensors (Basel)* 2019 Dec 13;19(24):5517 [FREE Full text] [doi: [10.3390/s19245517](https://doi.org/10.3390/s19245517)] [Medline: [31847216](https://pubmed.ncbi.nlm.nih.gov/31847216/)]
10. Ates S, Haarman CJW, Stienen AHA. SCRIPT passive orthosis: design of interactive hand and wrist exoskeleton for rehabilitation at home after stroke. *Auton Robot* 2016 Jul 12;41(3):711-723. [doi: [10.1007/s10514-016-9589-6](https://doi.org/10.1007/s10514-016-9589-6)]
11. Pollock A, Baer G, Campbell P, Choo PL, Forster A, Morris J, et al. Physical rehabilitation approaches for the recovery of function and mobility after stroke. *Stroke* 2014 Oct;45(10):e202. [doi: [10.1161/strokeaha.114.006275](https://doi.org/10.1161/strokeaha.114.006275)]
12. Iemmi V, Gibson L, Blanchet K, Kumar KS, Rath S, Hartley S, et al. Community-based rehabilitation for people with disabilities in low-and middle-income countries: a systematic review. *Campbell Syst Rev* 2015 Sep 1;11(1):1-177. [doi: [10.4073/csr.2015.15](https://doi.org/10.4073/csr.2015.15)]
13. Mauritz KH. Gait training in hemiplegia. *Eur J Neurol* 2002 May 25;9 Suppl 1(s1):23-29; discussion 53. [doi: [10.1046/j.1468-1331.2002.0090s1023.x](https://doi.org/10.1046/j.1468-1331.2002.0090s1023.x)] [Medline: [11918646](https://pubmed.ncbi.nlm.nih.gov/11918646/)]
14. Korzeniewska E, Krawczyk A, Mróz J, Wyszńska E, Zawisła R. Applications of smart textiles in post-stroke rehabilitation. *Sensors (Basel)* 2020 Apr 22;20(8):2370 [FREE Full text] [doi: [10.3390/s20082370](https://doi.org/10.3390/s20082370)] [Medline: [32331218](https://pubmed.ncbi.nlm.nih.gov/32331218/)]
15. Kim D. The effects of hand strength on upper extremity function and activities of daily living in stroke patients, with a focus on right hemiplegia. *J Phys Ther Sci* 2016 Sep;28(9):2565-2567 [FREE Full text] [doi: [10.1589/jpts.28.2565](https://doi.org/10.1589/jpts.28.2565)] [Medline: [27799695](https://pubmed.ncbi.nlm.nih.gov/27799695/)]
16. Everard G, Declerck L, Detrembleur C, Leonard S, Bower G, Dehem S, et al. New technologies promoting active upper limb rehabilitation after stroke: an overview and network meta-analysis. *Eur J Phys Rehabil Med* 2022 Aug;58(4):530-548 [FREE Full text] [doi: [10.23736/S1973-9087.22.07404-4](https://doi.org/10.23736/S1973-9087.22.07404-4)] [Medline: [35666491](https://pubmed.ncbi.nlm.nih.gov/35666491/)]
17. Kilbride C, Scott DJM, Butcher T, Norris M, Ryan JM, Anokye N, et al. Rehabilitation via home based gaming exercise for the upper-limb post stroke (RHOMBUS): protocol of an intervention feasibility trial. *BMJ Open* 2018 Nov 21;8(11):e026620 [FREE Full text] [doi: [10.1136/bmjopen-2018-026620](https://doi.org/10.1136/bmjopen-2018-026620)] [Medline: [30467137](https://pubmed.ncbi.nlm.nih.gov/30467137/)]
18. Myers M. Rehab device enables stroke survivors with arm disabilities to do more training. Imperial College London. 2021 Sep 1. URL: <https://tinyurl.com/wxydv9dc> [accessed 2022-02-11]
19. Mutalib SA, Mace M, Seager C, Burdet E, Mathiowetz V, Goldsmith N. Modernising grip dynamometry: inter-instrument reliability between GripAble and Jamar. *BMC Musculoskelet Disord* 2022 Jan 24;23(1):80 [FREE Full text] [doi: [10.1186/s12891-022-05026-0](https://doi.org/10.1186/s12891-022-05026-0)] [Medline: [35073887](https://pubmed.ncbi.nlm.nih.gov/35073887/)]

20. Rinne P, Mace M, Nakornchai T, Zimmerman K, Fayer S, Sharma P, et al. Democratizing neurorehabilitation: how accessible are low-cost mobile-gaming technologies for self-rehabilitation of arm disability in stroke? *PLoS One* 2016 Oct 5;11(10):e0163413 [FREE Full text] [doi: [10.1371/journal.pone.0163413](https://doi.org/10.1371/journal.pone.0163413)] [Medline: [27706248](https://pubmed.ncbi.nlm.nih.gov/27706248/)]
21. Kerr A, Grealy MA, Kuschmann A, Rutherford R, Rowe P. A co-creation centre for accessible rehabilitation technology. *Front Rehabil Sci* 2021 Jan 7;2:820929 [FREE Full text] [doi: [10.3389/freesc.2021.820929](https://doi.org/10.3389/freesc.2021.820929)] [Medline: [36188853](https://pubmed.ncbi.nlm.nih.gov/36188853/)]
22. Mace M, Rinne P, Liardon JL, Uhomobhi C, Bentley P, Burdet E. Elasticity improves handgrip performance and user experience during visuomotor control. *R Soc Open Sci* 2017 Feb;4(2):160961 [FREE Full text] [doi: [10.1098/rsos.160961](https://doi.org/10.1098/rsos.160961)] [Medline: [28386448](https://pubmed.ncbi.nlm.nih.gov/28386448/)]
23. Garry MI, Loftus A, Summers JJ. Mirror, mirror on the wall: viewing a mirror reflection of unilateral hand movements facilitates ipsilateral M1 excitability. *Exp Brain Res* 2005 May 8;163(1):118-122. [doi: [10.1007/s00221-005-2226-9](https://doi.org/10.1007/s00221-005-2226-9)] [Medline: [15754176](https://pubmed.ncbi.nlm.nih.gov/15754176/)]
24. Johnson M. Transcutaneous electrical nerve stimulation: mechanisms, clinical application and evidence. *Rev Pain* 2007 Aug;1(1):7-11 [FREE Full text] [doi: [10.1177/204946370700100103](https://doi.org/10.1177/204946370700100103)] [Medline: [26526976](https://pubmed.ncbi.nlm.nih.gov/26526976/)]
25. Mahmood A, Veluswamy SK, Hombali A, Mullick A, Solomon JM. Effect of transcutaneous electrical nerve stimulation on spasticity in adults with stroke: a systematic review and meta-analysis. *Arch Phys Med Rehabil* 2019 Apr;100(4):751-768. [doi: [10.1016/j.apmr.2018.10.016](https://doi.org/10.1016/j.apmr.2018.10.016)] [Medline: [30452892](https://pubmed.ncbi.nlm.nih.gov/30452892/)]
26. Hamzah N, Giban NI, Mazlan M. Robotic upper limb rehabilitation using Armeo®Spring for chronic stroke patients at University Malaya Medical Centre (UMMC). 2017 Dec 7 Presented at: 2nd International Conference for Innovation in Biomedical Engineering and Life Sciences; December 10-13, 2017; Penang, Malaysia p. 225-230. [doi: [10.1007/978-981-10-7554-4_39](https://doi.org/10.1007/978-981-10-7554-4_39)]
27. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. *Cochrane Database Syst Rev* 2017 Nov 20;11(11):CD008349 [FREE Full text] [doi: [10.1002/14651858.CD008349.pub4](https://doi.org/10.1002/14651858.CD008349.pub4)] [Medline: [29156493](https://pubmed.ncbi.nlm.nih.gov/29156493/)]
28. Bui J, Luauté J, Farnè A. Enhancing upper limb rehabilitation of stroke patients with virtual reality: a mini review. *Front Virtual Real* 2021 Nov 8;2:595771. [doi: [10.3389/frvir.2021.595771](https://doi.org/10.3389/frvir.2021.595771)]
29. da Silva ESM, Ocamoto GN, Santos-Maia GLD, de Fátima Carreira Moreira Padovez R, Trevisan C, de Noronha MA, et al. The effect of priming on outcomes of task-oriented training for the upper extremity in chronic stroke: a systematic review and meta-analysis. *Neurorehabil Neural Repair* 2020 Jun 26;34(6):479-504. [doi: [10.1177/1545968320912760](https://doi.org/10.1177/1545968320912760)] [Medline: [32452242](https://pubmed.ncbi.nlm.nih.gov/32452242/)]
30. Buma FE, Raemaekers M, Kwakkel G, Ramsey NF. Brain function and upper limb outcome in stroke: a cross-sectional fMRI study. *PLoS One* 2015 Oct 6;10(10):e0139746 [FREE Full text] [doi: [10.1371/journal.pone.0139746](https://doi.org/10.1371/journal.pone.0139746)] [Medline: [26440276](https://pubmed.ncbi.nlm.nih.gov/26440276/)]
31. Wilson N, Howel D, Bosomworth H, Shaw L, Rodgers H. Analysing the Action Research Arm Test (ARAT): a cautionary tale from the RATULS trial. *Int J Rehabil Res* 2021 Jun 01;44(2):166-169 [FREE Full text] [doi: [10.1097/MRR.0000000000000466](https://doi.org/10.1097/MRR.0000000000000466)] [Medline: [33741815](https://pubmed.ncbi.nlm.nih.gov/33741815/)]
32. Nair KPS, Marsden J. The management of spasticity in adults. *BMJ* 2014 Aug 05;349:g4737. [doi: [10.1136/bmj.g4737](https://doi.org/10.1136/bmj.g4737)] [Medline: [25096594](https://pubmed.ncbi.nlm.nih.gov/25096594/)]
33. Sommerfeld DK, Eek EU, Svensson A, Holmqvist LW, von Arbin MH. Spasticity after stroke. *Stroke* 2004 Jan;35(1):134-139. [doi: [10.1161/01.str.0000105386.05173.5e](https://doi.org/10.1161/01.str.0000105386.05173.5e)]
34. Francisco GE, Wissel J, Platz T, Li S. Post-stroke spasticity. In: Platz T, editor. *Clinical Pathways in Stroke Rehabilitation*. Cham, Switzerland: Springer; Jan 15, 2021:149-173.
35. Li S. Spasticity, motor recovery, and neural plasticity after stroke. *Front Neurol* 2017 Apr 03;8:120 [FREE Full text] [doi: [10.3389/fneur.2017.00120](https://doi.org/10.3389/fneur.2017.00120)] [Medline: [28421032](https://pubmed.ncbi.nlm.nih.gov/28421032/)]
36. Raghavan P. Upper limb motor impairment after stroke. *Phys Med Rehabil Clin N Am* 2015 Nov;26(4):599-610 [FREE Full text] [doi: [10.1016/j.pmr.2015.06.008](https://doi.org/10.1016/j.pmr.2015.06.008)] [Medline: [26522900](https://pubmed.ncbi.nlm.nih.gov/26522900/)]
37. National Clinical Guideline for Stroke for the UK and Ireland. London, United Kingdom: Intercollegiate Stroke Working Party; 2023 May 4. URL: <https://www.strokeguideline.org> [accessed 2023-09-12]
38. Nijland RHM, van Wegen EEH, Harmeling-van der Wel BC, Kwakkel G, Early Prediction of Functional Outcome After Stroke Investigators. Accuracy of physical therapists' early predictions of upper-limb function in hospital stroke units: the EPOS Study. *Phys Ther* 2013 Apr;93(4):460-469. [doi: [10.2522/ptj.20120112](https://doi.org/10.2522/ptj.20120112)] [Medline: [23139424](https://pubmed.ncbi.nlm.nih.gov/23139424/)]
39. Lang CE, Bland MD, Bailey RR, Schaefer SY, Birkenmeier RL. Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. *J Hand Ther* 2013;26(2):104-114;quiz 115 [FREE Full text] [doi: [10.1016/j.jht.2012.06.005](https://doi.org/10.1016/j.jht.2012.06.005)] [Medline: [22975740](https://pubmed.ncbi.nlm.nih.gov/22975740/)]
40. Fardipour S, Hadadi M. Investigation of therapeutic effects of wearable robotic gloves on improving hand function in stroke patients: a systematic review. *Curr J Neurol* 2022 Apr 04;21(2):125-132 [FREE Full text] [doi: [10.18502/cjn.v21i2.10496](https://doi.org/10.18502/cjn.v21i2.10496)] [Medline: [38011474](https://pubmed.ncbi.nlm.nih.gov/38011474/)]
41. Kang M, Yun SJ, Lee SY, Oh B, Lee HH, Lee S, et al. Effects of upper-extremity rehabilitation using smart glove in patients with subacute stroke: results of a prematurely terminated multicenter randomized controlled trial. *Front Neurol* 2020 Nov 9;11:580393 [FREE Full text] [doi: [10.3389/fneur.2020.580393](https://doi.org/10.3389/fneur.2020.580393)] [Medline: [33240205](https://pubmed.ncbi.nlm.nih.gov/33240205/)]

42. Harb A, Kishner S. Modified Ashworth Scale. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2024.

Abbreviations

ARAT: Action Research Arm Test

Edited by T Leung; submitted 31.10.23; peer-reviewed by J Quinzaños, A Perez Sanpablo; comments to author 15.02.24; revised version received 10.04.24; accepted 01.06.24; published 26.06.24.

Please cite as:

Wodu CO, Sweeney G, Slachetka M, Kerr A

Stroke Survivors' Interaction With Hand Rehabilitation Devices: Observational Study

JMIR Biomed Eng 2024;9:e54159

URL: <https://biomedeng.jmir.org/2024/1/e54159>

doi: [10.2196/54159](https://doi.org/10.2196/54159)

PMID: [38922668](https://pubmed.ncbi.nlm.nih.gov/38922668/)

©Chiomia Obinuchi Wodu, Gillian Sweeney, Milena Slachetka, Andrew Kerr. Originally published in JMIR Biomedical Engineering (<http://biomsedeng.jmir.org>), 26.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Classifying Residual Stroke Severity Using Robotics-Assisted Stroke Rehabilitation: Machine Learning Approach

Russell Jeter^{1,2*}, PhD; Raymond Greenfield^{1*}, MSci; Stephen N Housley^{2,3}, PT, DPT, PhD; Igor Belykh^{1,4}, Prof Dr, PhD

¹Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, United States

²Motus Nova, LLC, Atlanta, GA, United States

³Laboratory for Sensorimotor Integration, Georgia Institute of Technology, Atlanta, GA, United States

⁴Neuroscience Institute, Georgia State University, Atlanta, GA, United States

*these authors contributed equally

Corresponding Author:

Igor Belykh, Prof Dr, PhD

Department of Mathematics and Statistics

Georgia State University

PO Box 4110

Atlanta, GA, 30302 410

United States

Phone: 1 404 413 6411

Fax: 1 404 413 6403

Email: ibelykh@gsu.edu

Abstract

Background: Stroke therapy is essential to reduce impairments and improve motor movements by engaging autogenous neuroplasticity. Traditionally, stroke rehabilitation occurs in inpatient and outpatient rehabilitation facilities. However, recent literature increasingly explores moving the recovery process into the home and integrating technology-based interventions. This study advances this goal by promoting in-home, autonomous recovery for patients who experienced a stroke through robotics-assisted rehabilitation and classifying stroke residual severity using machine learning methods.

Objective: Our main objective is to use kinematics data collected during in-home, self-guided therapy sessions to develop supervised machine learning methods, to address a clinician's autonomous classification of stroke residual severity-labeled data toward improving in-home, robotics-assisted stroke rehabilitation.

Methods: In total, 33 patients who experienced a stroke participated in in-home therapy sessions using Motus Nova robotics rehabilitation technology to capture upper and lower body motion. During each therapy session, the Motus Hand and Motus Foot devices collected movement data, assistance data, and activity-specific data. We then synthesized, processed, and summarized these data. Next, the therapy session data were paired with clinician-informed, discrete stroke residual severity labels: "no range of motion (ROM)," "low ROM," and "high ROM." Afterward, an 80%:20% split was performed to divide the dataset into a training set and a holdout test set. We used 4 machine learning algorithms to classify stroke residual severity: light gradient boosting (LGB), extra trees classifier, deep feed-forward neural network, and classical logistic regression. We selected models based on 10-fold cross-validation and measured their performance on a holdout test dataset using F_1 -score to identify which model maximizes stroke residual severity classification accuracy.

Results: We demonstrated that the LGB method provides the most reliable autonomous detection of stroke severity. The trained model is a consensus model that consists of 139 decision trees with up to 115 leaves each. This LGB model boasts a 96.70% F_1 -score compared to logistic regression (55.82%), extra trees classifier (94.81%), and deep feed-forward neural network (70.11%).

Conclusions: We showed how objectively measured rehabilitation training paired with machine learning methods can be used to identify the residual stroke severity class, with efforts to enhance in-home self-guided, individualized stroke rehabilitation. The model we trained relies only on session summary statistics, meaning it can potentially be integrated into similar settings for real-time classification, such as outpatient rehabilitation facilities.

(*JMIR Biomed Eng* 2024;9:e56980) doi:[10.2196/56980](https://doi.org/10.2196/56980)

KEYWORDS

stroke; rehabilitation robotics; machine learning; artificial intelligence; physical therapy; neuroplasticity

Introduction

Stroke is a leading cause of mortality and disability worldwide, and the economic costs of treatment and poststroke care are substantial [1]. In 2019, there were 12.2 million incident cases of stroke, 101 million prevalent stroke cases, and 6.55 million deaths from stroke [2]. The severity of a stroke can range from mild to severe, with severe strokes often leading to long-term disability or even death. Stroke rehabilitation typically involves a team of health care professionals, including doctors, nurses, therapists, and other specialists. The specific goals and interventions of stroke rehabilitation vary depending on the individual's needs and abilities. They may include physical therapy to improve mobility; occupational therapy to improve the ability to perform daily activities; speech therapy to improve communication skills; and cognitive therapy to improve memory, problem-solving, and other cognitive abilities. While traditionally recovery has taken place in inpatient and outpatient rehabilitation facilities, there is growing recent literature about moving the recovery process into the home [3,4] and integrating technology-based interventions [5]. This study takes steps to achieve this goal of in-home and autonomous recovery for patients who experienced a stroke via robotics-assisted stroke rehabilitation and classification of stroke residual severity via machine learning methods.

Machine learning in health care and stroke rehabilitation is not a new concept (see Reyna et al [6], Alabi et al [7], Cerasa et al [8], and Harari et al [9] as notable examples of this vast research field and Campagnini et al [10] for a systematic review of machine learning methods for poststroke rehabilitation recovery prediction). In particular, multiple studies have been performed to predict outcomes in patient survival, locoregional recurrences, and long-term outcomes in patients who experienced an ischemic stroke [11-15]. Similarly, studies focused on motor function have leveraged retrospective health care data and targeted predicting the short- and long-term functional ability [16-18]. Such studies represent an exciting step forward in stroke rehabilitation but have some limitations. These limitations include the use of health care data that are infrequently measured (sometimes entirely limited to admission data), which can hamper the performance of models that rely on large datasets for generalizability. Similarly, most studies limit their scope to predicting short- and long-term outcomes and may fail to capture some of the day-to-day changes survivors' who have experienced a stroke experience.

This study aims to overcome these limitations by quantifying the progress of patient improvement via in-home therapy sessions using Motus Nova robotics rehabilitation technology [19] that captures upper and lower body motion. The Motus Hand and Motus Foot devices are robotic therapeutic devices designed to be used by survivors who have experienced a stroke with residual upper and lower extremity impairments at home without needing help from a clinician or caregiver. The neuromotor mechanism by which the Motus Hand and Motus Foot help rehabilitate patients who have experienced a stroke

is rooted in the results from constraint-induced movement therapy studies [20,21] and focus on getting survivors of stroke high volumes of repetitive task practice. The Motus Hand and Motus Foot engage the affected wrist or ankle of the user, guiding them through various therapeutic exercises targeting various functional tasks (eg, gross motor control, fine motor control, and precision tracking). Earlier versions of the technology have been shown to have clinically significant improvements in depressive symptoms, functional independence, upper extremity use in functional tasks, distance walking, and gait speed [19,22,23].

Traditionally, to determine the functional ability of survivors of stroke, they will be assessed by a clinician during often infrequent clinical visits (whether through an outpatient rehabilitation facility, visiting a neurologist, or a primary care physician). The time scale of these assessments fails to capture the progress made during the recovery process when it happens. Using machine learning and therapy session, kinematic measurements promise to have a central role in rehabilitation decision-making in determining whether patient therapy is improving. Machine learning is the methodology that allows computers to learn from experience. By constructing and training supervised classifiers to learn decision rules from data, automatic solutions can be exploited to make predictions on new data [24,25]. As in many health care, disease, or machine learning research applied in a clinical setting, labeling of patient data by a clinician is necessary [6]. This study applies the same heuristic methodologies. Our goal is to use kinematics data collected during in-home, self-guided therapy sessions to construct supervised machine learning methods to address the autonomous classification of stroke residual severity-labeled data toward improving in-home, robotics-assisted, individualized stroke rehabilitation.

Methods

Therapeutic Intervention Description

The Motus Hand and Motus Foot each consist of 2 major components: a peripheral (see the bottom panel of Figure 1 for a close-up of the Motus Hand peripheral) that the patient attaches to their affected limb and a console that guides their therapy routine and assessment using a video game interface. The peripherals have a pneumatic actuator that can dynamically provide assistance or resistance by filling an air muscle in the peripheral that moves the wrist or ankle joint. The wrist or ankle joint of the peripheral has an embedded angle and pressure sensor that transmits live angle and pressure data to the console. This allows the console to give the user immediate visual feedback of their movement through avatars in a video game on the screen. The therapeutic video game activities can provide a dynamic feedback loop consisting of in-game goals (eg, ships to shoot or coins to collect) that drive user movements, which correspond to movement on the screen, allowing the console to react and set new goals or obstacles. This feedback loop is designed to promote sensory motor function.

A therapy session with the Motus Hand or Motus Foot consists of stretching, gross motor control, fine motor control, and endurance exercises, depending on the patient's needs. This process is depicted in Figure 1, where a Motus Hand user is playing "Cosmic Tennis," a gross motor control exercise that plays like the classic arcade game Pong [26]. The user's wrist or ankle movement corresponds to the movement of the paddle

on the right-hand side of the screen, and the goal is to hit the ball back and forth to score on the artificial intelligence (AI)-controlled opponent. Because of the user-guided nature of a therapy session with the Motus Hand or Motus Foot, therapy sessions can vary greatly in length. In the data collected, therapy sessions range from 5 to 60 minutes and between 1 and 10 therapeutic activities.

Figure 1. Patients do therapy sessions with the Motus Hand or Motus Foot using a pneumatically driven exogenous robotic device worn on the affected hand, arm, or foot (the Motus Hand is depicted in the bottom panel). The peripheral acts as a game controller (through an angle sensor embedded in the wrist joint) that allows users to play therapeutic video games that dynamically adapt to their needs and provide the requisite assistance or resistance (computer screen in the bottom panel).



Study Design

The Motus Hand and Motus Foot collect high-resolution angle and pressure data from sensors embedded in the wrist or ankle joints and the pressure management system. These high-resolution data are collected at a frequency of 30 Hz and stored in a time series database. Other information collected during a therapy session includes score, peripheral type (Motus Hand or Motus Foot), and current game (therapeutic activity). This study used anonymous data collected from 33 patients who experienced a stroke. In total, those patients performed 32,902 therapeutic activities (ie, each unique activity performed in each therapy session). These therapy session data are then divided using an 80%:20% split into a training dataset and a holdout test set. The training set is used for training the classification models, and the test set is reserved for the final model evaluation.

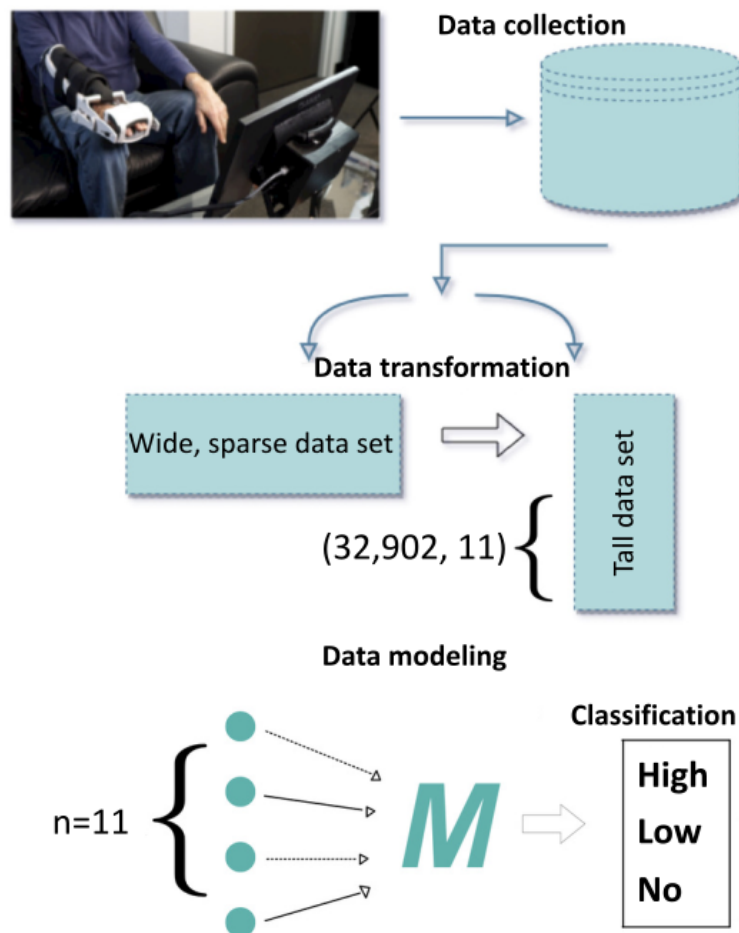
To use the data collected during a therapy session to classify a patient's stroke residual severity autonomously, each patient was given a guided assessment with a clinician using the Motus Hand or Motus Foot to classify them as having a high range of motion (ROM), low ROM, or no ROM. These classification levels are intentionally chosen to be coarse to mimic the environment in a rehabilitation therapy session.

To find an ideal classifier, we use to consider the training and performance of 4 machine learning algorithms: light gradient boosting (LGB) [27], extra trees classifier [28], deep feed-forward neural network (DNN) [29], and multiclass logistic regression (LR) [30]. A practical model is then constructed using the most common data measured in each session based on the maximum score per session per patient. Unsupervised learning methods are then applied to the training dataset, such as the correlation matrix and principal component analysis

(PCA), to show that all variables collected are relevant to the study. After performing dimensionality reduction analysis, the models are selected using 10-fold cross-validation on the training dataset with the mean and SD of accuracy from each computational experiment. Afterward, the following metrics determine the model's performance, including the accuracy, precision, and recall from the confusion matrix. The

macroaverage F_1 -score was used to judge the efficacy of the models, as this is a multiclassification problem [31], and as such, accuracy would be an insufficient measure. Figure 2 provides a high-level overview of the data collection, analysis, processing, and modeling that ultimately produces the final classification results.

Figure 2. Concept diagram of the overall data analysis and modeling. In total, 33 patients perform in-home therapy using the Motus Hand and Motus Foot rehabilitation devices. Sensors in the devices capture live angle and pressure data. These data are then processed and summarized to provide summary statistics of 32,902 therapeutic activities. This provides the base dataset for the analysis presented in this paper, with 11 features and 32,902 points. These data are then split and prepared for use in training a supervised machine learning model to classify the stroke severity of the patient.



Details of Data Collection

Throughout a therapy session using the Motus Hand or Motus Foot, live angle data (measured in degrees from a natural midpoint in wrist or ankle placement) are collected from the sensor embedded in the wrist or ankle joint at 30 readings per second. These “raw” angle sensor data are then stored in a time series database (InfluxDB [32]). In addition to the high-resolution angle data, pressure readings (measured in PSI) are taken from the pressure management system at 30 measurements per second. While these readings are not high resolution compared to state-of-the-art kinematics technology [33], it is significantly higher resolution than what a typical physician would have access to during assessments in a normal physical therapy visit.

Each therapy session for a patient includes a selection of about 30 activities that focus on several types of motor function, including gross motor control, fine motor control, flexor tone

reduction, endurance, reaction time, and tracking. A patient can participate in more than 1 video game (therapeutic activity) during a patient session. The score is recorded and stored once the patient completes the video game. The scores for each game are not necessarily standardized. This means a score of 100 in one game can represent a dramatically different performance than a score of 100 in another. The score is collected each time a player performs an action in the game that would increase or decrease the score, so this field is collected more irregularly and infrequently than angle and pressure data. Gender and other biometric data such as age, height, and weight are not included in the patient description or the analysis.

Clinician Labeling

To train a classifier for determining stroke residual severity, our dataset must have appropriate labels corresponding to the patient's level of function around the time the data were collected. During a series of video calls and using the Motus

Hand and Motus Foot technology, a clinician met with each individual and performed a series of assessments. Remote assessment of extremity function using an external device has been studied and indicates that it is noninferior to in-person assessment when done properly [34]. The clinician used the potentiometer [35] embedded in the wrist or ankle joint of the Motus Hand or Motus Foot and a “clinician dashboard” interface to read live angle and pressure data from the patient and provide them the requisite assistance to stretch the patient’s wrist or ankle to collect passive and active ROM thresholds. With these

assessments, the clinician estimated each individual’s active ROM and passive ROM and characterized their level of function as “no ROM,” “low ROM,” or “high ROM.” While these labels are quite broad, the labeling process is hardly a simple algorithm. At the clinician’s discretion, quantitative and qualitative factors must apply an appropriate label. In particular, the clinician recorded the minimum or maximum angle reached for the assessment performed, the type of assessment performed, and a label summarizing the patient’s ROM level. These data are summarized in Table 1.

Table 1. Example patient label table assessed by a clinician during telehealth session.

Patient ID	Maximum angle	Minimum angle	Assessment	Clinician classification
1495	37	-20	Passive	No
2273	21	-16	Passive	No
2085	40	-15	Passive	No
2098	44	-9	Passive	No
1864	28	-12	Passive	No
2040	37	-18	Passive	No
2097	43	-18	Passive	Low
2356	-3	-17	Assisted	Low
2356	-3	-17	Assisted	Low
1688	52	-23	Assisted	Low
1876	54	-12	Passive	Low
2029	46	-20	Passive	Low
1458	30	-18	Passive	Low
1637	10	-12	Assisted	High
2282	8	-16	Assisted	High
1781	39	-15	Assisted	High
2360	10	-18	Assisted	High

All patients first were given a passive ROM assessment, in which they were stretched as far as their wrist or ankle would allow without experiencing pain or discomfort. Next, an active ROM assessment was conducted. In this assessment, the patient bends their wrist or ankle as far up and down as they can without any assistance from the Motus Hand or Motus Foot and without compensating with other parts of their bodies (hips, shoulders, etc). Depending on the patient’s assessed active ROM, an assisted ROM assessment was performed. This assessment consists of providing patients with varying amounts of assistance and recording their ROM in the presence of an upward force.

We define a patient as “assisted” or “passive” based on the most arduous assessment performed on the patient. The low ROM label contains a combination of patients who either did or did not have enough movement for the assisted ROM assessment. All patients who are classified with a high ROM (low residual stroke severity) were able to complete the assisted ROM assessment. This is important when noticing that patients with ID 2085 and 1781 (blue) have a similar total ROM (maximum angle–minimum angle), but patient ID 1781 requires clinician assistance to reach their maximum ROM. However, there is

ambiguity in some labels. For example, take patient ID 2356 (red), where it can be argued that the patient should have a high stroke residual severity (corresponding to low or no ROM), given the low total ROM with assistance. This is where the clinician has other outside factors that contribute to the final labeled classification of a patient. The clinician is visually able to assess the level of tone and spasticity that a patient may be exhibiting, which would not necessarily be captured in the minimum and maximum ROM values. The assessment results and labels were reviewed and confirmed by an additional expert.

Data Processing

To create a more manageable dataset for the labeling task, we generate summary statistics of the high-resolution data for each activity performed during a therapy session. First, to compensate for sensor reading issues, we smooth outliers out of the raw time series data (replacing data points in the 99th and 1st percentile with the value of the 99th and 1st percentile, respectively). Then, summarize the angle (relative to a reference midpoint in degrees) and pressure (in PSI) using the following variables: R_{min} , the minimum ROM for a game; R_{max} , the maximum ROM for a game; R_{mean} , the mean ROM for a game;

P_{min} , the minimum pressure for a game; F_{flex} , the maximum centripetal force generated while moving downward (flexion for upper extremities and plantar flexion for lower extremities); F_{ext} , the maximum centripetal force generated while moving upward (extension for upper extremities and dorsiflexion for lower extremities); P_{max} , the maximum pressure for a game; and P_{mean} , the mean pressure for a game. We finally pair these game-level summary statistics with the number of movements performed in the game (N_{mov}), the maximum score in the game (*Score*), and the total time spent playing that game during a therapy session (t_{game}).

This transformation from high-resolution data to game-level summary statistics provides a much more manageable dataset to which we can apply the clinician labels. A patient with low ROM (as labeled by the clinician) has little ROM during each game throughout a session. Using this idea, we construct a new dataset from each activity (game) a patient takes part in during a session, with each row having a unique (patient ID, session ID, and game ID) tuple. It is worth noting that a patient is unlikely to take part in every activity over the course of a therapy session; often, they gravitated to a few choice activities during each session. A summary of the data in each row in the described dataset is presented in Table 2.

After combining the data into this standardized dataset, the data then require sanitization, analysis, and normalization. The “role”

column indicates whether a variable is part of the feature set, the labels, or not used in the model training at all. To sanitize the data, we fill in missing values, correct invalid sensor values, and throw out data that did not represent a meaningful therapeutic exercise.

To isolate games with insufficient activity to draw meaningful conclusions, we restrict the number of movements, N_{mov} , performed during a game (therapeutic activity). A “movement” is any change of direction recorded in the angle sensor after noise is smoothed out of the time series. We remove any game (therapeutic activity) with fewer than 3 movements, as no significant therapeutic exercise can be performed with fewer than 3 movements (under assistance from the robotic Motus Hand or Foot).

Before performing any data analysis, the harmonized dataset is partitioned using an 80%:20% split into a training set and a holdout test set. The training set is used for exploratory data analysis and model training. All normalization and transformation techniques derived from the training set are then applied to the test set before the final predictive measures are computed. This is done to prevent data leakage from including the test set in the derivation of normalization and transformation techniques. The test dataset is reserved for the final performance measures in the Results section.

Table 2. Session game data dictionary.

Variable	Description	Role	Unit	Example
F_{flex}	Maximum centripetal force generated moving in the downward direction during an activity (computed from derivatives of angle data)	Feature	Newton	-3.047709105
F_{ext}	Maximum centripetal force generated moving in the upward direction during an activity (computed from derivatives of angle data)	Feature	Newton	3.251405759
N_{mov}	The number of completed movements during an activity	Feature	Integer	10
R_{min}	Absolute minimum angle detected by angle sensor during an activity	Feature	Degrees	-25
R_{max}	Absolute maximum angle detected during an activity	Feature	Degrees	46.41941
t_{game}	Total time spent performing therapy during an activity	Feature	Seconds	15
P_{min}	Minimum pressure applied by the sensor during an activity	Feature	PSI	-0.04511994
P_{max}	Maximum pressure applied by the sensor during an activity	Feature	PSI	10.30989
P_{mean}	Average pressure applied by the sensor during an activity	Feature	PSI	3.590553432
<i>Score</i>	Maximum score achieved during an activity	Feature	Integer	100
<i>h</i>	Peripheral type variable indicating the hand or foot	Feature	0, 1	Hand
<i>Class</i>	Designate stroke severity label by a clinician (high, low, and no)	Label	0, 1, 2	High
<i>g</i>	Unique identifier for each game (therapeutic activity) that is available on the Motus Hand or Foot	Not used	Integer	4
<i>p</i>	Anonymous identifier for each patient using the Motus Hand or Foot in this study	Not used	Integer	11
<i>s</i>	Unique identifier for each session performed on the Motus Hand or Foot	Not used	Integer	782,302,348,734

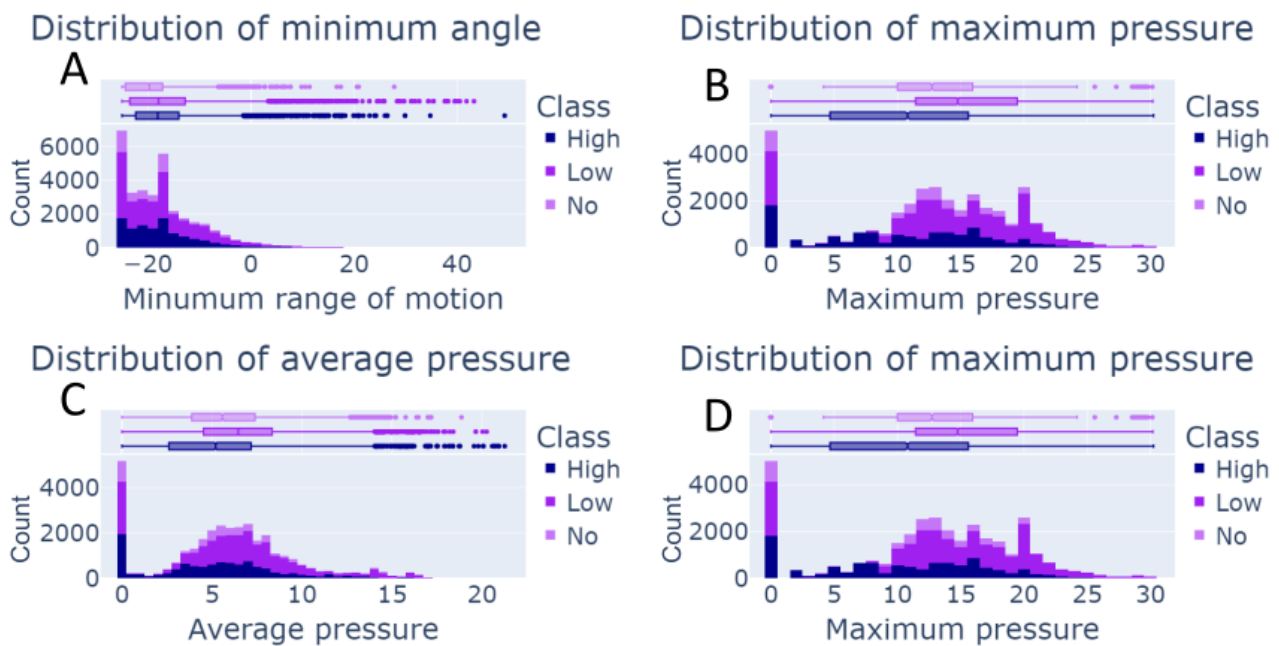
Exploratory Data Analysis

Data Distribution

It is well-known that proper data normalization is critical for maximizing model performance across machine learning applications and methods [36]. Knowing the proper

normalization technique for each feature requires a cursory dataset analysis. In Figure 3, we show representative distributions of the features that will be input variables for our comparative model analysis. While some variables are not normally distributed, assuming that the data are normally distributed is sufficient considering the results [37].

Figure 3. Example of the distribution plots for 4 of the random variables for each therapeutic activity color-coded by the class label from the clinician. The plots are (A) distribution of minimum angle, (B) distribution of maximum pressure, (C) distribution of average pressure, and (D) distribution of maximum pressure. An explanation for each variable in the dataset is given in Table 2. Subsequently, random variables are distributed normally, which is crucial for using the z score when inputting into a machine learning algorithm.



Correlation

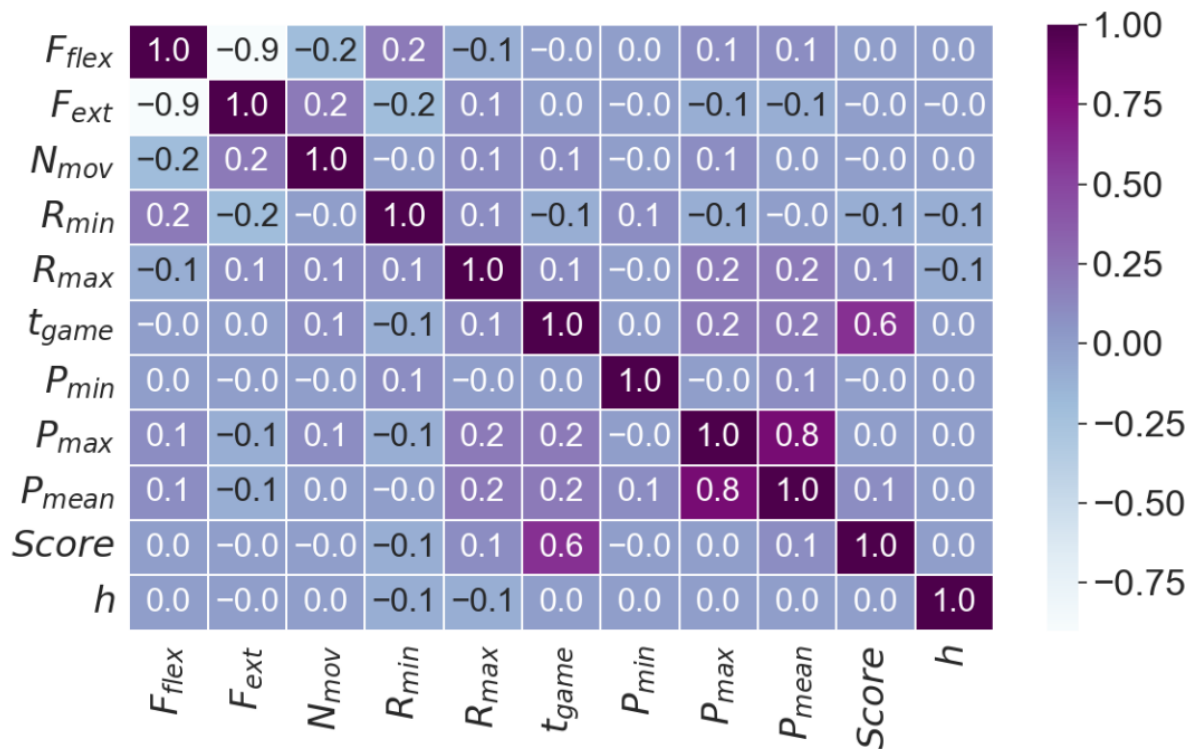
We analyze the correlation among the features in our dataset to identify potential redundancies. Then, we look at the principal component decomposition [38] to see if the variation in the data can be meaningfully reduced to a lower dimensional space. The correlation matrix for the feature set, constructed by computing the correlation between each pair of features in the dataset, is shown in Figure 4.

Because a correlation matrix points to potential relationships between features, it can indicate the feasibility of dimensionality reduction when preparing a dataset for building a classifier. If 2 variables are highly correlated, that is, $|Cor(X, Y)| > 0.9$, Shin and Park [39] suggest that one of those variables can be dropped from the analysis. We use this threshold of 0.9 where appropriate.

There exists a strong negative correlation between F_{ext} and F_{flex} approximately at -0.9 . However, we chose not to exclude either variable from our analysis due to their relevance in neuromotor recovery. For survivors who have experienced a stroke with upper extremity impairment, hypertonia often results in distinct patterns of volitional flexion (downward pushing force) and extension (upward pushing force) improvement [40].

The correlation between the mean pressure for an activity, P_{mean} , and maximum pressure for an activity, P_{max} , with the value of 0.80, indicates that the Motus Hand or Foot applied more pressure on average in each activity; however, because this correlation fails to surpass the threshold of 0.90, we do not drop either variable. Similarly, the correlation (0.60) between game time, t_{game} , and game score, $Score$, is intuitive: the longer a patient plays a game, the higher their score. Unfortunately, this correlation also does not meet the threshold for exclusion in the final feature set.

Figure 4. Correlation matrix for the feature set, excluding identification variables such as patient ID, session ID, game ID, and start time. A correlation threshold of $|Cor(X, Y)| > 0.9$ was used for variable exclusion in dimensionality reduction. Notably, F_{flex} and F_{ext} show a strong negative correlation. Despite this, both were included in the analysis because the development of downward pushing strength (flexion in upper extremities and plantar flexion in lower extremities) does not imply the development of upward pushing strength (extension in upper extremities and dorsiflexion in lower extremities).



Dimensionality Reduction

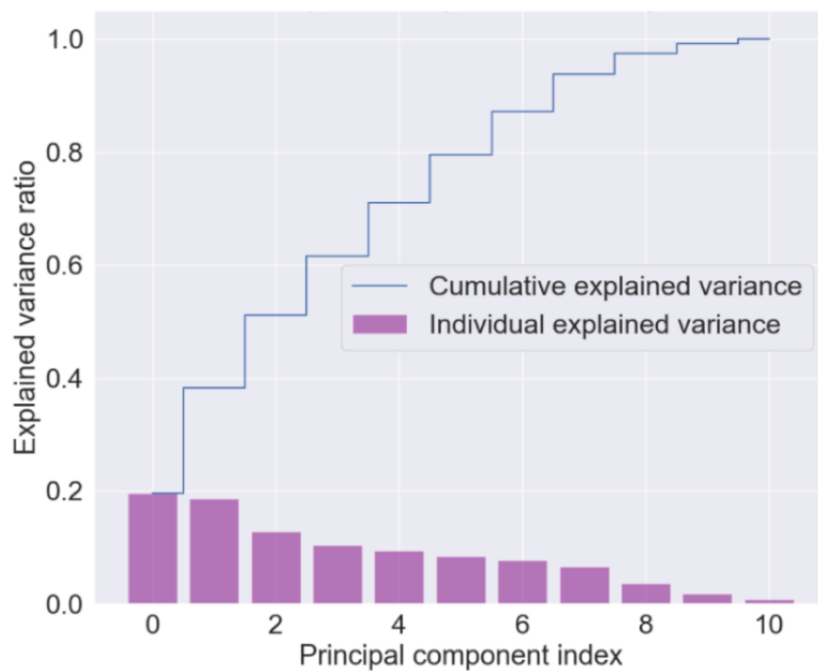
Another informative approach for analyzing the potential for dimensionality reduction in a feature set is PCA. Principal components are new variables constructed as linear combinations of the initial variables. These linear combinations ensure that the new variables (ie, principal components) are uncorrelated and that as few components as possible contain most of the information from the initial variables. Explained variance is a statistical measure of how much variation in a dataset is attributable to each principal component (eigenvectors) generated by the PCA method [41]. Explained variance thus

allows us to rank the components in order of importance and to focus on the most important ones when interpreting the results of our analysis.

In Figure 5, we show the explained variance of each principal component contributes to the total variation in the feature set. No component can be described as dominant, as none accounts for more than 20% of the variance in the initial dataset.

Given this and the results from our correlation analysis, no variables present in the principal dataset were excluded from the feature set.

Figure 5. Depiction of the principal components with the explained variance ratio. As shown, 95% of the explained variance is contributed by all principal components. As a result, all variables are used in the machine learning model for the analysis.



Model Description

Here, we provide a brief overview of the models compared in the *Results* section. LR is a classical statistical technique for binary classification. The technique consists of mapping the probability of an event happening to a logistic curve with the model inputs as dependent variables. LR is still widely used and is a common first model when performing classification because it is easy to implement and interpret.

Gradient boosting decision tree (GBDT) is a widely used machine learning algorithm due to its efficiency, accuracy, and interpretability [27]. The algorithm uses smaller “weaker classifiers” with a number of leaves. By taking a weighted average of these several “weaker classifiers,” we then can construct a “stronger classifier” [42]. By training several weaker models, this process is known as AdaBoosting. It results in a stronger model by adding more leaves to the decision tree and taking a weighted combination of these weaker models, where the weights are determined by the performance [43].

The DNN is a high-performance deep learning model with varying hidden layers. Several architectures were tested on the training dataset to see if there was an increase in performance by adding hidden layers (from 4 to 8) or a reduction in nodes in each input layer [44]. The rectified linear unit activation function was implemented into the model instead of the sigmoid function. Both were tried. Accuracy results from the computational experiment could surpass 80%, regardless of adding more layers, changing the hidden layer input size, or changing the activation function. The best-performing DNN trained in our analysis has 3 hidden layers with the input size of the hidden layers as 8, 5, and 8, respectively. Layer size, learning rates, batch size, and epoch size were all hyperparameters tuned during the training process.

The extra trees classifier is an ensemble learning method for classification. Ensemble learning is a machine learning technique that combines the predictions of multiple individual models to produce a more accurate and robust final prediction. The basic idea is to train multiple models independently, each with a different algorithm or set of hyperparameters, and then combine their predictions at the end [45]. This is similar to the AdaBoosting concept with LGB, where models can be combined by averaging or weighting their predictions [46]. The model uses entropy as the splitting criterion for the trees, with 100% of the features considered at each split. The maximum number of leaf nodes for each tree is 8717, and the model is comprised of 42 trees [46].

Ethical Considerations

This study was approved by the institutional review board of Georgia State University (IRB H24270). This research involves the analysis of preexisting, nonidentifiable data. No direct interaction or intervention with human participants occurred during the course of this study. The study relies solely on data generated by the commercial company Motus Nova for nonresearch purposes. All methods followed relevant guidelines and regulations approved by the Georgia State University institutional review board that waived the informed consent and designated this study as no human participants research.

Results

The original harmonized dataset (described in Table 2) contained all the scores, the minimum and maximum ROM, and minimum and maximum pressure, and we took the maximum score per game per session. Due to the smaller dataset, the training and testing were split on the 80%:20% principle, where for 80% of the data, the training set was used to train the models (with a subset of this set being used for training hyperparameters where

appropriate). For the remaining 20%, the holdout test set was used to compare the performance of the models after training.

Table 3 shows a performance based on 10-fold cross-validation for each machine learning classification algorithm applied to the training set. K-fold cross-validation is used to verify that a high-accuracy model does not necessarily overfit the training data. The training dataset is randomly divided into 10 different subsets or “folds” [47]. Each of these folds is then used as the new training data, while another is used as the new testing data for fitting a new model. We then take the mean and SD of the model accuracy across the 10 folds.

From Table 3, it is clear that both the neural network and LR perform poorly compared to the tree-based methods (extra trees classifier and LGB). The poor performance for LR is likely due to the assumption that there is a linear relationship between the features and the labels, that is, the points corresponding to each label can be nicely separated by a hyperplane (the N-dimensional extension of a line in 2D or plane in 3D). On the other hand, neural networks tend to perform poorly on small datasets like the therapy dataset we have compiled. This is because, while able to capture nonlinear decision boundaries, neural networks are prone to overfitting the training dataset. Tree-based methods provide an excellent combination of low bias but are still able to capture a nonlinear decision boundary.

Figure 6 presents the confusion matrix of each of the supervised learning methods. A confusion matrix is used to represent the algorithm’s performance visually. Each row of the matrix represents the instances in an actual class, while each column represents the instances in a predicted class or vice versa. We represent the percentage over the exact numeric number for display purposes. Three performance metrics come from the confusion matrix: precision, recall, and the F_1 -score. Accuracy measures the proportion of predicted positives that are truly positive. Recall measures the proportion of predicted negatives that are truly negative. To compare the performance of each model, we use the F_1 -score. The F_1 -score is the harmonic mean of the precision and recall [31]. In this case, this is macroaveraging (treating all classes equally important).

A full breakdown of the performance measures (precision, recall, and F_1 -score) for all models on the holdout test set is shown in Table 4. It is important to notice that while the extra trees classifier has a comparable accuracy (picking the correct label) with the LGB method, LGB performs reliably better than all of the other models when also weighing false positives and false negatives (precision, recall, and F_1 -score). Remarkably, the LGB model best fits the dataset with a weighted average F_1 -score of 96.70% compared to LR (55.82%), extra trees classifier (94.81%), and DNN (70.11%).

Table 3. Training set cross-validation algorithm accuracy.

Algorithm	Accuracy (%), mean (SD)
Extra trees classifier	96.40 (0.4)
Light gradient boosting	94 (0.4)
Neural network	71.70 (0.7)
Logistic regression	61.20 (0.5)

Figure 6. Confusion matrices for (A) light gradient boosting (LGB) and (B) logistic regression. Considering the false negative column of the no classification, it is seen that the LGB model greatly improved this classification. This is especially important when classifying a patient as having “no” stroke severity when they are actually a “high” severity. Misclassifications can be particularly dangerous, ranging from providing inadequate therapy to a patient with high ROM to injuring a user with low ROM with therapy designed for a user with high ROM. ROM: range of motion.

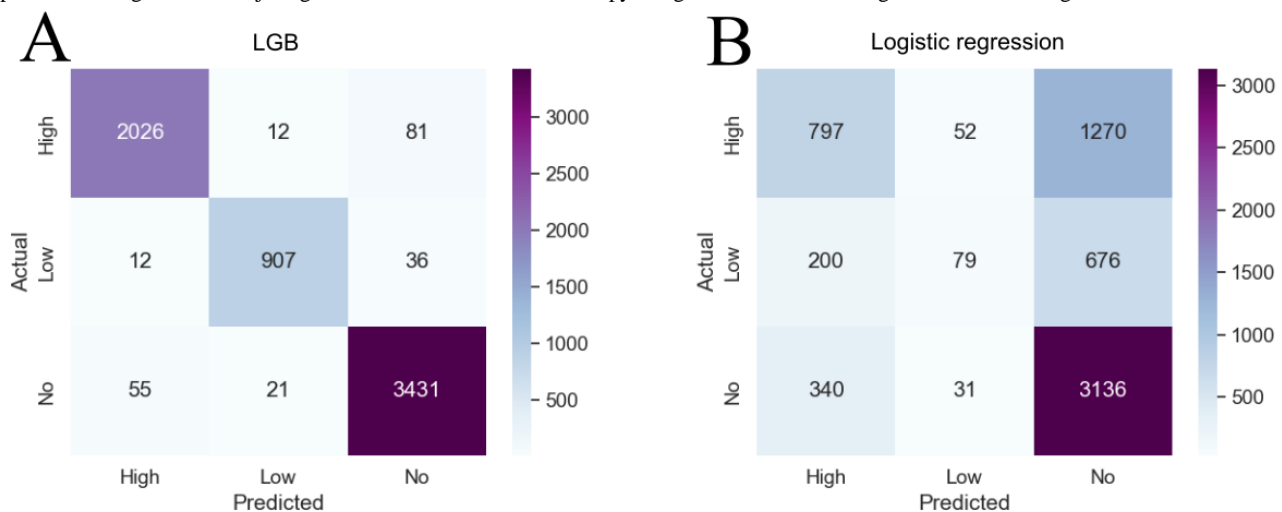


Table 4. Performance measures on the holdout test dataset for each model for each label type.

Model and label	Performance metric		
	Precision (%)	Recall (%)	F_1 -score (%)
Extra trees classifier			
Low	95.44	92.83	94.12
High	94.46	91.10	92.75
No	94.55	97.04	95.78
Weighted average	94.82	94.84	94.81
Light gradient boosting			
Low	96.80	95.61	96.20
High	96.49	94.97	95.73
No	96.70	97.83	97.26
Weighted average	96.70	96.70	96.70
Deep feed-forward neural network			
Low	74.34	64.94	69.32
High	61.93	28.27	38.82
No	71.64	87.71	78.86
Weighted average	71.14	71.94	70.11
Logistic regression			
Low	59.61	37.61	46.12
High	48.77	8.27	14.15
No	61.71	89.42	73.02
Weighted average	59.15	60.96	55.82

Discussion

Principal Findings

We have demonstrated that objectively measured rehabilitation training combined with machine learning methods can be effectively used to identify residual stroke severity classes. This approach aims to enhance in-home, self-guided, individualized stroke rehabilitation. We have tackled several challenges commonly faced in health care applications of machine learning, such as processing data with varying physical quantities, handling errors in sensory data, and addressing ambiguous classifications due to human error.

Comparison to Prior Work

Previous studies have largely focused on predicting short- and long-term functional ability based on clinical variables from an inpatient hospital stay immediately after stroke [16-18] or they have used robotic measurements to predict clinical measurement scores [48,49]. Meanwhile, our study focuses on data collected during rehabilitation in the in-home setting that is used to predict residual stroke severity.

In those previous studies, the algorithms most frequently used were linear and LR. However, these methods showed poor accuracy (less than 80%) with our dataset, prompting us to explore different approaches. We found that the LGB method provided substantially higher accuracy, despite being applied

to a relatively small dataset. LGB is advantageous for real-time autonomous stroke residual severity classification; it is known to be easily transferable and requires relatively little computational resources [27].

Strengths and Limitations

Our study design ensures that the model can make decisions based on summary statistics typically available in an outpatient rehabilitation setting. This design enables retraining of the model in an outpatient environment, similar to an in-home setting. Consequently, the model can offer a second opinion on a patient's stroke residual severity or potentially replace a clinician's assessment. This capability allows for a more targeted therapy routine based on the stroke residual severity classification.

A notable limitation is that the model may need retraining to accommodate specific data collected in each outpatient setting, accounting for differences in the data. Such data collection can be challenging and costly, particularly for outpatient facilities with low patient volumes.

Future Directions

Future work involves building an expanded and more sophisticated dataset. Real-time processing of sensor data will enable a classifier to interact with users in real time, recognizing and classifying subtle changes in their motor function. This

capability will allow clinicians (AI or otherwise) to prescribe personalized, targeted interventions that are most impactful.

Additionally, integrating real-time understanding of a patient's needs with an in-home robotic therapy device like the Motus Hand and Motus Foot will provide immediate feedback. An AI in therapeutic games can detect patient needs, such as fatigue, during a therapy session and adapt its strategy accordingly. Further research could also explore finer-grained severity classifications, such as labeling patients based on their total ROM or amount of tone, which would require more labeled data to train the machine learning model properly.

Conclusions

Autonomous classification is becoming more important for successful rehabilitation, as rehabilitation begins to move out of the clinical setting. Still, it faces challenges with the accessibility and volume of appropriate clinical data for training models and model access to user data for classification.

By leveraging the in-home stroke rehabilitation robotics provided by the Motus Hand and Motus Foot, we have made significant progress in addressing these issues that prevent

adequate training of an autonomous classification model. With the data collected from self-guided, in-home therapy sessions, we could train a classification model to identify the stroke residual severity in 33 patients. We compared 4 different models: extra trees classifier, LGB, DNN, and LR, finding the LGB method to outscore the other 3 with an average F_1 -score of 94%. The LGB method is a particularly powerful model for this case because it combines interpretability and portability.

Because our model relies only on therapy session summary statistics, the proposed method is expected to be successful when applied to comparable rehabilitation datasets. Once trained, the model is highly portable and can be integrated into similar rehabilitation settings, such as outpatient rehabilitation facilities with appropriate technological resources, to provide an autonomous real-time classification of stroke residual severity. Additionally, when paired with something like the Motus Hand and Motus Foot technology, our classifier provides the opportunity to develop personalized training based on the stroke residual severity of the individual and adapt the therapy exercises to each patient's needs. The efficacy of real-time classification and adaptation remains a subject of future study.

Acknowledgments

The authors would like to thank Motus Nova for providing the raw sensor data used in this study. This work was supported by the National Science Foundation (United States; grant CMMI-1953135).

Data Availability

The datasets generated and analyzed during this study are not publicly available because they are owned by Motus Nova, LLC, but are available from the corresponding author on reasonable request.

Authors' Contributions

RJ and IB designed and conceptualized the study and contributed to project administration. IB curated the data. RG and RJ contributed to the methodology, investigation, and formal analysis. All authors interpreted the results and drafted and edited the manuscript.

Conflicts of Interest

The authors acknowledge no conflicts of interest. Although the access to in-home stroke rehabilitation data, collected using the Motus Hand and Motus Foot devices and provided by Motus Nova, was crucial for the conception of this study, the company did not influence the methods, results, or discussions presented herein. All analyses and interpretations were conducted independently by the research team to ensure unbiased and objective findings.

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet Neurol* 2021;20(10):795-820 [FREE Full text] [doi: [10.1016/S1474-4422\(21\)00252-0](https://doi.org/10.1016/S1474-4422(21)00252-0)] [Medline: [34487721](https://pubmed.ncbi.nlm.nih.gov/34487721/)]
2. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019;18(5):439-458 [FREE Full text] [doi: [10.1016/S1474-4422\(19\)30034-1](https://doi.org/10.1016/S1474-4422(19)30034-1)] [Medline: [30871944](https://pubmed.ncbi.nlm.nih.gov/30871944/)]
3. Epalte K, Grjadovoj A, Bērziņa G. Use of the digital assistant vigo in the home environment for stroke recovery: focus group discussion with specialists working in neurorehabilitation. *JMIR Rehabil Assist Technol* 2023;10:e44285 [FREE Full text] [doi: [10.2196/44285](https://doi.org/10.2196/44285)] [Medline: [37058334](https://pubmed.ncbi.nlm.nih.gov/37058334/)]
4. Arntz A, Weber F, Handgraaf M, Lällä K, Korniloff K, Murtonen K, et al. Technologies in home-based digital rehabilitation: scoping review. *JMIR Rehabil Assist Technol* 2023;10:e43615. [doi: [10.2196/43615](https://doi.org/10.2196/43615)] [Medline: [37253381](https://pubmed.ncbi.nlm.nih.gov/37253381/)]

5. Broderick M, O'Shea R, Burridge J, Demain S, Johnson L, Bentley P. Examining usability, acceptability, and adoption of a self-directed, technology-based intervention for upper limb rehabilitation after stroke: cohort study. *JMIR Rehabil Assist Technol* 2023;10:e45993 [FREE Full text] [doi: [10.2196/45993](https://doi.org/10.2196/45993)] [Medline: [37603405](https://pubmed.ncbi.nlm.nih.gov/37603405/)]
6. Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemati S, et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Crit Care Med* 2020;48(2):210-217 [FREE Full text] [doi: [10.1097/CCM.0000000000004145](https://doi.org/10.1097/CCM.0000000000004145)] [Medline: [31939789](https://pubmed.ncbi.nlm.nih.gov/31939789/)]
7. Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *Int J Med Inform* 2020;136:104068. [doi: [10.1016/j.ijmedinf.2019.104068](https://doi.org/10.1016/j.ijmedinf.2019.104068)] [Medline: [31923822](https://pubmed.ncbi.nlm.nih.gov/31923822/)]
8. Cerasa A, Tartarisco G, Bruschetta R, Ciancarelli I, Morone G, Calabrò RS, et al. Predicting outcome in patients with brain injury: differences between machine learning versus conventional statistics. *Biomedicines* 2022;10(9):2267 [FREE Full text] [doi: [10.3390/biomedicines10092267](https://doi.org/10.3390/biomedicines10092267)] [Medline: [36140369](https://pubmed.ncbi.nlm.nih.gov/36140369/)]
9. Harari Y, O'Brien MK, Lieber RL, Jayaraman A. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. *J Neuroeng Rehabil* 2020;17(1):71 [FREE Full text] [doi: [10.1186/s12984-020-00704-3](https://doi.org/10.1186/s12984-020-00704-3)] [Medline: [32522242](https://pubmed.ncbi.nlm.nih.gov/32522242/)]
10. Campagnini S, Arienti C, Patrini M, Liuzzi P, Mannini A, Carrozza MC. Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review. *J Neuroeng Rehabil* 2022;19(1):54 [FREE Full text] [doi: [10.1186/s12984-022-01032-4](https://doi.org/10.1186/s12984-022-01032-4)] [Medline: [35659246](https://pubmed.ncbi.nlm.nih.gov/35659246/)]
11. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke* 2019;50(5):1263-1265. [doi: [10.1161/STROKEAHA.118.024293](https://doi.org/10.1161/STROKEAHA.118.024293)] [Medline: [30890116](https://pubmed.ncbi.nlm.nih.gov/30890116/)]
12. Saber H, Somai M, Rajah GB, Scalzo F, Liebeskind DS. Predictive analytics and machine learning in stroke and neurovascular medicine. *Neurol Res* 2019;41(8):681-690. [doi: [10.1080/01616412.2019.1609159](https://doi.org/10.1080/01616412.2019.1609159)] [Medline: [31038007](https://pubmed.ncbi.nlm.nih.gov/31038007/)]
13. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS One* 2020;15(6):e0234722 [FREE Full text] [doi: [10.1371/journal.pone.0234722](https://doi.org/10.1371/journal.pone.0234722)] [Medline: [32530947](https://pubmed.ncbi.nlm.nih.gov/32530947/)]
14. Fernandez-Lozano C, Hervella P, Mato-Abad V, Rodríguez-Yáñez M, Suárez-Garaboa S, López-Dequidt I, et al. Random forest-based prediction of stroke outcome. *Sci Rep* 2021;11(1):10071 [FREE Full text] [doi: [10.1038/s41598-021-89434-7](https://doi.org/10.1038/s41598-021-89434-7)] [Medline: [33980906](https://pubmed.ncbi.nlm.nih.gov/33980906/)]
15. Someeh N, Mirfeizi M, Asghari-Jafarabadi M, Alinia S, Farzipoor F, Shamshirgaran SM. Predicting mortality in brain stroke patients using neural networks: outcomes analysis in a longitudinal study. *Sci Rep* 2023;13(1):18530 [FREE Full text] [doi: [10.1038/s41598-023-45877-8](https://doi.org/10.1038/s41598-023-45877-8)] [Medline: [37898678](https://pubmed.ncbi.nlm.nih.gov/37898678/)]
16. Lin W, Chen C, Tseng Y, Tsai Y, Chang C, Wang H, et al. Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. *Int J Med Inform* 2018;111:159-164. [doi: [10.1016/j.ijmedinf.2018.01.002](https://doi.org/10.1016/j.ijmedinf.2018.01.002)] [Medline: [29425627](https://pubmed.ncbi.nlm.nih.gov/29425627/)]
17. Gupta VP, Garton AL, Sisti JA, Christophe BR, Lord AS, Lewis AK, et al. Prognosticating functional outcome after intracerebral hemorrhage: the ICHOP score. *World Neurosurg* 2017;101:577-583 [FREE Full text] [doi: [10.1016/j.wneu.2017.02.082](https://doi.org/10.1016/j.wneu.2017.02.082)] [Medline: [28242488](https://pubmed.ncbi.nlm.nih.gov/28242488/)]
18. Kim JK, Choo YJ, Chang MC. Prediction of motor function in stroke patients using machine learning algorithm: development of practical models. *J Stroke Cerebrovasc Dis* 2021;30(8):105856. [doi: [10.1016/j.jstrokecerebrovasdis.2021.105856](https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105856)] [Medline: [34022582](https://pubmed.ncbi.nlm.nih.gov/34022582/)]
19. Butler A, Housley SN, Chen YA, Wolf S. Increasing access to cost effective home-based robotic telerehabilitation for stroke survivors. 2017 Presented at: 2017 International Symposium on Wearable Robotics and Rehabilitation (WeRob); November 5-8, 2017; Houston, TX URL: <https://ieeexplore.ieee.org/document/8383873>
20. Sawaki L, Butler AJ, Leng X, Wassenaar PA, Mohammad YM, Blanton S, et al. Constraint-induced movement therapy results in increased motor map area in subjects 3 to 9 months after stroke. *Neurorehabil Neural Repair* 2008;22(5):505-513 [FREE Full text] [doi: [10.1177/1545968308317531](https://doi.org/10.1177/1545968308317531)] [Medline: [18780885](https://pubmed.ncbi.nlm.nih.gov/18780885/)]
21. Wolf SL, Blanton S, Baer H, Breshears J, Butler AJ. Repetitive task practice: a critical review of constraint-induced movement therapy in stroke. *Neurologist* 2002;8(6):325-338 [FREE Full text] [doi: [10.1097/01.nrl.0000031014.85777.76](https://doi.org/10.1097/01.nrl.0000031014.85777.76)] [Medline: [12801434](https://pubmed.ncbi.nlm.nih.gov/12801434/)]
22. Butler AJ, Bay C, Wu D, Richards KM, Buchanan S, Yepes M. Expanding tele-rehabilitation of stroke through in-home robot-assisted therapy. *Int J Phys Med Rehabil* 2014;02(02):1-11. [doi: [10.4172/2329-9096.1000184](https://doi.org/10.4172/2329-9096.1000184)]
23. Wolf SL, Sahu K, Bay RC, Buchanan S, Reiss A, Linder S, et al. The HAAP (Home Arm Assistance Progression Initiative) trial: a novel robotics delivery approach in stroke rehabilitation. *Neurorehabil Neural Repair* 2015;29(10):958-968 [FREE Full text] [doi: [10.1177/1545968315575612](https://doi.org/10.1177/1545968315575612)] [Medline: [25782693](https://pubmed.ncbi.nlm.nih.gov/25782693/)]
24. Cai S, Li G, Zhang X, Huang S, Zheng H, Ma K, et al. Detecting compensatory movements of stroke survivors using pressure distribution data and machine learning algorithms. *J Neuroeng Rehabil* 2019;16(1):131 [FREE Full text] [doi: [10.1186/s12984-019-0609-6](https://doi.org/10.1186/s12984-019-0609-6)] [Medline: [31684970](https://pubmed.ncbi.nlm.nih.gov/31684970/)]

25. Thakkar HK, Liao WW, Wu CY, Hsieh YW, Lee TH. Predicting clinically significant motor function improvement after contemporary task-oriented interventions using machine learning approaches. *J Neuroeng Rehabil* 2020;17(1):131 [FREE Full text] [doi: [10.1186/s12984-020-00758-3](https://doi.org/10.1186/s12984-020-00758-3)] [Medline: [32993692](https://pubmed.ncbi.nlm.nih.gov/32993692/)]
26. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. ArXiv Preprint posted online on December 19, 2013. [doi: [10.48550/arXiv.1312.5602](https://doi.org/10.48550/arXiv.1312.5602)]
27. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. 2017 Presented at: Advances in Neural Information Processing Systems 30 (NIPS 2017); December 04, 2017; Red Hook, NY p. 3149-3157 URL: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
28. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn* 2006;63(1):3-42. [doi: [10.1007/s10994-006-6226-1](https://doi.org/10.1007/s10994-006-6226-1)]
29. Wang S, Cao G, Shang Z. Deep neural network classifier for multi-dimensional functional data. ArXiv. Preprint posted online on May 17, 2022 2022. [doi: [10.48550/arXiv.2205.08592](https://doi.org/10.48550/arXiv.2205.08592)]
30. Xie P, Kim JK, Zhou Y, Ho Q, Kumar A, Yu Y, et al. Distributed machine learning via sufficient factor broadcasting. ArXiv. Preprint posted online on September 7, 2015 2015. [doi: [10.48550/arXiv.1409.5705](https://doi.org/10.48550/arXiv.1409.5705)]
31. Düntsch I, Gediga G. Confusion matrices and rough set data analysis. *J Phys Conf Ser* 2019;1229(1):012055. [doi: [10.1088/1742-6596/1229/1/012055](https://doi.org/10.1088/1742-6596/1229/1/012055)]
32. Naqvi SNZ, Yfantidou S. Time series databases and InfluxDB. Universite Libre de Bruxelles. 2017. URL: https://cs.ulb.ac.be/public/media/teaching/influxdb_2017.pdf [accessed 2024-09-06]
33. Porciuncula F, Roto AV, Kumar D, Davis I, Roy S, Walsh CJ, et al. Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances. *PM R* 2018;10(9 Suppl 2):S220-S232 [FREE Full text] [doi: [10.1016/j.pmrj.2018.06.013](https://doi.org/10.1016/j.pmrj.2018.06.013)] [Medline: [30269807](https://pubmed.ncbi.nlm.nih.gov/30269807/)]
34. Gopal A, Hsu WY, Allen DD, Bove R. Remote assessments of hand function in neurological disorders: systematic review. *JMIR Rehabil Assist Technol* 2022;9(1):e33157 [FREE Full text] [doi: [10.2196/33157](https://doi.org/10.2196/33157)] [Medline: [35262502](https://pubmed.ncbi.nlm.nih.gov/35262502/)]
35. Maceira-Elvira P, Popa T, Schmid A, Hummel FC. Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment. *J Neuroeng Rehabil* 2019;16(1):142 [FREE Full text] [doi: [10.1186/s12984-019-0612-y](https://doi.org/10.1186/s12984-019-0612-y)] [Medline: [31744553](https://pubmed.ncbi.nlm.nih.gov/31744553/)]
36. Sola J, Sevilla J. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans Nucl Sci* 1997;44(3):1464-1468 [FREE Full text]
37. Goldstein ML, Morris SA, Yen GG. Problems with fitting to the power-law distribution. *Eur Phys J B* 2004;41(2):255-258. [doi: [10.1140/epjb/e2004-00316-5](https://doi.org/10.1140/epjb/e2004-00316-5)]
38. Bro R, Smilde AK. Principal component analysis. *Anal Methods* 2014;6(9):2812-2831. [doi: [10.1039/c3ay41907j](https://doi.org/10.1039/c3ay41907j)]
39. Shin YJ, Park CH. Analysis of correlation based dimension reduction methods. *Int J Appl Math Comput Sci* 2011;21(3):549-558 [FREE Full text] [doi: [10.2478/v10006-011-0043-9](https://doi.org/10.2478/v10006-011-0043-9)]
40. Doussoulin A, Bacco JL, Rivas C, Saiz JL. Association between postural patterns of spastic upper extremity and functional independence after TBI and stroke. *NeuroRehabilitation* 2020;46(4):551-559. [doi: [10.3233/NRE-203042](https://doi.org/10.3233/NRE-203042)] [Medline: [32508335](https://pubmed.ncbi.nlm.nih.gov/32508335/)]
41. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemom Intell Lab Syst* 1987;2(1-3):37-52. [doi: [10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)]
42. Schapire RE. The strength of weak learnability. *Mach Learn* 1990;5(2):197-227. [doi: [10.1007/bf00116037](https://doi.org/10.1007/bf00116037)]
43. Chengsheng T, Huacheng L, Bing X. AdaBoost typical Algorithm and its application research. *MATEC Web Conf* 2017;139(2):00222. [doi: [10.1051/mateconf/201713900222](https://doi.org/10.1051/mateconf/201713900222)]
44. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT, Alinejad-Rokny H. A review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues. *J Biomed Inform* 2021;113:103627 [FREE Full text] [doi: [10.1016/j.jbi.2020.103627](https://doi.org/10.1016/j.jbi.2020.103627)] [Medline: [33259944](https://pubmed.ncbi.nlm.nih.gov/33259944/)]
45. Ganaie M, Hu M, Malik A, Tanveer M, Suganthan P. Ensemble deep learning: a review. *Eng Appl Artif Intell* 2022;115:105151. [doi: [10.1016/j.engappai.2022.105151](https://doi.org/10.1016/j.engappai.2022.105151)]
46. Arya M, Sastry G, Motwani A, Kumar S, Zaguia A. A novel extra tree ensemble optimized DL framework (ETEODL) for early detection of diabetes. *Front Public Health* 2022;9:797877 [FREE Full text] [doi: [10.3389/fpubh.2021.797877](https://doi.org/10.3389/fpubh.2021.797877)] [Medline: [35242738](https://pubmed.ncbi.nlm.nih.gov/35242738/)]
47. Shao J. Linear model selection by cross-validation. *J Am Stat Assoc* 1993;88(422):486-494. [doi: [10.1080/01621459.1993.10476299](https://doi.org/10.1080/01621459.1993.10476299)]
48. Zariffa J, Myers M, Coahran M, Wang RH. Smallest real differences for robotic measures of upper extremity function after stroke: implications for tracking recovery. *J Rehabil Assist Technol Eng* 2018;5:2055668318788036 [FREE Full text] [doi: [10.1177/2055668318788036](https://doi.org/10.1177/2055668318788036)] [Medline: [31191947](https://pubmed.ncbi.nlm.nih.gov/31191947/)]
49. Mostafavi SM, Glasgow JI, Dukelow SP, Scott SH, Mousavi P. Prediction of stroke-related diagnostic and prognostic measures using robot-based evaluation. 2013 Presented at: 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR); June 24-26, 2013; Seattle, WA URL: http://vigir.missouri.edu/~gdesouza/Research/Conference_CDs/IEEE_ICORR_2013/contents/papers/249.pdf

Abbreviations

AI: artificial intelligence
DNN: deep feed-forward neural network
LGB: light gradient boosting
LR: logistic regression
PCA: principal component analysis
ROM: range of motion

Edited by S Rizvi, T Leung; submitted 01.02.24; peer-reviewed by A Hassan, R Bidkar, X Cheng; comments to author 25.04.24; revised version received 22.05.24; accepted 31.07.24; published 07.10.24.

Please cite as:

Jeter R, Greenfield R, Housley SN, Belykh I

Classifying Residual Stroke Severity Using Robotics-Assisted Stroke Rehabilitation: Machine Learning Approach

JMIR Biomed Eng 2024;9:e56980

URL: <https://biomedeng.jmir.org/2024/1/e56980>

doi: [10.2196/56980](https://doi.org/10.2196/56980)

PMID: [39374054](https://pubmed.ncbi.nlm.nih.gov/39374054/)

©Russell Jeter, Raymond Greenfield, Stephen N Housley, Igor Belykh. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 07.10.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

<https://www.jmirpublications.com/>