JMIR Biomedical Engineering

Engineering for health technologies, medical devices, and innovative medical treatments and procedures
Volume 10 (2025) ISSN 2561-3278 Editor in Chief: Javad Sarvestan, PhD

Contents

Corrigenda and Addenda

Reviews

Original Papers

Research Letter

Corrigenda and Addenda

Correction: Can Artificial Intelligence Diagnose Knee Osteoarthritis?

Mihir Tandon[1], BA; Nitin Chetla[2], BS; Adarsh Mallepally[3]; Botan Zebari[4], BS; Sai Samayamanthula[2], BA; Jonathan Silva[1], BS; Swapna Vaja[5], BS; John Chen[1], BS; Matthew Cullen[1], BS; Kunal Sukhija[6], MD

[1]Albany Medical College, Albany, NY, United States
[2]University of Virginia School of Medicine, Charlottesville, VA, United States
[3]School of Medicine, Virginia Commonwealth University, Richmond, VA, United States
[4]St. James School of Medicine, Binghamton, NY, United States
[5]Rush Medical College, Chicago, IL, United States
[6]Kaweah Health, Visalia, CA, United States

Corresponding Author:
Mihir Tandon, BA
Albany Medical College
43 New Scotland Ave
Albany, NY, 12208
United States
Phone: 1 3322488708
Email: tandonm@amc.edu

Related Article:

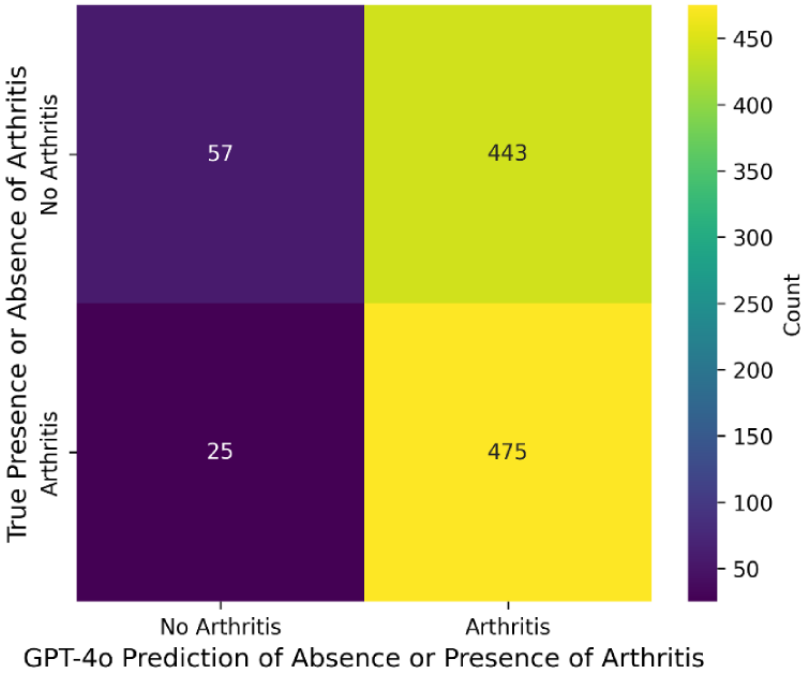
Correction of: https://biomedeng.jmir.org/2025/1/e67481

(JMIR Biomed Eng 2025;10:e82980) doi:10.2196/82980

In “Can Artificial Intelligence Diagnose Knee Osteoarthritis?” (JMIR Biomed Eng 2025;10:e67481), the authors made two corrections.

In the originally published version, Figure 1 displayed two of the Y-axis labels incorrectly. The label *“Arthritis”* was placed next to the row representing X-rays without arthritis, and the label *“No Arthritis”* was placed next to the row representing X-rays with arthritis.

Figure 1. Sensitivity and specificity of Chat-GPT4o in analyzing knee osteoarthritis X-rays.



The text in Figure 1 has been corrected so that the Y-axis labels align with the data:

- The top row is labeled *"No Arthritis"*, representing X-rays of knees without arthritis.
- The bottom row is labeled *"Arthritis"*, representing X-rays of knees with arthritis.

The X-axis label, *"GPT-4o Prediction of Absence or Presence of Arthritis"* and the Y-axis label, *"True Presence or Absence of Arthritis"*, have been reformatted to span the entire length of the figure rather than being stacked to improve both readability and overall appearance.

Additionally, an **Authors' Contributions** section has been added to the manuscript using the CREdiT taxonomy format:

Conceptualization: NC (lead), MT (equal), KS (equal)

Data curation: AM (lead), MT (equal), SS (supporting), SV (supporting), JC (supporting)

Formal analysis: JC (lead), JS (supporting), MC (supporting), SV (supporting), AM (supporting)

Funding acquisition: KS (lead)

Investigation: SS (lead), KS (equal), BZ (supporting), SV (supporting)

Methodology: MT (lead), NC (equal), KS (equal), AM (supporting)

Resources: SV (lead), JC (supporting)

Software: JC (lead), AM (supporting)

Supervision: KS (lead), MT (equal), NC (equal)

Validation: JS (lead), JC (equal), MC (equal)

Visualization: MT (lead), MC (equal), SS (supporting)

Writing – original draft: MT (lead), NC (equal), BZ (supporting), SS (supporting), AM (supporting)

Writing – review & editing: JS (lead), SV (equal), JC (equal), MC (supporting), KS (supporting)

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.08.25; this is a non–peer-reviewed article;accepted 28.08.25; published 12.09.25.

Please cite as:
Tandon M, Chetla N, Mallepally A, Zebari B, Samayamanthula S, Silva J, Vaja S, Chen J, Cullen M, Sukhija K
Correction: Can Artificial Intelligence Diagnose Knee Osteoarthritis?
JMIR Biomed Eng 2025;10:e82980
URL: https://biomedeng.jmir.org/2025/1/e82980
doi:10.2196/82980
PMID:

Advancing Brain-Computer Interface Closed-Loop Systems for Neurorehabilitation: Systematic Review of AI and Machine Learning Innovations in Biomedical Engineering

Christopher Williams[1], BS; Fahim Islam Anik[2], MS; Md Mehedi Hasan[3], MS; Juan Rodriguez-Cardenas[3], MS; Anushka Chowdhury[4], HD; Shirley Tian[3], PhD; Selena He[3], PhD; Nazmus Sakib[3], PhD

[1]Department of Computer Science, Troy University, Troy, AL, United States

[2]Department of Mechanical Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh

[3]College of Computing and Software Engineering, Kennesaw State University, Marietta, 1100 South Marietta Pkwy SE, Marietta, GA, United States

[4]Lambert High School, Suwanee, GA, United States

Corresponding Author:
Nazmus Sakib, PhD
College of Computing and Software Engineering, Kennesaw State University, Marietta, 1100 South Marietta Pkwy SE, Marietta, GA, United States

Abstract

Background: Brain-computer interface (BCI) closed-loop systems have emerged as a promising tool in health care and wellness monitoring, particularly in neurorehabilitation and cognitive assessment. With the increasing burden of neurological disorders, including Alzheimer disease and related dementias (AD/ADRD), there is a critical need for real-time, noninvasive monitoring technologies. BCIs enable direct communication between the brain and external devices, leveraging artificial intelligence (AI) and machine learning (ML) to interpret neural signals. However, challenges such as signal noise, data processing limitations, and privacy concerns hinder widespread implementation.

Objective: The primary objective of this study is to investigate the role of ML and AI in enhancing BCI closed-loop systems for health care applications. Specifically, we aim to analyze the methods and parameters used in these systems, assess the effectiveness of different AI and ML techniques, identify key challenges in their development and implementation, and propose a framework for using BCIs in the longitudinal monitoring of AD/ADRD patients. By addressing these aspects, this study seeks to provide a comprehensive overview of the potential and limitations of AI-driven BCIs in neurological health care.

Methods: A systematic literature review was conducted following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, focusing on studies published between 2019 and 2024. We sourced research articles from PubMed, IEEE, ACM, and Scopus using predefined keywords related to BCIs, AI, and AD/ADRD. A total of 220 papers were initially identified, with 18 meeting the final inclusion criteria. Data extraction followed a structured matrix approach, categorizing studies based on methods, ML algorithms, limitations, and proposed solutions. A comparative analysis was performed to synthesize key findings and trends in AI-enhanced BCI systems for neurorehabilitation and cognitive monitoring.

Results: The review identified several ML techniques, including transfer learning (TL), support vector machines (SVMs), and convolutional neural networks (CNNs), that enhance BCI closed-loop performance. These methods improve signal classification, feature extraction, and real-time adaptability, enabling accurate monitoring of cognitive states. However, challenges such as long calibration sessions, computational costs, data security risks, and variability in neural signals were also highlighted. To address these issues, emerging solutions such as improved sensor technology, efficient calibration protocols, and advanced AI-driven decoding models are being explored. In addition, BCIs show potential for real-time alert systems that support caregivers in managing AD/ADRD patients.

Conclusions: BCI closed-loop systems, when integrated with AI and ML, offer significant advancements in neurological health care, particularly in AD/ADRD monitoring and neurorehabilitation. Despite their potential, challenges related to data accuracy, security, and scalability must be addressed for widespread clinical adoption. Future research should focus on refining AI models, improving real-time data processing, and enhancing user accessibility. With continued advancements, AI-powered BCIs can revolutionize personalized health care by providing continuous, adaptive monitoring and intervention for patients with neurological disorders.

(JMIR Biomed Eng 2025;10:e72218) doi:10.2196/72218

KEYWORDS

Brain-Computer Interface; closed-loop systems; Artificial Intelligence; AI; Machine Learning; neurorehabilitation; cognitive monitoring; real-time monitoring; healthcare technology; biomedical signal processing; human-computer interaction; PRISMA

Introduction

The adoption of technology in health care and wellness monitoring has grown significantly in recent years [1,2]. As of 2024, more than 1.3 billion people worldwide relied on digital health tools such as fitness trackers, smartwatches, and virtual doctor consultations. In the United States alone, 43% of the population actively used health apps [3,4]. This surge in digital health adoption is further reflected in the health care IT market, which is projected to expand from US$360 billion in 2024 to over US$730 billion by 2029 [5]. A recent survey revealed that 80% of Americans own at least one such device, including blood pressure monitors (45%), electric toothbrushes (39%), and fitness trackers or pedometers (24%) [6]. These devices play a crucial role in early detection and management of health conditions; notably, 28% of users reported receiving alerts about potential health issues from their devices, leading to successful diagnoses after consulting with health care professionals [6].

As technological advancements continue to reshape health care, their role in the early detection and management of Alzheimer disease and related dementias (AD/ADRD) is becoming increasingly critical [7]. AD is known as a neurological disorder characterized by memory loss, cognitive decline, and impaired motor skills [8]. It damages brain cells responsible for important mental functions and enables the cells themselves to degenerate and die. The degeneration begins from cognitive impairments, with motor functions still intact. Gradually, over time, this progresses into neuronal degeneration in several areas of the brain, including the hippocampus and mediotemporal cortex [9]. The disease is most commonly found in older adult populations; the prevalence of all dementias is known to increase for people aged 60 - 90 years, making aging the biggest risk factor for AD [10]. While the disease is irreversible and has no cure, early detection and continuous monitoring can significantly improve patient outcomes. However, up to a third of dementia cases remain undiagnosed, and existing diagnostic methods are often slow and inaccurate [11]. The integration of technology—through wearable devices, advanced diagnostic tests, and AI-driven analysis—enables continuous monitoring and early identification of cognitive decline.

A promising innovation in this landscape is brain-computer interfaces (BCIs), which have the potential to revolutionize the diagnosis and management of neurodegenerative diseases like AD/ADRD [12,13]. BCIs have been the subject of significant research due to their correlation to decoding neural activity and use by people with disabilities. The BCI closed-loop system directly connects the human brain and the outside environment [14], allowing for direct communication between a person and a computer. It enables users the ability to operate external devices through their brain activity and translate brain signals, strictly produced by the central nervous system, into commands that carry out a desired action [15]. The “closed-loop” aspect allows for the use of real-time data to monitor and adjust updates based on the patient’s condition. In particular, BCI applications have been initially designed to help people with disabilities and enhance neuroplasticity, characterized as the capacity of the brain to change or adapt its morphology in response to experiences [16]. The system may also help in rehabilitation for people with strokes, head trauma, and other disorders [15]. Broadly, a BCI system consists of 4 standard, sequential components: signal acquisition, feature extraction, feature translation, and device output [15]. Within each component, there exist several methods and techniques that have been reviewed that effectively execute the goal of detecting and qualifying features of brain signals. There are many parameters that the BCI closed-loop system seeks to measure, with the intention of collecting large and diverse datasets; performance metrics heavily influence the quality of BCI research, which several methods of BCI closed-loop systems depend on.

BCIs facilitate direct communication between the brain and external devices, allowing real-time monitoring of neural activity and cognitive function. This technology is particularly valuable for detecting early neurophysiological changes that precede noticeable cognitive decline, offering a more objective and continuous assessment than traditional diagnostic methods [17,18]. By integrating BCIs with artificial intelligence (AI) and machine learning (ML), researchers can analyze brain signals to identify patterns associated with Alzheimer progression, potentially enabling earlier and more accurate diagnoses. Furthermore, BCIs hold promise for enhancing cognitive rehabilitation and assistive communication for patients in later stages of the disease. As the demand for advanced neurological monitoring grows, BCIs represent a critical step toward personalized and proactive dementia care, bridging the gap between early intervention and improved patient outcomes [19-21]. Therefore, in the context of neuroscience and AI, the BCI is a proposed solution for identifying and providing neurorehabilitation methods through decoding electroencephalogram (EEG) signals. This can prove to be of great significance for the detection and diagnosis of several neurological disorders, such as Alzheimer disease, through exploiting the use of neuron devices and stimulating biological sensory neurons [22]. The ultimate motivation is to integrate AI models and BCI systems in order to allow for personalized treatment plans and contribute greatly to breakthroughs in health care.

However, many limitations are associated with BCI-based closed-loop systems that can hinder the systems’ performance and efficacy. For instance, BCI applications must recalibrate the system in order to account for each user/participant due to the high variability in brain signals [23].

The model must be trained from scratch each time there is a new subject. This contributes to significant financial expenses. Furthermore, the limited size of datasets can lead to overfitting, which occurs when a model fits too closely to its training data rather than including new data [23]. When using an EEG to capture brain signals, several limitations exist with using the

method. EEG-based BCI systems measure the average activity of neurons with electrodes located on the surface of the brain [23]. These generally produce a low signal-to-noise ratio (SNR); a low SNR indicates that the signal is corrupted by noise and therefore makes it difficult to interpret brain signals. This review analyzes several solutions to these challenges with the use of machine learning algorithms and networks that can easily decode complex brain data. However, this field of research is not limited to current knowledge and there is still more to explore regarding the use of machine learning and deep learning in BCI closed-loop systems.

In the exploration of BCI systems and artificial intelligence algorithms, our research aims to address a range of critical questions and topics that are integral to advancing this field, as shown in Figure 1. By investigating the following research questions, we will gain a comprehensive understanding of the real-world applications of BCIs, uncovering insights that could lead to innovative opportunities and improvements in the monitoring of AD/ADRD patients.

RQ 1. What specific methods and parameters are used in the BCI closed-loop system?

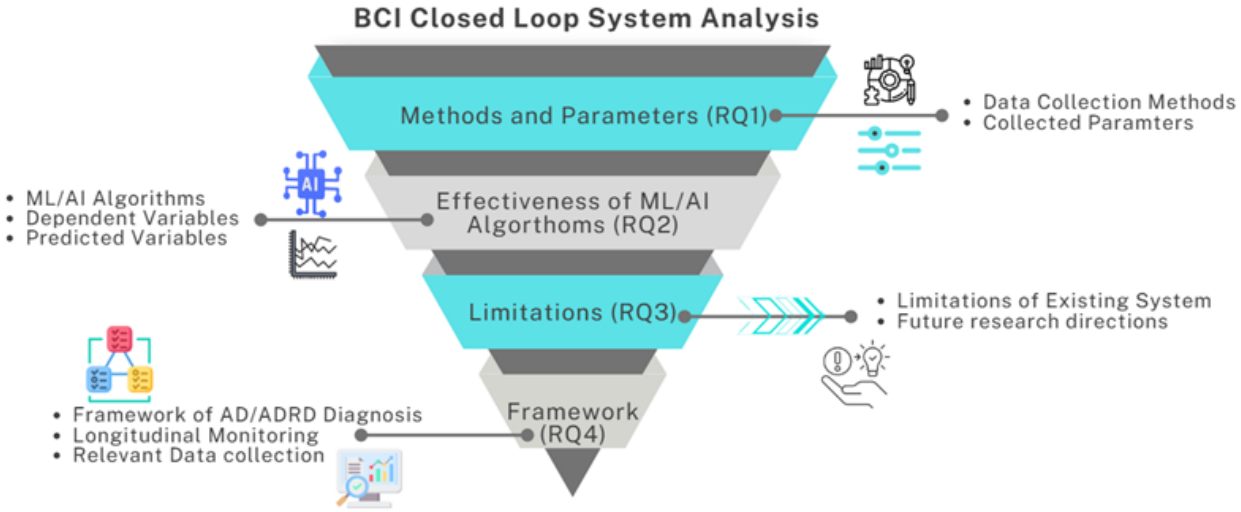
RQ 2. How effective are the different ML and AI algorithms used in the BCI closed-loop system?

RQ 3. How can we critically investigate the limitations in the development and implementation of the BCI closed-loop system?

RQ 4. How can we design a BCI closed-loop system-based framework for longitudinal monitoring of AD/ADRD patients?

The remainder of the paper is organized into 5 key sections. Section 2 outlines the methodology, comprising three subsections that detail the scoping criteria, literature search strategy, and data analysis procedures. Sections 3, 4, 5, and 6 address the 4 research questions in depth. Finally, Section 7 concludes the literature review, summarizing the key findings and their implications.

Figure 1. Brain-computer interface closed-loop systems overview in health care and wellness monitoring. AD/ADRD: Alzheimer disease/Alzheimer disease and related dementias; AI: artificial intelligence; BCI: brain-computer interface; ML: machine learning;



Methods

Overview

Our research approach centered on a comprehensive evaluation of the literature exploring the integration of AI—particularly its subset, ML—within BCI closed-loop systems in health care. The goal was to synthesize current knowledge on the methodologies, algorithms, outcomes, limitations, and emerging directions that define this interdisciplinary field. To achieve this, we developed a targeted search strategy using relevant keywords and Boolean operators, enabling us to identify both theoretical advancements and real-world applications of AI- and ML-enhanced BCIs. This method allowed for a focused analysis of how these technologies are transforming neurological monitoring, cognitive rehabilitation, and personalized patient care.

Scoping Criteria

Our scoping criteria focused on the specific domain of BCI closed-loop systems integrated with ML and AI in health care. We prioritized studies published between 2019 - 2024 to ensure the relevance and timeliness of our findings. Our approach included not only technological advancements but also practical challenges and developments in BCI closed-loop systems in health care. Specifically, we reviewed studies that examined the various methods and parameters collected in BCI closed-loop systems (RQ1), ensuring a comprehensive understanding of data acquisition, preprocessing, and real-time feedback mechanisms. We also investigated the ML and AI algorithms used, and the outcomes obtained (RQ2), identifying the overall effectiveness of these algorithms in clinical and experimental settings. In addition, we focused on studies discussing the limitations encountered in current BCI closed-loop systems and proposed future research directions (RQ3), aiming to understand the barriers to implementation, ethical considerations, and technological limitations (RQ4). Exclusion variables from some

papers were added, as through our search, we filtered out papers that were not relevant to our goals, any research conducted on animals and not humans, and a lack of focus on Machine Learning.

Systematic Literature Search

Our systematic approach involved gathering, critical assessing, integrating, and presenting findings from various research papers on BCI closed-loop systems integrated with ML and AI in health care. We followed a detailed procedure to conduct and report systematic literature reviews, ensuring a rigorous selection process. Initially, we developed a carefully crafted search query to refine our search effectively, using terms such as “BCI OR brain computer interface,” “AND Machine Learning OR AI OR algorithm,” “AND Alzheimer OR Dementia.” Boolean operators like “AND” and “OR” were used strategically to narrow our search. This search spanned 4 major databases: PubMed, IEEE, ACM (Association for Computing Machinery), and Scopus. From these databases, we identified a total of 220 papers: 43 from PubMed, 22 from IEEE, 114 from ACM, and 41 from Scopus. After removing 8 duplicate records, 212 unique records were screened. During the screening phase, 179 records were excluded for reasons such as being out of context (n=84), not relevant to the research questions (n=94), or inaccessible (n=1). Subsequently, the titles and abstracts of the 212 screened records were assessed for eligibility, resulting in 33 full-text articles being reviewed. Of these, 15 reports were excluded due to being theses or books (n=9), report articles (n=4), or of poor quality (n=2). Ultimately, 18 studies met all inclusion criteria and were included in the final review. These papers were selected based on their focus on BCI closed-loop systems in health care, the integration of ML and AI, and their relevance to our study. We prioritized papers displaying rigorous methodologies, including empirical studies, surveys, case studies, experiments, and systematic literature reviews, showcasing innovative approaches, novel insights, or significant findings. In addition to the primary search, we cross-referenced each article’s citations to identify other pertinent papers, ultimately including any that fit our criteria.

Results

Study Selection and Characteristics

Our selection process, guided by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [24,25] guidelines as shown in Figure 2, allowed for a comprehensive understanding of the current state and future potential of BCI closed-loop systems in health care. We evaluated sources based on their methodology, innovation, significant findings, and overall relevance.

Our data analysis approach used a systematic data extraction method to rigorously analyze literature focused on BCI closed-loop systems integrated with ML and AI. This approach covered essential aspects such as the methods and parameters used in BCI systems, the ML algorithms used, challenges encountered, proposed solutions, and future research directions. Initially, we conducted an extensive literature review to identify pertinent studies. From this review, we developed a structured extraction matrix aimed at comprehensively capturing thematic elements critical to our study. The matrix included categories such as Title, Methods, Parameters, Machine Learning Algorithms, Challenges/Limitations, Proposed Solutions, Future Research Directions, and Title and Abstract Screening Score (0 - 3). These scores would be averaged out among a panel of 3 researchers with a 2 being a “Yes” to our paper list. To validate our methodology, we conducted several validation steps. First, we pilot-tested the matrix with a small sample of 10 papers to ensure it effectively captured relevant information while excluding irrelevant details. Second, we aligned the matrix variable with our research questions to ensure clarity in data extraction. The finalized matrix, formatted in Microsoft Excel, allowed for a smooth, systematic, and comparative analysis across selected papers, including Full Text Screening Score (0 - 3). This methodical approach enabled us to extract and synthesize data methodically, allowing anomalies and patterns to naturally emerge. Our synthesis and evaluation of articles were guided by their direct relevance to our study’s focus areas. This systematic approach ensured a robust analysis and provided a solid foundation for our literature review, as reflected in Table 1.

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

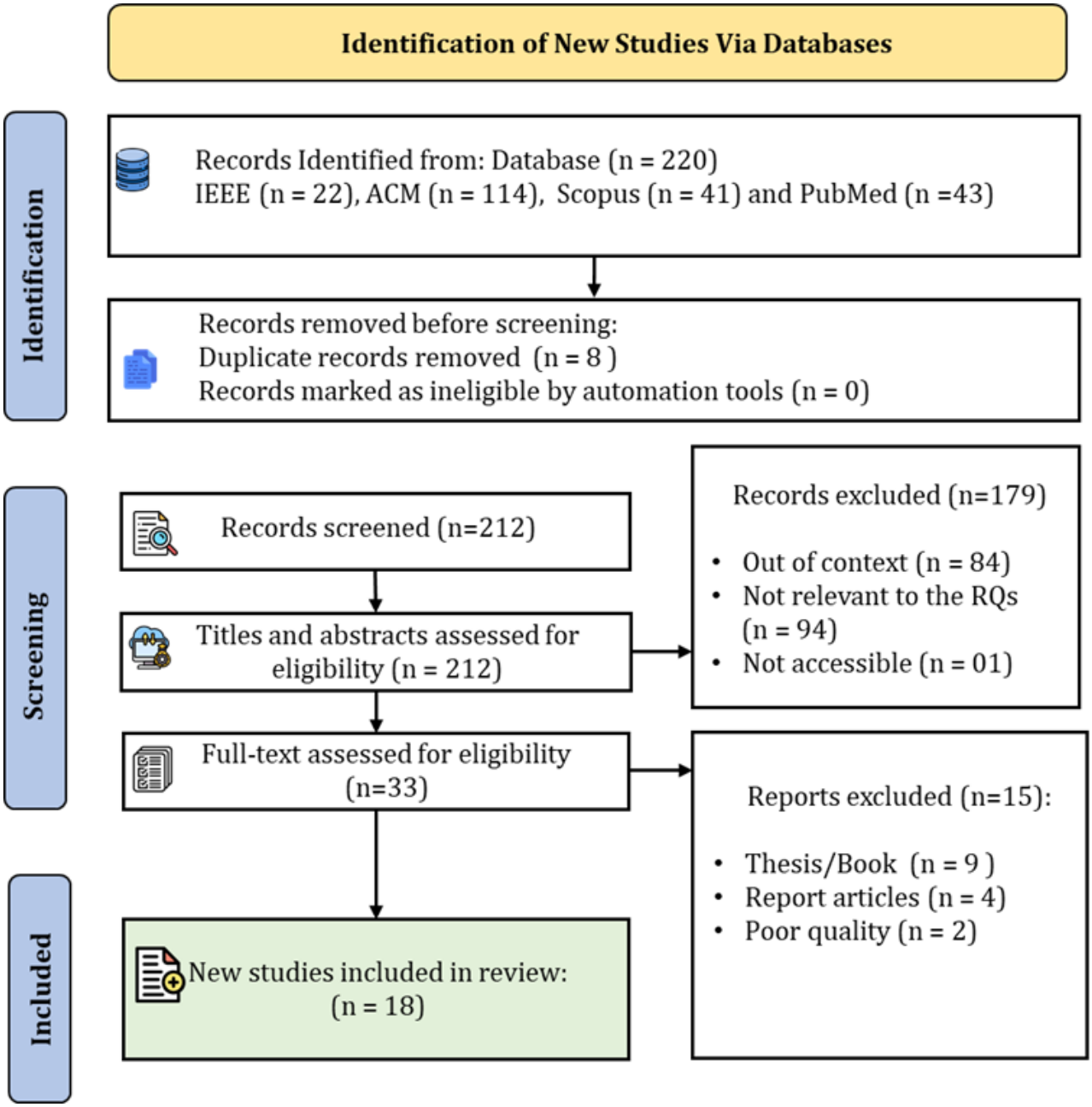


Table . Key algorithms and techniques commonly used in brain-computer interfaces.

Algorithm/technique	Role in BCIs[a]	Key applications	Advantages	References
Transfer learning (TL)	Feature extraction	Data alignment, spatial filtering	Improves robustness and accuracy	Shanechi [26]
SVM[b]	Classification	EEG[c] signal classification	High performance in high-dimensional space	Gu et al [27]
LDA[d]	Classification	EEG signal classification	Computational simplicity, good performance	Gu et al [27]
ICA[e]	Preprocessing	Artifact removal	Isolates artifact components from neural signals	Tsai et al [28]
CNN[f]	Feature extraction	Emotion recognition, workload estimation	High accuracy in classifying brain activities	Mughal et al [29]
TSNN[g]	Feature extraction, classification	Neural activity classification	Effective in high-dimensional data	Shin et al [30]
RBM[h]	Dimensionality reduction	Mental state recognition	Learns underlying data structures	Wang et al [31]
Fuzzy models	Classification	EEG pattern classification	Handles uncertainty and imprecision	Wu et al [32]
GANs[i]	Data augmentation	Augmented data generation	Increases data robustness and accuracy	Tsai et al [28]

[a]BCI: brain-computer interface.
[b]SVM: support vector machine.
[c]EEG: electroencephalography.
[d]LDA: linear discriminant analysis.
[e]ICA: independent component analysis.
[f]CNN: convolutional neural network.
[g]TSNN: tree-structured neural network.
[h]RBM: restricted Boltzmann machine.
[i]GAN: generative adversarial network.

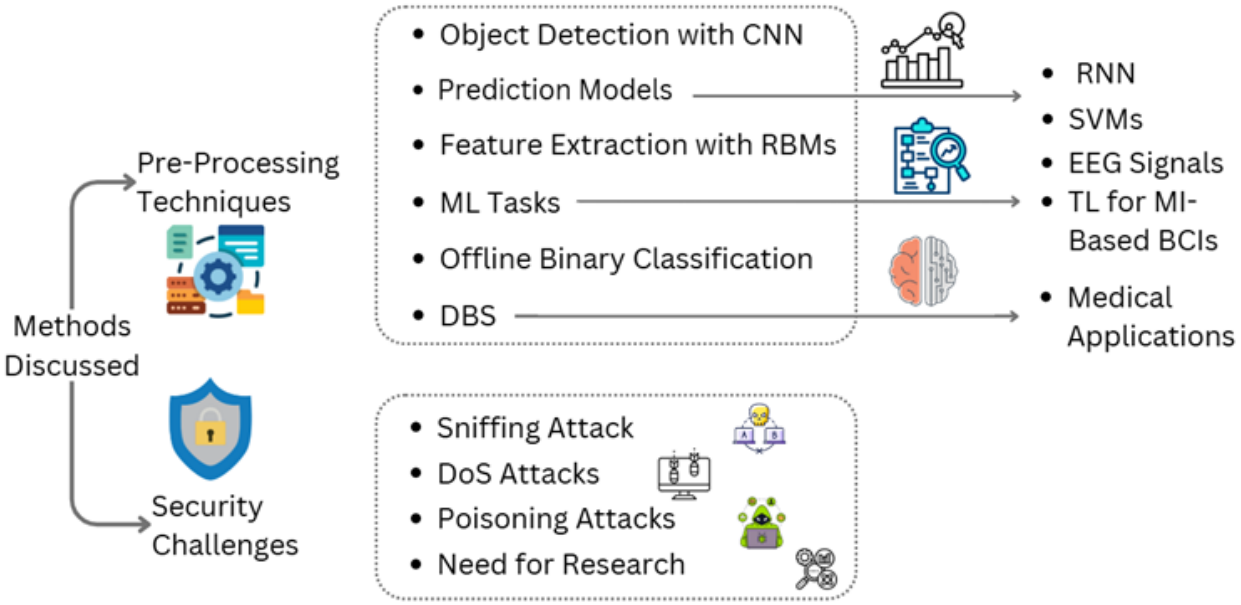
Methods and Parameters Used in the BCI Closed Loop System (RQ1)

Many studies have explored BCIs with closed-loop systems, but a comprehensive survey focusing on the challenges associated with methods and parameters used in these systems is still lacking. This section addresses this gap by reviewing various preprocessing techniques and the parameters used in BCI closed-loop systems, highlighting their implications for neural activity monitoring and intervention.

Preprocessing Techniques

The review identifies several effective methods and parameters that have demonstrated significant potential, as summarized in Figure 3. For instance, object detection is the paradigm for recognizing patterns using convolutional neural networks (CNN) [31], where it learns from more than a million images and can classify downstream objects in an image with high accuracy. This approach improves the ability to intermittently support real-time detection of nuanced neural activity and thus intervention. Likewise, Restricted Boltzmann Machines (RBMs) have been used to extract features for large-scale datasets [31]. Recurrent neural network (RNN) is a class of artificial neural network models that produce more accurate predictions than preferred direction and other systems like neuron-level readout methods including Poisson Process Velocity Tuning or generalized linear models (GLM). RNN can generate realistic simulations [33]. Support Vector Machines (SVMs) have been successfully used in small datasets, but their improvement to a larger accuracy level may be enhanced with Particle Swarm Optimization, particularly on the understanding of brain signals by means of EEG. BCI technology has been further refined by the categorization of different brain errors with SVMs. Motor Imagery (MI) is a mental process. MI starts from the thought of the movement of a body part. This activates different areas of the motor cortex and is commonly adopted for EEG-based BCIs. MI tasks performed by the users are sensed as EEG signals. TL makes use of source domain data to improve calibration in the target domain, which is a well-established technique used for improving MI-based BCIs [34]. In addition, offline binary classification is used to classify trials from target subjects. Currently, deep brain stimulation (DBS) is established as an effective treatment for conditions such as tremors, dystonia, and Parkinson disease. DBS also has shown promise in treating certain other types of chronic pain and psychiatric conditions, including neuropsychological tribulations. DBS is also being looked at as a possible pathway to the infusion of memory circuits and treatment avenues for dementia and Alzheimer disease.

Figure 3. Different methods and parameters in the brain-computer interface closed-loop system. CNN: convolutional neural network; DBS: deep brain stimulation; DoS: denial of service; EEG: electroencephalography; MI: motor imagery; ML: machine learning; RNN: recurrent neural network; SVM: Support Vector Machine; TL: transfer learning;



Security Challenges

These technologies, hereafter referred to as potential neuromodulatory treatments for symptoms, have been demonstrated to be capable of driving neural signals. In fact, BCIs have a high risk of sniffing attacks, where an attacker can eavesdrop on network channels and preview unencrypted data. This vulnerability can be used to affect denial-of-service attacks, which, in the case of implanted BCIs, target battery depletion [35]. Poisoning attacks alter the behavior of a BCI machine learning system by providing it with malicious input. These inputs are generated to lead the respective outputs of a system into misleading neural signaling patterns. These types of attacks have catastrophic consequences, such as failing to trigger an alarm for a seizure. The resolution of these security challenges is paramount in ensuring the safe and efficient roll-out of BCIs. Further research and development are needed to improve the privacy/security properties of these systems so people with neurological conditions would be able to heavily rely on them.

Effectiveness of the ML and AI Algorithms Used in the BCI Closed-Loop System (RQ2)

The effectiveness of ML and AI algorithms in BCI closed-loop systems is crucial for enhancing patient outcomes, particularly in applications related to neurorehabilitation and cognitive monitoring. These algorithms play a pivotal role in accurately interpreting neural signals, enabling real-time feedback and adaptive responses tailored to individual user needs. Their ability to analyze complex patterns in brain activity allows for improved signal classification and feature extraction, which are essential for ensuring reliable communication between the brain and external devices. Furthermore, the integration of effective ML and AI algorithms facilitates continuous learning and adaptation, ensuring that the BCI system evolves alongside the user's cognitive state. This adaptability not only enhances the overall user experience but also promotes better engagement and efficacy in therapeutic interventions, making the technology a powerful tool in managing neurological disorders.

Figure 4 illustrates the key machine learning techniques used in BCI closed-loop systems. It categorizes these techniques into preprocessing (eg, Independent Component Analysis [ICA] for noise reduction), data augmentation (eg, generative adversarial networks [GANs] for expanding training data diversity), feature extraction (eg, CNN and transfer learning [TL] for identifying critical signal patterns), and classification (eg, SSVMs and linear discriminant analysis [LDA] for categorizing neural signals). These methods collectively improve the system's effectiveness by refining the input data, enhancing model training with more varied data, extracting meaningful features, and accurately classifying neural patterns. This multistep approach enables closed-loop BCIs to achieve reliable real-time monitoring and intervention, making them more effective for health care and wellness applications.

In addition, Table 1 outlines the key algorithms and techniques commonly used in BCI systems, while Table 2 offers a detailed comparative evaluation of these machine learning approaches in closed-loop frameworks, emphasizing their applications, adaptability to neurological conditions, performance metrics, and computational complexity, supported by relevant literature.

Figure 4. Machine learning techniques in brain-computer interface closed-loop systems [27,28,30-35]. BCI: brain-computer interface; LDA: Linear Discriminant Analysis; SVM: Support Vector Machine; TSNN: tree-structured neural network;

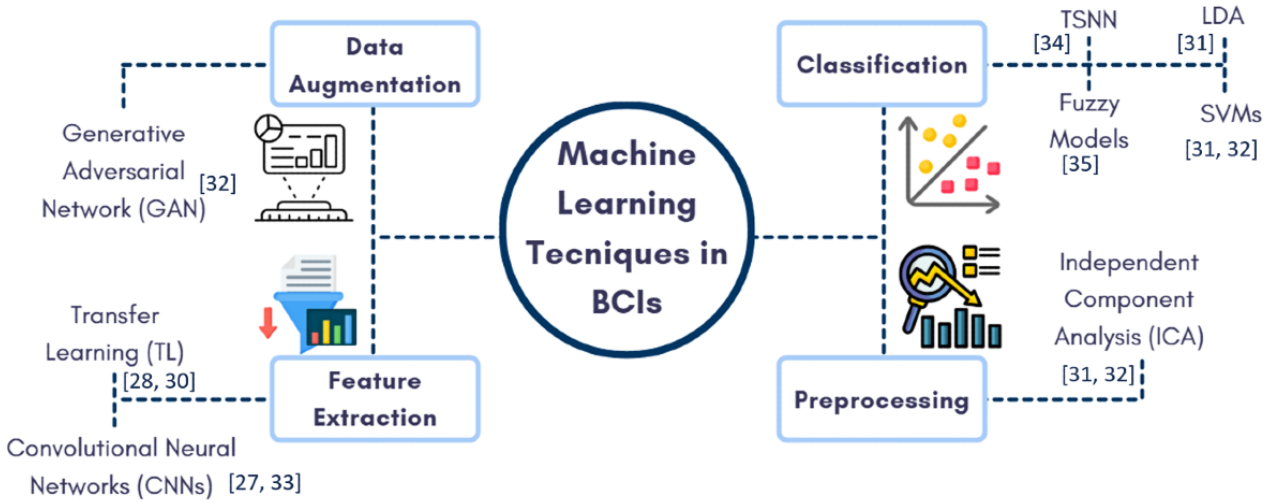


Table . Comparative performance of machine learning techniques in brain-computer interface closed-loop systems.

Algorithm/ technique	Key use cases in BCI[a] systems	Adaptability to neurological conditions	Avg accuracy / error Rate	Processing time / complexity	Reference(s)
SVM[b]	Motor imagery (MI), emotion recognition, EEG[c] classification	Moderate adaptability; sensitive to intersubject variability	78% - 90% (low error in MI classification)	Fast on small datasets; efficient for real-time binary tasks	Gu et al, Tsai et al [27,28]
CNN[d]	Emotion detection, mental workload, EEG-fNIRS[e] hybrid classification	High adaptability across subjects/sessions; handles complex patterns	>90% for workload/emotion tasks	High processing cost; ~300 - 500 ms latency unless optimized	Mughal et al, Liang and Kao [29,33]
TL[f]	MI classification, cross-session calibration, cognitive decline monitoring	Highly adaptable; ideal for changing patient conditions (eg, AD[g]/ADRD[h])	Reduces error up to 15% across domains	Moderate training cost; speeds up cross-subject adaptation	Shanechi, Belkacem et al [26,34]
LDA[i]	Basic EEG classification, passive BCI	Suitable for real-time low-power systems	~75% - 85% in EEG classification	Very low latency (<100 ms); lightweight	Gu et al [27]
ICA[k]	Noise reduction, preprocessing EEG/fNIRS	Improves SNR[j], crucial for low-signal patients (eg, dementia)	Preprocessing only (not classifier)	Fast artifact removal; boosts downstream model accuracy	Gu et al, Tsai et al [27,28]
RBM[l]	Mental state recognition, feature learning	Good for poorly labeled, noisy data (common in AD/ADRD)	~80% in unsupervised tasks	Medium complexity; good for dimensionality reduction	Wang et al, Golshan et al [31,36]
Fuzzy models	EEG pattern classification, aBCIs	Handles uncertainty well; ideal for imprecise EEG from late-stage dementia	70% - 85% (context dependent)	Low to medium; interpretable rule-based outputs	Wu et al [32]
GAN[m]	Data augmentation for EEG/BCI model training	Improves performance in data-scarce or imbalanced domains	Indirectly improves downstream model accuracy	High training time; not used in real-time inference	Tsai et al [28]
TSNN[n]	Neurological disorder detection, adaptive BCI	Effective in high-dimensional, complex datasets	~88% - 92% in neural activity classification	Moderate-to-high, but hierarchical structure improves learning	Shin et al [30]
RNN[o]	Cognitive state prediction, BCI simulations	Well-suited for time-series EEG signal modeling	~85% - 90% (task dependent)	Computationally intensive; not ideal for all real-time apps	Liang and Kao [33]

[a]BCI: brain-computer interface.
[b]SVM: support vector machine.
[c]EEG: electroencephalography.
[d]CNN: convolutional neural network.
[e]fNIRS: functional near-infrared spectroscopy.
[f]TL: transfer learning.
[g]AD: Alzheimer disease
[h]ADRD: Alzheimer disease and related dementia
[i]LDA: linear discriminant analysis.
[j]SNR: signal-to-noise ratio.
[k]ICA: independent component analysis.
[l]RBM: restricted Boltzmann machine.
[m]GAN: generative adversarial network.
[n]TSNN: tree-structured neural network.
[o]RNN: recurrent neural network.

Different ML Techniques in BCI Closed-Loop Systems

Machine learning algorithms have significantly enhanced the performance of BCIs. In that respect, one of the most influential techniques would be TL, which borrows knowledge from a source domain to perform better in a target domain. This is of particular importance to BCI, as the collection of data across sessions and subjects is often limited or variable [26]. TL has been successfully applied to data alignment, spatial filtering, feature selection, and classification tasks, dramatically

improving the robustness and accuracy of BCIs across different conditions and subjects.

Complementing TL, SVMs have turned out to perform very well in high-dimensional spaces and have performed both linear and nonlinear classification using the kernel trick on data transformation [27]. SVMs have found wide applications in BCI applications, more specifically in classifying EEG signals. The SVM finds the optimal hyperplane that allows separation of different classes, thus always showing good performance in most BCI tasks, such as motor imagery classification and emotion recognition [28].

On the other side, LDA finds out the best separation between several classes by maximizing such a separation by choosing an appropriate linear combination of features. LDA has a nice balance between computational simplicity and performance for BCIs [27].

The role of ICA is paramount in preprocessing methods. ICA is one of the key tools that attempt to separate multivariate signals into additive, independent components [27]. Especially with BCIs, it is very good at isolating artifact components from the neural signals, hence improving the quality of data used for subsequent classification tasks. This step in preprocessing appreciably improves the accuracy of a number of BCI applications.

Moving to more complex models, CNNs have been very triumphant in visual and spatial data, including EEG and functional near-infrared spectroscopy (fNIRS) signals. CNNs are known for their emotional ability to build hybrid brain images that classify the activities taking place in the brain in a very accurate manner for the detection and interpretation of any complex neural pattern [29]. This is critical in applications like mental workload estimation and emotion recognition, where a spatial hierarchy in neural data may be critical for appropriate classification and analysis [33] .

Tree-structured neural networks (TSNN) combine decision trees and neural networks to provide the possibility of hierarchical feature extraction and classification. More importantly, these networks work quite effectively in relation to data: complex and high-dimensional. TSNNs are therefore able to yield promising results on the classification of neural activities and detecting symptoms of neurological disorders with a rich set of neural biomarkers [30]. This fills a gap in the field by providing an optimal balance between accuracy and computational efficiency, needed for real-time BCI applications.

RBMs have made some very great contributions to unsupervised scenes, where instances of the objective are to learn underlying structures of data [31]. In this case, RBMs learn with effective features and reduce dimensionality to improve the performance of classifiers on mental state recognition and motor imagery classification tasks [36].

On the other side, fuzzy models represent the uncertainty and imprecision of EEG data using fuzzy logic. Such models generate rules that are much closer to those resulting from human reasoning and hence are very suitable for processing nonlinear and nonstationary signals. Fuzzy models applied in BCIs include the so-called fuzzy inference systems (FIS) and fuzzy neural networks (FNNs) for classifying EEG patterns, offering both accuracy and interpretability [32].

Another extension to the toolkit of BCIs is GANs. It consists of two neural networks: a generator and a discriminator. These networks counteract in a framework, and each has an opposite goal in a zero-sum game setup. GANs' application in BCIs is in augmented data generation for improving classifier training, more so when there is not enough data, as GANs increase the robustness, hence the accuracy, of BCI systems by bringing forth more training data. Improvements in these machine learning algorithms have increased the potential of BCIs not only on grounds of performance but also by opening new avenues for possible clinical and practical applications [28].

Table 1 outlines key methods used in BCIs, detailing their roles, applications, and benefits. Techniques like TL, SVM, LDA, and ICA enhance data preprocessing and classification, improving signal quality and performance. CNNs and TSNNs excel in feature extraction and classification of complex neural data, while RBMs and Fuzzy Models handle dimensionality reduction and uncertainty in EEG signals. GANs support data augmentation, boosting robustness and accuracy. These methods collectively optimize the processing of neural signals in closed-loop BCI systems.

TL, SVM, LDA, ICA, CNN, TSNN, RBMs, fuzzy models, and GAN techniques have helped in making BCIs effective and reliable. These algorithms help improve the capability of BCIs to better handle the user's requirements, reduce calibration time, and realize more accurate and robust control of artificial limbs and other devices.

Limitations in the Development and Implementation of the BCI Closed-Loop System (RQ3)

Some of the limitations facing BCIs' development and implementation can be summarized as ranging from decoding algorithms through neural and behavioral measurements to computational constraints, as shown in Table 3. The table outlines various challenges associated with BCI technology and proposes corresponding solutions to address these issues. It covers aspects like improving neural signal decoding, enhancing sensor accuracy, and increasing the precision of behavioral measurements. These limitations show requirements for further research in terms of target setting and orientation of work for increasing effectiveness. Further explanations about these are as follows:

Table . Challenges and solutions in brain-computer interface development.

References	Associated problems	Proposed solutions
Bryan et al [37]	Neural signal decoding	Develop sophisticated algorithms
Jiang et al [38]	Accurate neural measurements	Advanced sensor technologies
Jiang et al [38]	Behavioral measurements	Improve granularity and precision
Gu et al [27]	Computational cost	Optimize algorithms and hardware
Gu et al [27]	Long calibration sessions	Develop efficient calibration methods
Merk et al [39]	Electrode design	Enhance ergonomic and reliable designs
Yue et al [35]	Decoding and encoding algorithm heterogeneity	Standardize methodologies
Wu et al [32]	Lack of long-term studies	Conduct long-term validation studies
Mughal et al [29]	Hardware limitations	Develop scalable hardware
Golshan et al [36]	Model generalization	Ensure models generalize to closed-loop conditions
Xavier et al [40]	Privacy and security	Implement robust security measures

Neural Signal Decoding

The challenging part of BCIs is the neural signal decoding into meaningful commands. This is due to the large array of neural signals requiring high accuracy; this becomes very challenging, especially in scenarios where they are either too noisy or highly variable across different cognitive states. This variability calls for sophisticated algorithms that can adapt to these changes and guarantee real-time performance [37]. Successful BCIs require accurate neural measurements; conventional methods generally have spatial and temporal resolution that is inadequate.

Behavioral Measurements

Behavioral measurements correlated with specific neural activities often experience imprecision and lack the fine detail necessary for comprehensive analysis. This limitation stems from the complexity of human behavior and the intricate relationship between neural processes and external actions. Standard measurement techniques may fail to capture the subtleties of these interactions, resulting in a loss of crucial information that could deepen our understanding of brain-behavior dynamics. Compounding this issue is the challenge posed by the time scale of behavioral dynamics; neural activities can change rapidly, often within milliseconds, while corresponding behavioral responses may take longer to manifest. This discrepancy makes it difficult to capture and analyze real-time correlations, as a sudden shift in brain activity may not immediately lead to observable changes in behavior, creating potential misalignments in data interpretation [38]. Consequently, the inability to accurately synchronize these fast-changing neural activities with their associated behaviors can hinder our understanding of cognitive processes and impair the effectiveness of interventions in areas like neurorehabilitation and BCIs. Addressing these challenges necessitates the development of advanced measurement techniques and analytical frameworks capable of capturing the nuances of both neural dynamics and behavioral responses.

Computational Cost

The high computational cost associated with processing and analyzing neural data presents a significant challenge in the development and implementation of brain-computer interface systems. TL techniques, when integrated with active BCIs, can incur substantial computational expenses due to the high-dimensional nature of neural data and the complexity of the models involved. This complexity poses a considerable burden on the real-time applicability of BCI systems, limiting their responsiveness and efficiency in practical scenarios. Furthermore, the current applications of TL in BCI research have primarily focused on binary MI classification problems, which restrict the versatility and scope of TL methods in broader contexts. As a result, the limitations of TL not only affect the computational feasibility of BCIs but also hinder their potential for more complex tasks, such as multi-class classification or real-time adaptive learning.

Long Calibration Sessions

One significant challenge associated with most MI-based BCIs is the extensive calibration sessions required before they can operate effectively. These lengthy calibration processes diminish the overall usability and practicality of BCIs, particularly in real-world applications where quick deployment is essential. To enhance the applicability of TL in everyday situations, it is crucial to develop more efficient calibration methods that can streamline the setup process and reduce the time commitment for users [27].

Electrode Design

An integral aspect of neural signal acquisition in BCIs is the design and fabrication of electrodes. Current electrode designs face significant challenges related to mechanical and electrical reliability, flexibility, and the speed at which they can accommodate various configurations. These issues can hinder the overall performance of BCIs, as unreliable electrodes may lead to inconsistent signal quality and compromised data accuracy. In addition, the pressure exerted by BCI headsets on the user's head can result in discomfort during prolonged use, underscoring the need for improved ergonomic designs.

Enhanced ergonomic considerations not only promote user comfort but also facilitate longer monitoring sessions, which are crucial for effective neural signal acquisition. By addressing these challenges in electrode design and headset ergonomics, researchers can significantly improve the functionality and user experience of BCIs, ultimately expanding their applications in clinical settings and enhancing the quality of life for individuals who rely on this technology [39].

Decoding and Encoding Algorithm Heterogeneity

The heterogeneity of decoding and encoding algorithms used in BCIs represents a significant challenge in the field. This diversity complicates comparisons across different closed-loop BCIs, as variations in purpose, methodology, and outcomes hinder the establishment of standardized benchmarks and best practices. Furthermore, the majority of existing studies tend to focus narrowly on cognitive neural features, often neglecting affective aspects of BCIs. This limited scope underscores the pressing need for larger, more comprehensive studies that encompass a broader range of neural activities and scenarios. By addressing the issues of algorithmic heterogeneity and expanding the research focus, the BCI community can enhance the comparability of findings, foster innovation, and ultimately improve the effectiveness and applicability of BCIs across various domains [35]. This will facilitate a deeper understanding of how different neural signals can be decoded and encoded, paving the way for more nuanced applications in both cognitive and affective realms.

Lack of Long-Term Studies

The absence of long-term studies significantly undermines the effectiveness of training BCI systems. Establishing a robust definition of a reinforcement signal is crucial, yet it raises ethical concerns, particularly when involving human participants. To mitigate these ethical dilemmas, it may be more appropriate to conduct initial experiments in nonhuman models, thereby sidestepping potential ethical issues. In addition, there is no assurance that human participants will interpret the feedback provided to them as a reward, complicating the training process further. This variability in interpretation can lead to inconsistent learning outcomes, making it challenging to develop reliable and effective BCI systems. Therefore, conducting comprehensive long-term studies is essential for refining training protocols, ensuring ethical compliance, and ultimately enhancing the overall effectiveness and applicability of BCIs in real-world scenarios [32].

Hardware Limitations

Hardware limitations pose significant challenges to the therapeutic effectiveness of BCIs, primarily through the need for higher channel counts and improved scalability. These requirements can result in the loss of critical information due to downsampling and channel selection processes, which may eliminate relevant neural signals necessary for accurate interpretation. In addition, there is often a considerable disparity between the sampling rate and the number of channels in EEG and fNIRS data, complicating the data analysis process. While proposed methodologies to address these issues aim to enhance data integrity, they frequently come with high computational costs and complexities that hinder their applicability in real-world settings. Consequently, overcoming these hardware limitations is crucial for advancing BCI technology, ensuring that it can deliver reliable and effective therapeutic outcomes for users [29].

Model Generalization

Another significant challenge in the development of BCIs is ensuring that models trained on open-loop data can effectively generalize to closed-loop conditions. The experiments necessary for this validation are often prohibitively expensive and time-consuming, which limits their widespread implementation. This highlights the critical need for real-time applicability of these models to facilitate the validation of adaptive deep brain stimulation (aDBS) systems [36]. Without the ability to efficiently transfer knowledge gained from open-loop scenarios to real-time closed-loop environments, the effectiveness and reliability of BCIs in practical applications remain in question. Thus, enhancing model generalization is essential for advancing BCI technology and ensuring its successful integration into therapeutic settings.

Privacy and Security

The issues surrounding privacy, security, and ethics are of paramount importance in the context of BCIs [41-44]. These systems are susceptible to various data breaches and cyberattacks, including cryptographic attacks, denial-of-service attacks, and sniffing attacks, which can compromise sensitive neural data and user information. Such vulnerabilities underscore the urgent need for robust privacy protection and comprehensive security measures to safeguard both the integrity of the data and the users' personal information. In addition, ethical considerations surrounding the use of BCIs are critical, particularly regarding user privacy and informed consent. It is essential that users are fully aware of how their data will be used and are able to provide consent without coercion. Addressing these privacy, security, and ethical concerns is vital for the responsible development and deployment of BCI technologies, ensuring that they benefit users while minimizing potential risks and harms [40].

Ongoing collaborative research efforts are actively addressing the critical limitations identified in current BCI closed-loop systems. Among the most promising directions is the development of more sophisticated decoding algorithms capable of accommodating the inherent variability in neural signals across individuals and cognitive states. Improvements in neural and behavioral measurement precision—through advanced sensor technologies, multimodal signal integration, and robust signal processing methods—are also contributing to more accurate and responsive BCI systems. A key advancement involves the integration of TL with active BCIs beyond traditional binary classification, allowing systems to adapt across sessions and users while minimizing lengthy calibration times. In parallel, the design of more comfortable and reliable electrodes, alongside expanded studies into cognitive and affective dimensions of brain activity, is broadening the applicability of BCIs in both clinical and non-clinical environments. Furthermore, enhancing hardware scalability and addressing data loss due to downsampling remain essential for

the therapeutic efficacy and widespread deployment of these systems. Ethical implementation, including user-informed consent and privacy-preserving frameworks, must be embedded into system design to ensure trust and adoption.

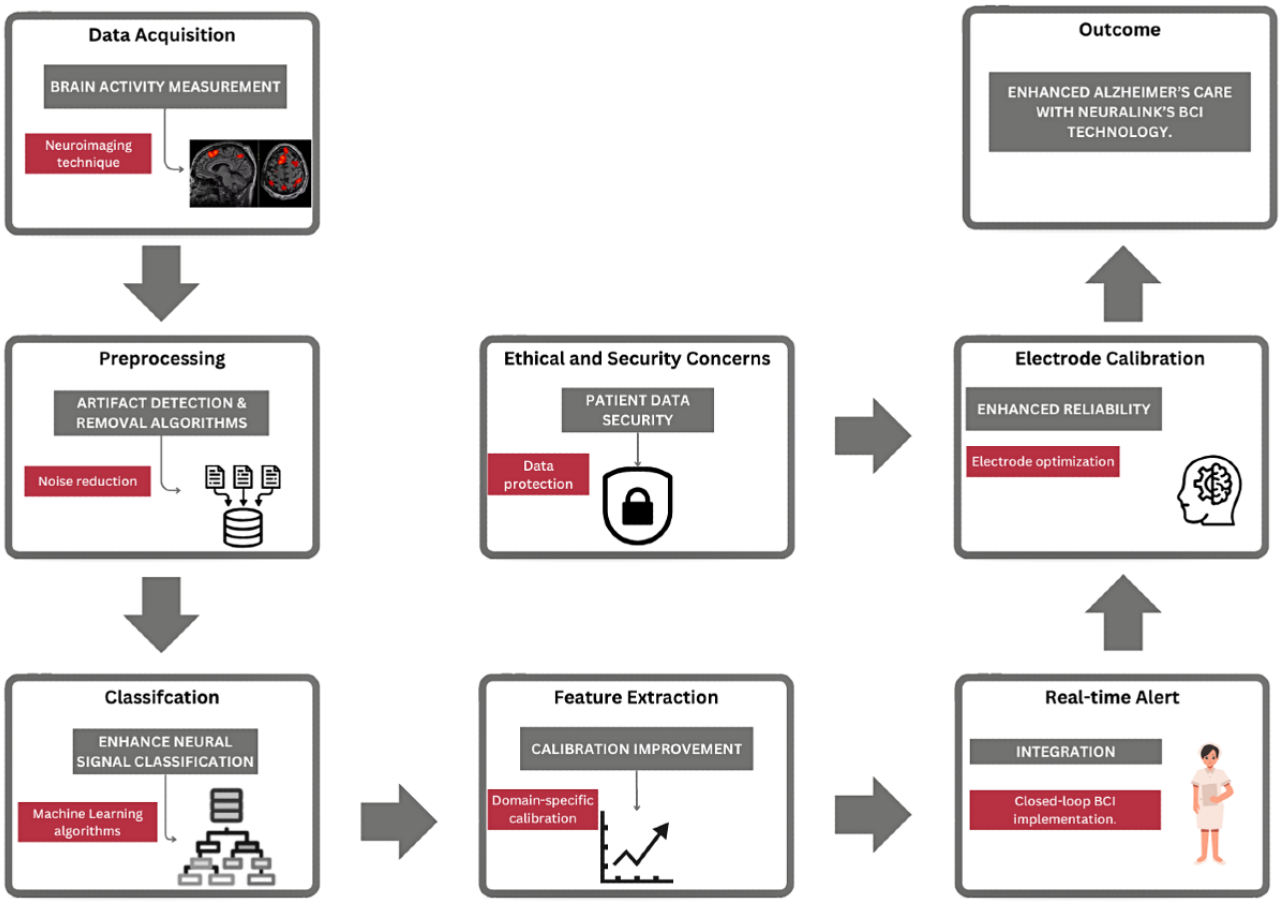
Recent advancements underscore how these challenges are being met through innovative and applied research. For instance, studies using TL and One-Shot Learning demonstrate that calibration requirements can be drastically reduced by reusing training data across users and sessions, enabling more efficient deployment in real-world environments [26,28]. In addressing cybersecurity concerns, researchers have proposed advanced encryption protocols and privacy-preserving neural computation strategies to mitigate sniffing, poisoning, and denial-of-service attacks—ensuring the confidentiality and integrity of neural data [35,40]. Notably, real-world applications such as Neuralink's adaptive BCI and Tsai et al's [28] secure closed-loop brain-machine interface exemplify successful responses to these challenges. These platforms leverage online tuning algorithms, secure data pipelines, and adaptive feedback systems to maintain robust performance while safeguarding patient data in both clinical and home care settings [35,45]. Together, these advancements highlight the growing maturity of BCI technologies and point toward a future in which user-friendly, secure, and scalable BCI systems are a practical reality.

BCI Closed-Loop System-Based Framework for Longitudinal Monitoring of AD/ADRD Patients (RQ4)

AD/ADRD is a progressive neurological disorder, and one of the major causes of death among the older adults [46,47]. Therefore, it is important to acquire new solutions that can enhance the quality of life of patients and their caregivers as the number of people affected increases yearly. The use of BCI technology is considered one of the most promising approaches to this challenge, using state-of-the-art neuroimaging techniques and machine learning algorithms for continuous monitoring and diagnosis.

The proposed framework shown in Figure 5 overcomes the complexity of decoding neural activity in AD/ADRD patients, who often lack the cognitive control normally required by conventional BCIs. It illustrates a closed-loop BCI framework for longitudinal monitoring of AD/ADRD patients, beginning with the acquisition of brain activity through neuroimaging techniques to capture critical neural signals. The collected data undergoes preprocessing, where noise is reduced using artifact detection and removal algorithms, ensuring high-quality signals. Ethical and security concerns are addressed by implementing robust data protection measures to safeguard sensitive patient information. Machine learning algorithms classify the neural signals, accurately distinguishing between various mental states, while feature extraction and domain-specific calibration improve the system's precision. Electrode calibration ensures reliable signal acquisition, enhancing system performance. Real-time alerts integrate closed-loop monitoring for continuous observation, enabling timely interventions. Ultimately, this framework aims to improve Alzheimer patient care by leveraging Neuralink's BCI technology [45] for better monitoring and intervention strategies. The different stages are further elaborated below:

Figure 5. Closed-loop brain-computer interface framework for longitudinal monitoring of Alzheimer disease and related dementias patients.



Data Acquisition and Processing

The proposed framework is empowered for the correct interpretation of neural signals through sophisticated classifiers such as SVMs and LDA; it thus helps in the achievement of reliable monitoring and communication. Once brain activity data is collected through neuroimaging techniques like fMRI and EEG, preprocessing steps are crucial for artifact detection and noise reduction [48]. Advanced algorithms are used to filter out irrelevant data, ensuring that the subsequent analysis focuses on meaningful neural patterns.

Classification, Feature Extraction, and Electrode Calibration

In the classification phase, machine learning algorithms improve the deciphering of neural signals. In doing so, SVMs and LDA techniques can be used to classify the different mental states, giving insights into the patient responses and cognitive conditions. In this step, it is important to distinguish different neural activities and their behaviors. Feature extraction works on refining the accuracy and reliability of the classifications. These domain-specific calibration methods enable the tuning of the analysis so that the system learns the characteristics of the individual's needs and variations in neural activity [48,49]. At this stage, electrode calibration is crucial for ensuring BCI system reliability. Optimized placement and configuration of electrodes ensures consistent data acquisition, reducing errors and thus enhancing the system's overall performance.

Real-Time Alert

Real-time alerting of caregivers is another important factor in the framework. Data from both AD patients can be used to develop the alert system, ensuring that it can accurately identify deviations from normal neural activity. For instance, the classification accuracy reported in [48] indicates that the system can reliably distinguish between different mental states, which is crucial for triggering real-time alerts. Thus, this feature serves significantly in terms of preventing accidents or ensuring timely assistance from a medical point of view. This framework is aimed at enhancing the care of Alzheimer disease by integrating Neuralink's BCI technology. The framework offers a strong and trustworthy tool for the betterment of the patient outcome and elongation of time that patients can spend with their loved ones by addressing the challenges of neural signal decoding, data security, and real-time monitoring.

Ethical and Security Concern

After analyzing data, references indicate the most prominent symptoms of AD are severe deficits in communication, cognitive decline, and behavioral changes. Traditional BCIs require active participation; hence, they cannot be used with AD patients. For example, in [48] the authors highlight that traditional BCIs, requiring active control, are not suitable for AD patients due to their cognitive impairments. Instead, the study emphasizes the use of passive BCIs, which leverage preserved affective responses to facilitate basic communication and continuous monitoring [48]. This method makes use of detailed neuroimaging and machine learning components for cognitive

and behavioral changes in order to provide such monitoring. Such continual assessment allows for early diagnosis and timely intervention that are crucial in managing AD progression [49]. In that respect, the BCI systems are vulnerable to data breaches and ethical misuses. The security aspects are core parts of the proposed BCI framework, evident in the position of ethical and security considerations at the very center of the framework diagram in Figure 5 [50]. This framework would hardwire strong measures for security against breaches and ethical misuse of information related to a patient. This involves encryption methods, secure storage, and tight access control that ensures only authorized personnel can view sensitive information. Despite technical possibilities, data protection and ethical guidelines are argued to exist, whereas BCI systems are weak in cryptographic, denial-of-service, and sniffing cyber-attacks [50]. Therefore, strong protection measures must be implemented. Implantable BCI devices give real-time data, thereby allowing caregivers to receive instant alerts on a patient's condition. This approach will improve the quality of care and prevent emergencies.

The envisioned BCI framework addresses the most critical challenges of caring for the patient experiencing Alzheimer disease by efficiently combining advanced neuroimaging techniques with machine learning algorithms as shown in Table 4. In this regard, this approach is relevant for improving patient clinical outcomes while assisting caregivers in handling the complexities of Alzheimer disease management by enhancing neural signal classification, guaranteeing data security, and real-time monitoring [50].

Table . Mapping of framework components to challenges and supporting literature.

Framework component	Challenge addressed	Description	Supporting literature
Neural Signal Acquisition (EEG[a]/fMRI[b])	Low signal-to-noise ratio (SNR); variability across sessions	Uses EEG and fMRI for high-resolution brain activity monitoring; requires preprocessing for noise	Gu et al, Liberati et al [27,48]
Preprocessing (eg, ICA)	Artifact contamination; real-time signal distortion	Independent Component Analysis (ICA) removes artifacts to improve signal clarity	Gu et al, Tsai et al [27,28]
Feature extraction and classification (SVM[c], LDA)[d]	Inaccurate decoding of mental states	SVM and LDA used to classify neural patterns for real-time state detection	Shanechi, Gu et al [26,27]
Transfer learning (TL)	Lengthy calibration sessions; cross-user variability	Reduces setup time by leveraging previously trained models from similar domains	Shanechi, Belkacem et al [26,34]
Domain-specific calibration	Adaptability to individual neural profiles	Fine-tunes BCI[e] parameters to individual characteristics	Shanechi, Liberati et al [26,48]
Real-time alert system	Lack of timely caregiver intervention	Monitors patient state continuously and sends alerts to caregivers during anomalies	Pisarchik et al, Liberati et al [45,49]
Passive BCI design	Limited cognitive engagement in AD[f]/ADRD[g] patients	Enables nonintrusive monitoring based on implicit neural responses	Liberati et al, Liberati et al [48,49]
Security and ethical framework	Privacy risks; cyber threats; informed consent	Implements encryption, access control, and ethical safeguards for neural data	Yue et al, Xavier Fidêncio et al [35,40]
Scalable hardware integration	Usability and long-term deployment	Incorporates ergonomic, wearable sensors for home and clinical environments	Mughal et al, Merk et al [29,39]

[a]EEG: electroencephalography.

[b]fMRI: functional magnetic resonance imaging.

[c]SVM: support vector machine.

[d]LDA: linear discriminant analysis.

[e]BCI: brain-computer interface.

[f]AD: Alzheimer disease.

[g]ADRD: Alzheimer disease and related dementia.

Discussion

Principal Findings

This systematic review synthesized the current evidence from 18 studies on the integration of AI and ML within BCI closed-loop systems for neurorehabilitation, with a specific focus on AD/ADRD. The findings indicate that ML techniques such as TL, CNNs, and SVMs significantly enhance the performance of BCI systems by improving real-time signal classification, feature extraction, and cross-session adaptability. However, the translation of these technological advancements into widespread clinical practice is hampered by significant

challenges, including signal variability, computational demands, lengthy calibration, and profound privacy concerns. The proposed framework for longitudinal AD/ADRD monitoring represents a promising, patient-centric application that leverages passive BCI paradigms to circumvent the cognitive demands of traditional systems.

General Interpretation in the Context of Existing Evidence

Our findings on the efficacy of ML algorithms like CNNs and TL in BCI systems are strongly supported by the broader literature on AI in digital health. The high accuracy (>90%) of CNNs in classifying complex neural patterns for mental workload and emotion recognition aligns with their proven success in other pattern recognition domains, such as medical imaging. Similarly, the utility of TL in reducing calibration time and improving cross-subject generalization addresses a well-documented bottleneck in BCI research, echoing its successful application in other fields where data scarcity and individual variability are concerns. The review's identification of passive BCIs as a solution for AD/ADRD patients is particularly insightful. This approach is consistent with a growing trend in digital biomarkers, which seeks to leverage implicit, continuous data from wearables and other sensors for early disease detection and monitoring, moving beyond active user participation. Furthermore, the emphasis on real-time, closed-loop feedback for neurorehabilitation is supported by neuroscientific principles of neuroplasticity. The ability of AI-enhanced BCIs to provide immediate, adaptive intervention is theorized to strengthen neural pathways more effectively than open-loop systems, a hypothesis that is gaining traction in stroke and spinal cord injury rehabilitation. Thus, the results of this review are not isolated but are part of a convergent evolution across AI, neuroscience, and clinical medicine toward more adaptive, data-driven therapeutic interventions.

Limitations of the Included Evidence

While the reviewed studies demonstrate significant promise, the evidence base has several important limitations that temper the immediate readiness of these technologies for clinical deployment. The majority of included studies were small-scale, laboratory-based demonstrations. They often involved healthy participants or highly controlled patient groups, lacking the diversity and complexity of real-world clinical environments. This limits the generalizability of the reported high accuracy rates. In addition, as highlighted in the review, there is a pronounced heterogeneity in decoding algorithms, performance metrics, and experimental protocols across studies. The absence of standardized benchmarks makes it difficult to directly compare the performance of different ML models or BCI systems, hindering the identification of optimal approaches. Furthermore, there is a critical gap in long-term longitudinal studies. It remains largely unknown how these systems perform over months or years, how they adapt to disease progression, and whether improvements in signal classification accuracy translate into meaningful clinical outcomes, such as slowed cognitive decline or improved quality of life.

Limitations of the Review Process

This review itself is subject to certain methodological limitations that should be acknowledged. Limiting the search to studies published between 2019 and 2024, while ensuring timeliness, may have excluded foundational or highly relevant older studies. Furthermore, while major databases were consulted, the exclusion of other potential sources may have led to the omission of pertinent research. Next, the review likely reflects a positive publication bias, as studies with null or negative results are less frequently published. This may create an over-optimistic picture of the current capabilities and reliability of AI-driven BCIs.

Despite following PRISMA guidelines and using a panel of researchers, the processes of screening titles/abstracts and extracting data into a matrix involve a degree of subjective judgment, which could have influenced the final selection and synthesis of the studies.

Implications for Practice, Policy, and Future Research

The findings of this review have several critical implications across different domains:

For clinical practice: in the short term, AI-enhanced BCIs are most likely to find application as sophisticated diagnostic and monitoring tools in specialized neurology centers, aiding in the early and objective detection of cognitive impairment. The proposed framework for AD/ADRD provides a blueprint for developing caregiver alert systems, which could significantly reduce burden and improve patient safety in home-care settings. Clinicians should be aware of these emerging technologies to guide future patient care and manage expectations.

For policy and regulation: the security vulnerabilities and ethical dilemmas identified (eg, data privacy, informed consent for cognitively impaired users) demand urgent attention from policymakers and regulatory bodies like the FDA and EMA. New frameworks are needed to govern the security of neural data, which is arguably the most personal of all health information. Policies must be established to ensure equitable access and prevent misuse, defining clear guidelines for the ethical development and clinical validation of BCI technologies.

For future research: future work must transition from proof-of-concept to robust, clinically focused research. Key priorities should include rigorous, long-term trials with diverse AD/ADRD populations that are essential to validate efficacy and establish clinical utility. In addition, the BCI research community should collaborate to establish common data formats, reporting standards, and performance benchmarks to enable meaningful comparisons. Moreover, research must focus on developing more ergonomic, user-friendly, and low-power hardware that is suitable for prolonged use outside the lab. Creating interpretable ML models will be crucial for building trust among clinicians and patients, allowing them to understand the basis for the system's classifications and decisions.

Conclusion

This review systematically explored the role of BCI closed-loop systems in health care, with a specific focus on their potential to enhance neurological disorder detection and management

through advanced ML and AI techniques. Addressing RQ1, we analyzed various methods and parameters used in BCI closed-loop systems, including signal acquisition, feature extraction, classification, and device output. Key preprocessing techniques such as ICA and TL were identified as crucial for reducing noise and improving signal quality. DBS was also highlighted as a promising intervention for neuropsychological disorders like AD and ADRD.

In evaluating RQ2, we examined the effectiveness of ML and AI algorithms in BCI systems. Techniques like Support SVM, CNN, and RNN demonstrated significant improvements in decoding neural activity, enabling more accurate classification of cognitive states. TL, in particular, showed promise in reducing calibration time, making BCI systems more adaptive to individual users. In addition, BCIs have expanded beyond disease detection, playing a pivotal role in cognitive enhancement, neurofeedback training, and assistive communication.

Despite these advancements, RQ3 highlighted several challenges in the development and implementation of BCI closed-loop systems. Key limitations include high computational costs, long calibration sessions, signal variability across individuals, and security risks such as Poisoning Attacks that could compromise neural signal integrity. Ethical concerns surrounding data privacy and the potential misuse of BCIs also remain pressing issues. Addressing these challenges requires advancements in real-time signal processing, improved sensor technology, and robust cybersecurity frameworks to protect patient data.

To answer RQ4, we proposed a BCI-based framework for longitudinal monitoring of AD/ADRD patients, integrating real-time neural signal acquisition, feature extraction, and ML-based classification for early cognitive decline detection. This framework incorporates real-time alert systems to assist caregivers in proactive intervention, enhancing patient outcomes. In addition, passive BCIs were identified as a viable alternative for patients with severe cognitive impairments, enabling continuous monitoring without requiring active user engagement.

To answer RQ4, we proposed a BCI-based framework for longitudinal monitoring of AD/ADRD patients, integrating real-time neural signal acquisition, feature extraction, and ML-based classification for early cognitive decline detection. This framework incorporates real-time alert systems to assist caregivers in proactive intervention, enhancing patient outcomes. In addition, passive BCIs were identified as a viable alternative for patients with severe cognitive impairments, enabling continuous monitoring without requiring active user engagement.

Building on these advancements, future research should prioritize the refinement of machine learning algorithms to better support real-time signal processing and adaptive learning in dynamic environments. Ethical considerations—such as user consent, data ownership, and secure data handling—must remain central to system design. Continued progress in these areas will be essential for creating scalable, secure, and user-friendly BCI systems that integrate seamlessly into daily life. Ultimately, these innovations will position AI-powered BCIs as transformative tools in improving care, independence, and quality of life for individuals with neurological disorders, particularly those living with AD/ADRD.

Acknowledgments

This project is based on work supported by the National Science Foundation under Grant No. 22,44,450.

Authors' Contributions

CW, FIA, and MMH contributed to the conceptualization of the study. Data curation was performed by JRC and CW. Formal analysis was conducted by CW, FIA, and AC. Funding acquisition was led by ST and SH. Investigation was performed by CW, FIA, and MMH, while methodology was developed by CW, AC, and JRC. Project administration and validation were supervised by NS. Software development was completed by FIA. Visualization was done by FIA and CW. CW and FIA prepared the original draft, and FIA and NS reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 32 KB - biomedeng_v10i1e72218_app1.docx]

References

1. Hossein Motlagh N, Zuniga A, Thi Nguyen N, et al. Population digital health: continuous health monitoring and profiling at scale. Online J Public Health Inform 2024 Nov 20;16:e60261. [doi: 10.2196/60261] [Medline: 39565687]
2. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861. [doi: 10.2196/12861] [Medline: 31512582]
3. Digital health - statistics & facts. Statista n.d. URL: https://www.statista.com/topics/2409/digital-health/ [accessed 2025-02-02]

4. MHealth - statistics and facts. Statista. URL: https://www.statista.com/topics/2263/mhealth/?utm_source=chatgpt.com [accessed 2025-02-02]
5. Healthcare technology statistics & market size analysis. TATEEDA. URL: https://tateeda.com/blog/healthcare-technology-statistics-and-market-share?utm_source=chatgpt.com [accessed 2025-02-02]
6. Shocking amount of americans unaware of benefits of personal medical devices:poll. New York Post. URL: https://nypost.com/2024/10/28/lifestyle/shocking-amount-of-american-unaware-of-benefits-of-personal-medical-devices/?utm_source=chatgpt.com [accessed 2025-02-02]
7. Jolliff A, Holden RJ, Valdez R, et al. Investigating the best practices for engagement in remote participatory design: mixed methods analysis of 4 remote studies with family caregivers. J Med Internet Res 2024 Dec 3;26:e60353. [doi: 10.2196/60353] [Medline: 39626228]
8. Faisal M, Alharbi A, Alhamadi A, et al. Robot-based solution for helping Alzheimer patients. SLAS Technol 2024 Jun;29(3):100140. [doi: 10.1016/j.slast.2024.100140] [Medline: 38729525]
9. Frere S, Slutsky I. Alzheimer's disease: from firing instability to homeostasis network collapse. Neuron 2018 Jan 3;97(1):32-58. [doi: 10.1016/j.neuron.2017.11.028] [Medline: 29301104]
10. Fareed MM, Qasmi M, Aziz S, Völker E, Förster CY, Shityakov S. The role of clusterin transporter in the pathogenesis of Alzheimer's disease at the blood-brain barrier interface: a systematic review. Biomolecules 2022 Oct 10;12(10):1452. [doi: 10.3390/biom12101452] [Medline: 36291661]
11. Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of dementia: an observational study of patterns in diagnosis and awareness in US older adults. J Gen Intern Med 2018 Jul;33(7):1131-1138. [doi: 10.1007/s11606-018-4377-y] [Medline: 29508259]
12. Alder G, Taylor D, Rashid U, et al. A brain computer interface neuromodulatory device for stroke rehabilitation: iterative user-centered design approach. JMIR Rehabil Assist Technol 2023 Dec 11;10:e49702. [doi: 10.2196/49702] [Medline: 38079202]
13. He Y, Tang Z, Sun G, et al. Effectiveness of a mindfulness meditation app based on an electroencephalography-based brain-computer interface in radiofrequency catheter ablation for patients with atrial fibrillation: pilot randomized controlled trial. JMIR Mhealth Uhealth 2023 May 3;11:e44855. [doi: 10.2196/44855] [Medline: 37133926]
14. Mridha MF, Das SC, Kabir MM, Lima AA, Islam MR, Watanobe Y. Brain-computer interface: advancement and challenges. Sensors (Basel) 2021 Aug 26;21(17):5746. [doi: 10.3390/s21175746] [Medline: 34502636]
15. Mudgal SK, Sharma SK, Chaturvedi J, Sharma A. Brain computer interface advancement in neurosciences: applications and issues. Interdiscip Neurosurg 2020 Jun;20:100694. [doi: 10.1016/j.inat.2020.100694]
16. Värbu K, Muhammad N, Muhammad Y. Past, present, and future of EEG-based BCI applications. Sensors (Basel) 2022 Apr 26;22(9):3331. [doi: 10.3390/s22093331] [Medline: 35591021]
17. Kubben P. Invasive brain-computer interfaces: a critical assessment of current developments and future prospects. JMIR Neurotech ;3:e60151. [doi: 10.2196/60151]
18. Musk E, Neuralink. An integrated brain-machine interface platform with thousands of channels. J Med Internet Res 2019 Oct 31;21(10):e16194. [doi: 10.2196/16194] [Medline: 31642810]
19. Kale M, Wankhede N, Pawar R, et al. AI-driven innovations in Alzheimer's disease: integrating early diagnosis, personalized treatment, and prognostic modelling. Ageing Res Rev 2024 Nov;101:102497. [doi: 10.1016/j.arr.2024.102497] [Medline: 39293530]
20. Zhang H, Jiao L, Yang S, et al. Brain–computer interfaces: the innovative key to unlocking neurological conditions. Int J Surg 2024;110(9):5745-5762. [doi: 10.1097/JS9.0000000000002022]
21. Awuah WA, Ahluwalia A, Darko K, et al. Bridging minds and machines: the recent advances of brain-computer interfaces in neurological and neurosurgical applications. World Neurosurg 2024 Sep;189:138-153. [doi: 10.1016/j.wneu.2024.05.104] [Medline: 38789029]
22. Wang Y, Liu S, Wang H, Zhao Y, Zhang XD. Neuron devices: emerging prospects in neural interfaces and recognition. Microsyst Nanoeng 2022;8(1):1-13. [doi: 10.1038/s41378-022-00453-4]
23. Khademi Z, Ebrahimi F, Kordy HM. A review of critical challenges in MI-BCI: from conventional to deep learning methods. J Neurosci Methods 2023 Jan 1;383:109736. [doi: 10.1016/j.jneumeth.2022.109736] [Medline: 36349568]
24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
25. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021:n160. [doi: 10.1136/bmj.n160]
26. Shanechi MM. Brain-machine interfaces from motor to mood. Nat Neurosci 2019 Oct;22(10):1554-1564. [doi: 10.1038/s41593-019-0488-y] [Medline: 31551595]
27. Gu X, Cao Z, Jolfaei A, et al. EEG-based brain-computer interfaces (BCIs): a survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. IEEE/ACM Trans Comput Biol Bioinform 2021;18(5):1645-1666. [doi: 10.1109/TCBB.2021.3052811] [Medline: 33465029]

28. Tsai CW, Zhang M, Zhang L, Yoo J. A closed-loop brain-machine interface with one-shot learning and online tuning for patient-specific neurological disorder treatment. Presented at: 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS); Incheon, Korea, Republic of. [doi: 10.1109/AICAS54282.2022.9870001]
29. Mughal NE, Khan MJ, Khalil K, et al. EEG-fNIRS-based hybrid image construction and classification using CNN-LSTM. Front Neurorobot 2022;16:873239. [doi: 10.3389/fnbot.2022.873239] [Medline: 36119719]
30. Shin U, Ding C, Zhu B, et al. NeuralTree: A 256-Channel 0.227-µJ/Class versatile neural activity classification and closed-loop neuromodulation SoC. IEEE J Solid-State Circuits 2022;57(11):3243-3257. [doi: 10.1109/jssc.2022.3204508] [Medline: 36744006]
31. Wang X, Li S, Jin X. Synergizing optogenetics and artificial intelligence: a novel paradigm for advanced neuroscience investigationsai and optogenetics: advancing neuroscience. 2023 Oct 20 Presented at: ISAIMS 2023; Chengdu China. [doi: 10.1145/3644116.3644288]
32. Wu D, Lu BL, Hu B, Zeng Z. Affective brain–computer interfaces (aBCIs): a tutorial. Proc IEEE 2023;111(10):1314-1332. [doi: 10.1109/JPROC.2023.3277471]
33. Liang KF, Kao JC. Deep learning neural encoders for motor cortex. IEEE Trans Biomed Eng 2020 Aug;67(8):2145-2158. [doi: 10.1109/TBME.2019.2955722] [Medline: 31765302]
34. Belkacem AN, Jamil N, Khalid S, Alnajjar F. On closed-loop brain stimulation systems for improving the quality of life of patients with neurological disorders. Front Hum Neurosci 2023;17:1085173. [doi: 10.3389/fnhum.2023.1085173] [Medline: 37033911]
35. Yue C. Privacy and ethical concerns of brain-computer interfaces. Presented at: 2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom); Kyoto, Japan p. 134-138. [doi: 10.1109/MetaCom57706.2023.00036]
36. Golshan HM, Hebb AO, Mahoor MH. LFP-Net: a deep learning framework to recognize human behavioral activities using brain STN-LFP signals. J Neurosci Methods 2020 Apr 1;335:108621. [doi: 10.1016/j.jneumeth.2020.108621] [Medline: 32027889]
37. Bryan MJ, Preston Jiang L, P N Rao R. Neural co-processors for restoring brain function: results from a cortical model of grasping. J Neural Eng 2023 Jun 1;20(3):036004. [doi: 10.1088/1741-2552/accaa9]
38. Jiang Y, Jessee W, Hoyng S, et al. Sharpening working memory with real-time electrophysiological brain signals: which neurofeedback paradigms work? Front Aging Neurosci 2022;14:780817. [doi: 10.3389/fnagi.2022.780817] [Medline: 35418848]
39. Merk T, Peterson V, Köhler R, Haufe S, Richardson RM, Neumann WJ. Machine learning based brain signal decoding for intelligent adaptive deep brain stimulation. Exp Neurol 2022 May;351:113993. [doi: 10.1016/j.expneurol.2022.113993] [Medline: 35104499]
40. Xavier Fidêncio A, Klaes C, Iossifidis I. Error-related potentials in reinforcement learning-based brain-machine interfaces. Front Hum Neurosci 2022;16:806517. [doi: 10.3389/fnhum.2022.806517] [Medline: 35814961]
41. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 6;10(5):e33735. [doi: 10.2196/33735] [Medline: 35522465]
42. Andrews A. Integration of augmented reality and brain-computer interface technologies for health care applications: exploratory and prototyping study. JMIR Form Res 2022 Apr 21;6(4):e18222. [doi: 10.2196/18222] [Medline: 35451963]
43. Kelly JT, Campbell KL, Gong E, Scuffham P. The internet of things: impact and implications for health care delivery. J Med Internet Res 2020 Nov 10;22(11):e20135. [doi: 10.2196/20135] [Medline: 33170132]
44. Lake K, Mc Kittrick A, Desselle M, et al. Cybersecurity and privacy issues in extended reality health care applications: scoping review. JMIR XR Spatial Comput 2024;1:e59409-e59409. [doi: 10.2196/59409]
45. Pisarchik AN, Maksimenko VA, Hramov AE. From novel technology to novel applications: comment on "An Integrated Brain-Machine Interface Platform With Thousands of Channels" by Elon Musk and Neuralink. J Med Internet Res 2019 Oct 31;21(10):e16356. [doi: 10.2196/16356] [Medline: 31674923]
46. Albers EA, Mikal J, Millenbah A, et al. The use of technology among persons with memory concerns and their caregivers in the United States during the COVID-19 pandemic: qualitative study. JMIR Aging 2022 Mar 17;5(1):e31552. [doi: 10.2196/31552] [Medline: 35134748]
47. Fan Q, DuBose L, Ory MG, et al. Financial, legal, and functional challenges of providing care for people living with dementia and needs for a digital platform: interview study among family caregivers. JMIR Aging 2023 Sep 5;6:e47577. [doi: 10.2196/47577] [Medline: 37526513]
48. Liberati G, Veit R, Kim S, et al. Development of a binary fmri-BCI for Alzheimer patients: a semantic conditioning paradigm using affective unconditioned stimuli. 2013 Presented at: Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. [doi: 10.1109/ACII.2013.157]
49. Liberati G, Dalboni da Rocha JL, van der Heiden L, et al. Toward a brain-computer interface for Alzheimer's disease patients by combining classical conditioning and brain state classification. J Alzheimers Dis 2012;31 Suppl 3(s3):S211-S220. [doi: 10.3233/JAD-2012-112129] [Medline: 22451316]

50. Parikh PM, Venniyoor A. Neuralink and brain-computer interface-exciting times for artificial intelligence. South Asian J Cancer 2024 Jan;13(1):63-65. [doi: 10.1055/s-0043-1774729] [Medline: 38721102]

Abbreviations

ACM: Association for Computing Machinery
AD: Alzheimer disease
ADRD: Alzheimer disease and related dementias
AI: artificial intelligence
BCI: brain-computer interface
CNN: convolutional neural network
DBS: deep brain stimulation
EEG: electroencephalography
fNIRS: functional near-infrared spectroscopy
GAN: generative adversarial network
ICA: Independent Component Analysis
LDA: linear discriminant analysis
MI: motor imagery
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RBM: Restricted Boltzmann Machine
RNN: recurrent neural network
SVM: Support Vector Machine
TL: transfer learning
TSNN: tree-structured neural network

Edited by J Shaikh-Mohammed; submitted 05.02.25; peer-reviewed by R Cajo, X Liu; revised version received 21.07.25; accepted 05.08.25; published 05.11.25.

Please cite as:
Williams C, Anik FI, Hasan MM, Rodriguez-Cardenas J, Chowdhury A, Tian S, He S, Sakib N
Advancing Brain-Computer Interface Closed-Loop Systems for Neurorehabilitation: Systematic Review of AI and Machine Learning Innovations in Biomedical Engineering
JMIR Biomed Eng 2025;10:e72218
URL: https://biomedeng.jmir.org/2025/1/e72218
doi:10.2196/72218

Review

Cardiac Repair and Regeneration via Advanced Technology: Narrative Literature Review

Yugyung Lee[1], PhD; Sushil Shelke[1], BS; Chi Lee[1], PhD

Division of Pharmacology and Pharmaceutics Sciences, School of Pharmacy, University of Missouri Kansas City, Kansas City, MO, United States

Corresponding Author:
Chi Lee, PhD
Division of Pharmacology and Pharmaceutics Sciences
School of Pharmacy
University of Missouri Kansas City
5000 Holmes St
Kansas City, MO, 64110
United States
Phone: 1 8162352408
Fax: 1 8162355990
Email: leech@umkc.edu

Abstract

Background: Cardiovascular diseases (CVDs) are the leading cause of death globally, and almost one-half of all adults in the United States have at least one form of heart disease. This review focused on advanced technologies, genetic variables in CVD, and biomaterials used for organ-independent cardiovascular repair systems.

Objective: A variety of implantable and wearable devices, including biosensor-equipped cardiovascular stents and biocompatible cardiac patches, have been developed and evaluated. The incorporation of those strategies will hold a bright future in the management of CVD in advanced clinical practice.

Methods: This study employed widely used academic search systems, such as Google Scholar, PubMed, and Web of Science. Recent progress in diagnostic and treatment methods against CVD, as described in the content, are extensively examined. The innovative bioengineering, gene delivery, cell biology, and artificial intelligence–based technologies that will continuously revolutionize biomedical devices for cardiovascular repair and regeneration are also discussed. The novel, balanced, contemporary, query-based method adapted in this manuscript defined the extent to which an updated literature review could efficiently provide research on the evidence-based, comprehensive applicability of cardiovascular devices for clinical treatment against CVD.

Results: Advanced technologies along with artificial intelligence–based telehealth will be essential to create efficient implantable biomedical devices, including cardiovascular stents. The proper statistical approaches along with results from clinical studies including model-based risk probability prediction from genetic and physiological variables are integral for monitoring and treatment of CVD risk.

Conclusions: To overcome the current obstacles in cardiac repair and regeneration and achieve successful therapeutic applications, future interdisciplinary collaborative work is essential. Novel cardiovascular devices and their targeted treatments will accomplish enhanced health care delivery and improved therapeutic efficacy against CVD. As the review articles contain comprehensive sources for state-of-the-art evidence for clinicians, these high-quality reviews will serve as a first outline of the updated progress on cardiovascular devices before undertaking clinical studies.

(JMIR Biomed Eng 2025;10:e65366) doi:10.2196/65366

KEYWORDS

advanced technologies; genetics; biomaterials; bioengineering; medical devices; implantable devices; wearables; cardiovascular repair and regeneration; cardiac care; cardiovascular disease

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally, accounting for an estimated 17.9 million deaths in 2019 according to a report from the World Health Organization. Almost one-half of all adults in the United States have at least one form of heart disease [1]. Myocardial infarction (MI) is caused by ischemia in the coronary artery, primarily due to blocked arteries resulting from atherosclerosis [2]. This blockage damages the myocardium, reducing its contractile capacity, which leads to a decreased ejection fraction and, ultimately, heart failure [3]. In the United States, one healthy heart becomes infarcted every 40 seconds [4].

Preserving tissue and cellular function is crucial for maintaining heart functionality. Numerous signaling pathways and genetic factors associated with MI survival have been periodically reviewed [5-7]. There is a growing emphasis on understanding the mechanisms involved in myocardial repair and regeneration [8]. Reports from organizations such as the Transnational Alliance for Regenerative Therapies in Cardiovascular Syndromes highlight the importance of these mechanisms. Key principles affecting reparative and regenerative potential include survival and protection, cell-cell communication, angiogenesis and vascularization, cardiomyogenesis, molecular regulation of the cell cycle and proliferation, inflammation reduction, and cardiac aging [7,9].

An increase in reactive oxygen species (ROS) is a hallmark of ischemic cardiomyopathy [10]. ROS, such as hydrogen peroxide (H_2O_2) and hydroxyl radicals, play a significant role in MI and can be considered ideal regulators for patients post-MI [11]. The concentration of H_2O_2 in healthy cells is about 0.02 mM, whereas intracellular concentrations above 0.1 mM induce oxidative stress and cell death [12,13]. Given that extracellular H_2O_2 concentrations can be 10 to 100 times higher than intracellular levels [14], careful monitoring of H_2O_2 levels in cells is essential for prevention and treatment. As ROS play an integral role in platelet aggregation and vasodilation, inhibitors of vasodilation and platelet aggregation are commonly adapted as a therapeutic means against MI [15].

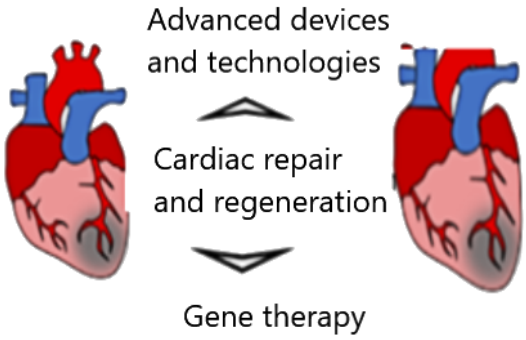
Regarding the treatment methods against CVD, organ transplant has been the most efficient strategy. Despite the preference for organ donor replacement in treating CVD, the shortage of organ donors has driven significant research into human-scale cardiovascular organs and functional tissue substitutes [16,17]. Challenges such as complex fabrication processes [18], poor mechanical properties [19], and biocompatibility and immunogenic issues [20] remain unresolved.

Designing prostheses requires fabricating matrix constructs with complex shapes and sizes for clinical applications [21]. Prostheses and implantable devices have varying requirements that are categorized into chemical, mechanical, electrical, and thermal characteristics [22]. Additionally, these devices must be biocompatible, be nonimmunogenic, and maintain functional capabilities within the body's biological environment [23]. Although serious infections or side effects from cardiovascular prostheses are rare, infected prostheses can be fatal [24].

Hydrogels, which are hydrophilic polymeric scaffolds with unique 3-dimensional structures, can absorb large amounts of water or biological fluids, making them potential candidates for cardiovascular tissue engineering [23]. Various synthetic and natural polymers are used in implantable hydrogels, with natural polymers like collagen offering higher immunity and biodegradable properties over synthetic ones.

This review focused on genetic variables in CVD, advanced technologies, and biomaterials for organ-independent cardiovascular repair systems (Figure 1). A variety of implantable and wearable devices, including biosensor-equipped cardiovascular stents and biocompatible cardiac patches, have been developed and evaluated. Finally, future research directions in the rapidly evolving fields of 3D-printed biomedical devices, artificial intelligence (AI), and multifunctional sensing devices are discussed.

Figure 1. Cardiac repair and regeneration via advanced technologies and gene therapy.



Advanced Assessment Technologies for Cardiac Image and Genetic Factors

Image Features Extracted From Imaging Modalities

Risk variables used for the classification of CVD progression include radiological imaging features and genetic factors. The complex nature of cardiovascular structures makes stenosis assessment from image modalities a serious challenge. In general, imaging features are considered radiomic-based biomarkers or indicators rather than pathological symptoms. An assessment of imaging features can serve as a quantitative index extractable from such imaging modalities as magnetic resonance imaging, computed tomography angiography (CTA), and optical coherence tomography [25]. Even though a semiquantitative estimation of coronary stenosis is feasible via

a thorough assessment of image features over an extended period, this process requires advanced technology expertise and labor-intensive effort.

In particular, coronary CTA, a noninvasive examination technique, plays an integral role in the evaluation and treatment of coronary artery disease (CAD) [26]. For instance, dual-source CTA allows for improved resolutions of implantable devices, including intrinsically higher-density stents, whose adversities are due to distortion reduction stemming from thick strut slices [27]. This approach makes it possible to conduct advanced cardiac imaging analysis, even though its invasive nature sometimes yields a high risk of fatality and complications [28-30].

As the number of images exponentially grows, the lack of ability to accurately label those images causes intrinsic limitations in the interpretation of the data [31]. A recent surge of AI techniques could serve as an ideal solution, enhancing the accuracy of a quantitative assessment of segmented features, including intima-media thickness ascertained by such computed algorithms as convolutional neural networks, UNet, UNet+, and DenseNet [32]. AI techniques and associated programmed models are for accurate identification of patterns, abnormalities, and defects in images, leading to enhanced efficiency and a reduction in errors inherent in human inspection [33].

Evolving Gene Therapy Against CVD

Genetic Factors in the Assessment of CVD Risk

Genes are involved in most cardiovascular functions, starting with the robustness of blood vessels to the way cells interact. People with a family history of heart disease could share common environmental factors, such as the intake habits of drinking water and daily food and exposure to chemicals, including carbon monoxide, in the air. As most cardiac disorders, including arrhythmias, congenital heart disease, cardiomyopathy, and high blood cholesterol, can be inherited [34], assessing genetic variants or biomarkers to identify at-risk individuals is integral to the prevention and treatment of CVD [35].

Genetic variations acquired by children from parents in the DNA of the eggs and sperm can influence every cell of a child's body, not only in the development process but also in the onset of heart disease [36]. An 8-year follow-up study found that CVD risk increased by 75% with a paternal history and about 60% with a maternal history of premature CVD, implying that certain genes can significantly enhance the risk of heart disease [37]. In the same study, a 16-year follow-up investigation found that a family history of premature CAD (age <50 years) marked a 44% higher risk of CVD mortality.

The pooled cohort equations for risk classification have been adapted based on genetic variants and medication decisions, including statins [38]. On the other hand, polygenic risk score (PRS) generation based on the relationships between the amount and frequency of genetic variants and the onset of specific diseases [39-41] has been explored for the assessment of genetic risk and extrapolation of individual outcomes [42]. The PRS could be accompanied by family history, lifestyle, and environmental factors [43,44] and fortified with emerging technologies, including proteomics, when determining an individual's genetic predisposition to CVD [45,46]. PRS mostly outperforms traditional risk scores in the prediction of individual outcomes, and additional AI-based transfer learning could further upgrade the relatively less accurate performance on translating PRS from ancestry to different ethnicities that are mostly unknown and unvalidated [47].

Genes that could reduce the development of plaque around infected regions would prevent neointimal formation [48]. The primary CVD endogenous biological variants include C-reactive protein, a liver protein released in response to inflammation [8,49], and plasma levels of low-density lipoprotein cholesterol [50], a seminal risk factor for the development of coronary heart disease. In addition, pro-inflammatory CD4+ cells with CD28 expression [49,51], cardiac troponin I [52], and the number of regulatory T lymphocytes [53] are frequently examined as specific biomarkers for the diagnosis of acute MI. Also, specific genes (eg, APOB, LDLR, and PCSK9 genes for familial hypercholesterolemia and BAG3, LMNA, MYH7, PLN, RBM20, SCN5A, TTN, TNNC1, TNNI3, TNNT2, and TPM1 genes for dilated cardiomyopathy) were recommended by the American Heart Association to be tested for the diagnosis of monogenic CVDs [54].

Along with those biological variants, pathological genetic factors or symptoms assessed for CVD include carotid intima thickness [55,56] and vascular function (which occur in the early stage of familial hypercholesterolemia) [57,58]. Detection of those genetic markers as part of familial cascade screening programs in familial hypercholesterolemia can lead to preventive effects, where subsequent medical therapy can lower long-term CVD risk [55,59]. A combined application of various genetic factors based on each patient's genetic profile may guarantee an efficient treatment strategy against CVD [35].

Even though genetic factors play a significant role in developing conditions of CVD, the screening processes including a health DNA test can only reveal certain genetic mutations that increase the risk and responses [60]. Subsequently, the relationship between genetic factors and risk scores is sometimes poor due to the fact that those having the genetic mutation do not necessarily have the same lifestyle factors, including basic health measures. Therefore, proper statistical approaches along with the results from clinical studies including model-based risk probability prediction from each or combined genetic variables are integral for genetic-based prediction of the CVD risk [61].

AI for Cardiovascular Gene Therapy

Genes (DNA, small interfering RNA, and microRNA) that could interfere with the development of plaque around infected regions are conjugated on biomedical devices like cardiovascular stents to prevent neointimal formation. An advanced monitoring process of genetic data and clinical data from electronic health records could lead to a fast and precise clinical decision and achieve customized treatment, eventually alleviating CVD via the detection of CVD symptoms at an early stage. However, the efficiency of cardiovascular gene therapy has been hampered by some obstacles, such as insufficient gene propagation, a lack of delivery mechanisms, and insufficient cell-vector interactions [62]. Moreover, health care providers may negatively influence

clinical outcomes due to the lack of discipline in the treatment algorithms and the absence of established regulations to handle early-onset data [63,64].

Combined AI models will address highly complicated cardiovascular clinical genetics [65]. AI profoundly apprehends complex patterns in imaging profiles and offers quantitative assessments of radiographic properties, serving as a valuable tool for enhancing imaging postprocessing. For instance, a combined convolutional neural network and recurrent neural network has achieved enhanced accuracy in predicting stenosis (≥50%) upon examining genetic variables grouped into training and testing samples [32,66]. This approach has obtained similar outcomes in the quantitative assessment of the growing number of segmented image features, including intima-media thickness for CVD [31,32].

In general, the advanced technology involved with AI is revolutionizing the method that ensures the accuracy, completeness, consistency, and validity of clinically applicable gene data [67]. In parallel, researchers should follow established guidance on using information from the digital world, as several guidelines have already been issued by institutional review boards to properly maintain genetic data integrity [68]. As a result of the increase in genetic testing and the fear of privacy breaches by health providers, employers, and society, the disciplines of ethics, public health, and genetics have also emerged. The health professional should make a compromise between providing proper arrangements for patient care and protecting personal privacy. In the near future, the adaptation of AI in radiomics will lead to precise and automated analysis of genetic variables involved with disease onset and progress.

Telehealth Genetic Counseling Between Patients and Genetic Counselors

To improve the efficacy of the diagnosis and assessment of CVD risk, the prediction tools, including telehealth systems, should assess endogenous genetic compounds involved with heart failure, atherosclerosis, and CAD [67,69]. Telehealth genetic counseling, including videoconferencing and telephone counseling, was compared with in-person genetic counseling for the degree of outcomes specific to patient experiences and accessibility to various treatment methods. The patients expressed the highest satisfaction with genetic counseling provided by media devices, such as telephone and video [70-72]. Moreover, telehealth genetic counseling is considered equitable to in-person genetic counseling across numerous domains, even though those studies were conducted with telehealth systems that were less robust and accurate than what is available today.

The benefits and limitations of telehealth from the perspectives of the patients and genetic counselors have been thoroughly examined to resolve potential uncertainty in the analysis processes [73-75]. Those limitations include technical challenges, difficulty in rapport and the subsequent psychosocial issues, and lack of clinical complement [74,76]. There needs to be some conceptual changes in the current status of telehealth approaches over time, providing continuous advancement in involved technologies [76,77].

Mobile Sensors for Cardiovascular Information Systems

Remote monitoring is considered the ambulatory tracing of vital signs and other medical indicators of a patient's health and recovery status via a telemedicine system without the patient meeting doctors or being present in the clinic (Figure 2) [78]. The Food and Drug Administration has recognized the importance of devices such as continuous temperature monitoring or continuous glucose monitoring devices that allow health care providers to remotely monitor patients, including those that measure body temperature, respiratory rate, heart rate, and blood pressure. In addition, a new approach based on advanced technologies for various physiological variables and biomarkers has performed continuous in-time monitoring as well as subsequent customized treatment strategies.

Figure 2. Schematic Representation of Remote Monitoring System of Biosensor/Cardiac Implantable Electronic Device.

The current roles of mobile sensors explored in telehealth technologies and further challenges in CVD will specifically emphasize (1) accurate assessment and diagnosis of vital signs or biomarkers from CVDs, (2) reliable and reproducible sensing systems to monitor the progress of a patient's disease status, and (3) wearable devices with maintenance of battery life and restoration of interaction sensitivity capable of assessing cardiovascular information of patients at risk [79-81].

The problem arises when analyzing data from mobile sensors due to a lack of normalization and implementation of proprietary interfaces to the respective device or platform. In daily life, numerous portals provided by each sensor manufacturer should be simultaneously traced and aggregated into the existing database for each cardiovascular patient [82]. Thus, the integration of data obtained from patients with heart failure or implantable cardiac devices needs to be properly conducted to store data in a structured and interoperable way for timely clinical and scientific evaluations [83,84].

Advanced Systems Currently Available for CVD

Biomaterials for Organ-Independent Cardiovascular Repair Systems

Required Properties for Organ-Independent Cardiovascular Repair Systems

The highly ordered myocardium capacity for electrical integrity and electrical conduction between healthy and infarcted cells starts to diminish as the relatively disordered fibrous scar tissue disposition increases in the myocardium, leading to systolic and diastolic dysfunction and cardiac arrhythmia [85]. As heart transplantation is limited due to a shortage of organ donors, organ-independent systems, including cardiac patches, grafts, and scaffolds, play an essential role in cardiac repair and treatment of MI [86].

Biomaterial systems function like normal cardiac tissues, providing excellent electrical conductivity, mechanical strength, and biological activities to infarcted heart tissues [87]. Novel biomaterial-based systems offer self-renewal and regeneration in the damaged heart, serving as various resources for cardiac tissue repair for those with CVD. For instance, cardiac patches provide mechanical support to the myocardial wall and passively prevent the infarcted myocardium following MI by reducing myocardial wall stress and preventing left ventricular dilation and remodeling [88].

Hydrogels for Organ-Independent Cardiovascular Repair Systems

Hydrogels are soft and moist injectable biomaterials with properties similar to those of human soft tissues. They are minimally invasive and serve as a vehicle for the delivery of therapeutic agents in situ [89,90]. Conductive hydrogel systems based on low-dimensional inorganic nanomaterials, such as carbon nanotubes and graphene derivatives [23], and simultaneously loaded with stem cells, growth factors, cytokines, or oligonucleotides, are found to alleviate cardiac casualties by promoting angiogenesis and cardiomyocyte proliferation and reducing fibrosis and apoptosis.

In addition, a complex hydrogel patch is produced by principles of fabrication via Fe^{+3}-induced ionic coordination between a homogeneous network of dopamine-gelatin conjugates and dopamine-functionalized polypyrrole [91]. The Schiff base reaction between oxidized sodium hyaluronic acid and hydrazided hyaluronic acid was explored to form an injectable hydrogel patch. Added bioactive peptides, a 7-amino acid peptide, loaded in collagen-based hydrogel reduced cell apoptosis, enhanced Sca-1+ recruitment and differentiation of stem cells, and enhanced neovascularization formation, which resulted in improved heart function in a mouse MI model [90].

Cardiac Patch

Therapeutic Patch as an Effective Strategy

All the delivery methods for MI recovery drugs, primarily via the oral route but occasionally via an intravenous route, direct injection to the heart, and drug-eluting stents, have their own limitations in resolving MI-induced loss of cardiomyocytes [92]. Advanced formulations, including cardiac patches, have demonstrated their efficiencies in functional recovery for drug carriers with targeted and local delivery of cardiovascular drugs, nutrients, and cells. Moreover, patches not only are capable of providing necessary mediators in multiple therapies to recover the affected area but also strengthen the damaged area with induced cell attachment and proliferation [93].

Types of Patches and Their Applications for MI Recovery

Therapeutic patches are divided into two types based on the presence or absence of cells: cell-based patches and acellular patches. As there is a lack of regeneration of cardiomyocytes, cells such as human-induced pluripotent stem cells, mesenchymal stem cells, and skeletal myoblasts are often introduced to restore cardiac function [94].

Newly introduced cells can lead to enhanced angiogenesis, lowered fibrosis, and apoptosis of cardiomyocytes [2]. Due to the inefficiency of generating new heart tissue from cardiomyocytes, acellular cardiac patches, which might include paracrine factors such as proteins, RNA, growth factors, or small molecules, are occasionally explored to accomplish cardioprotective effects [95].

The biocompatibility of the source biomaterial often entails a serious challenge in designing any implantable patches [96]. Moreover, the biomaterial should be similar to that of host tissues from the perspectives of biochemical, mechanical, and topographical properties [97,98]. For instance, poly(hydroxyethyl) methacrylate (pHEMA) polymer has demonstrated biocompatibility and has been used for biomedical applications, including drug delivery [99,100], contact lenses [101,102], and tissue engineering [103,104]. However, the low viscous nature of pHEMA makes it a challenging task to develop pHEMA-based biomedical devices, including a cardiac patch that is capable of successfully delivering agents like ROS scavengers against MI.

3D Printing Technology for Cardiac Patch Development

3D printing can be used to create patient-specific devices, such as organ implants and tissue models that mimic human physiology. 3D printing can generate surgical planning models and reduce the need for animal testing. 3D printing can be used to create personalized medicines and their delivery systems that specifically adapt to each patient's genetic makeup [105].

There are numerous methods, including electrospinning, solvent evaporation, and decellularization, used for the development of patches [106]. Each of these methods has its own challenges, such as material selectivity, limitations in complex shapes, and cost and time efficiency [107]. Additionally, 3D printing has emerged as a low-cost and fast method to develop patches produced from a vast range of materials with the utmost efficacy.

As previously mentioned, a novel patch based on biocompatible pHEMA polymers was developed with the aid of direct-light 3D printing technology. Stereolithography-based 3D printing, where the ink is placed on a platform, was successfully used to prepare 3D-printed acellular cardiac patches or cardiovascular stents [21]. In 3D-printed systems, the immunosuppressive drug, like sirolimus, dispersed within the patch matrix will be released when the linker, like an ROS-responsive thioketal linker, that connects the polymers is cleaved [108]. The ratio of the polymer and crosslinker can be customized to achieve controllable drug release.

3D- or 4D-Printable Smart Devices for CVD

3D printing provides geometric flexibility, which has been explored to produce metal or polymer-embedded 3D construct microsystems with high flexibility [23,109]. 3D-printed systems or smart devices use advanced materials with characteristics such as thermal and electrical conductivity and piezo-resistivity [110]. Electric units or components, including resistors, capacitors, inductors, circuits, and passive wireless sensors and batteries, have been incorporated into 3D-printed products for potential practical applications.

3D tactile sensors capable of detection and differentiation of human movements, including pulse monitoring and finger motions via detection of endogenous compounds, were fabricated using multimaterial, multiscale, and multifunctional approaches under ambient conditions conformally onto freeform surfaces [111]. As lactate levels have been associated with heart failure as well as diabetes, the portable luminometer, a disposable minicartridge produced by 3D printing and stored in cell phones, was used to detect chemiluminescence from enzyme-coupled reactions [112]. Lactate oxidase was coupled with horseradish peroxidase to noninvasively detect the lactate levels within 5 minutes at a detection limit of 0.5 mM/L and 0.1 mM/L in oral fluids and sweat, respectively.

By adapting AI to additive manufacturing, 3D designers can optimize cardiovascular biosensors or implants to be more efficient and robust. AI-mediated 3D printing tools can synchronize with high-quality imaging data, such as computed tomography and magnetic resonance imaging scans, and generate personalized designs, enabling thorough control over the otherwise unavoidably complicated, time-consuming, and exhaustive process [3].

An optimal combination of 3D printing based on novel or hybrid 3D printing methods and AI can achieve the next generation of cardiovascular systems [113]. Subsequently, advanced 3D or 4D printing, once nearly overcoming the cost and scalability barriers, could lead to more effective and targeted treatments against CVD, accomplishing improved treatment outcomes and enhanced health care delivery [67].

Advanced Cardiovascular Stents for CVD

Gene-Eluting Stents

Advanced biomedical gene carriers have been intensively explored in vascular cell biology and CVD treatment. The identification of critical regulators, such as noncoding RNAs, including microRNA, long noncoding RNA, and circular RNA presence in such cell types as vascular smooth muscle cells, endothelial cells, and macrophages, has served as an efficient therapeutic target in the field of CVD.

Among biomedical carriers, multifunctional gene-loaded stents and integrated stents equipped with self-reporting sensors are often explored as promising technologies against CVD, including atherosclerosis and MI [114,115]. Cardiovascular stents keep the vessel open and prevent it from re-occluding (ie, restenosis), but vessel injury by stent struts leads to the activation of platelets and mural thrombus formation, leading to the activation of circulating neutrophils and tissue macrophages [116-118]. As the cardiovascular stent produces late-stage restenosis [119,120], people with stents are at risk of high blood pressure. Therefore, it is integral to find a more advanced and sensitive stent capable of real-time monitoring of blood flow.

Gene-loaded stents coated with synthetic and natural polymers such as polylactic-polyglycolic acid (PLGA), collagen, hydroxyapatite, and matrix metalloproteins can overcome major limitations of cardiovascular gene therapy, including insufficient cell-vector interactions, a lack of delivery mechanisms, and insufficient gene propagation [121]. Gene-loaded stents also allow for maintaining a curative gene, serving as a carrier to convey the gene and administer the vector and avoiding immune responses [62].

The first successful in vivo transfection of green fluorescent protein plasmid DNA loaded in a DNA-PLGA coated stent was efficiently expressed in cell cultures (7.9%, SD 0.7% vs 0.6%, SD 0.2% control; $P<.001$) of rat aortic smooth muscle cells [122]. In addition, PLGA nanoparticle-coated stents encapsulated with vascular endothelial growth factor and paclitaxel [123] or Ang-1 proteins [124] were developed as an alternative therapy, reducing in-stent restenosis and accomplishing complete re-endothelialization. In addition, an Akt1 small interfering RNA–embedded stent alleviated restenosis, reducing cell growth via muting RNA [125,126]. Furthermore, bare-metal stents with a synthetic complex for reversible vector binding produced prominent green fluorescent protein positivity in A10 cells proximal to the strut after 72 hours in culture [127].

A collagen-coated stent covalently coupled with anti-DNA immunoglobulin M antibody and loaded with plasmid DNA was efficiently developed for localized gene delivery to smooth

muscle cells in an artery, accomplishing high-level protein production through reporter gene expression [125]. In addition, a stent coated with biomimetic hyaluronic acid and deposited with DNA/polyethylenimine polyplexes was explored to deliver plasmid DNA to the artery, exerting its efficacy in alleviating restenosis with a higher neointimal transfection rate while maintaining structural stability [128].

Stents Equipped With Cardiovascular Self-Reporting Sensors

Continuous blood flow surveillance can serve as screening, advanced detection, and alert for cardiovascular health using noninvasive technology that can be placed in the coronary arteries [129]. Remote monitoring of patient progress is feasible by creating an application-specific integrated circuit that features a voltage regulator and radio frequency power element loaded in biomedical devices, including cardiovascular stents.

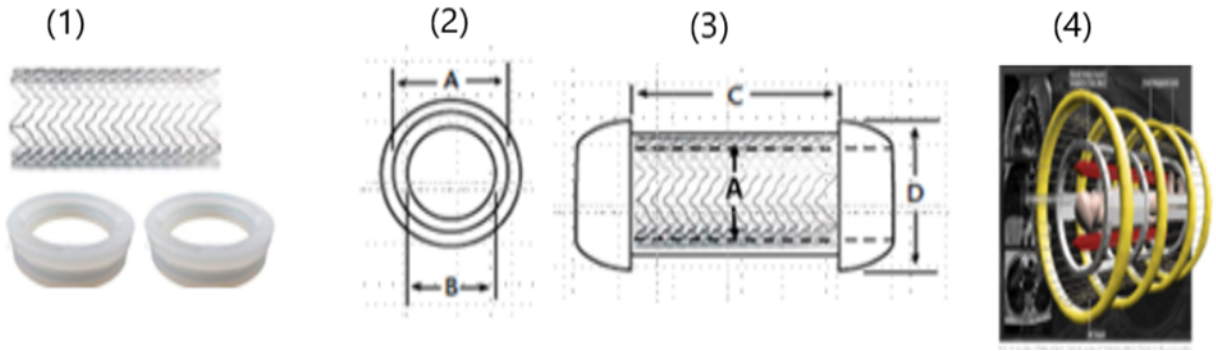
For instance, a remote monitoring stent was combined with a tiny heart pressure sensor as well as a wireless transmitter that continuously monitors vascular conditions and the status of implanted devices. To minimize the number of antenna components for the conservation of space, the stent was used as an inductive antenna to create a wireless network [130,131], transmitting quantified solubilization to the immediate neighborhood via a wireless telemetry transmitter [132]. Reviewing the admittance of an antenna close to the implant component and connected to it via electromagnetic coupling will enable this function [133]. A radio frequency–powering component was implanted on the chip in the finished device as an ideal power distribution feature. Microelectromechanical modules were crammed with an application-specific integrated circuit for data collection [134].

As shown in Figure 3, a blood flow sensor enclosed in graphene-embedded silicon rings subsequently equipped with a digital wireless transmission microchip was developed as a unit of the smart theranostic cardiovascular stent (Figure 3C). Numerous commercial devices, including pressure sensors, use the piezoresistive effect of silicon, whose gauge factors can be 2 orders of magnitude larger than those observed in most metals [135,136]. Thus, a flow sensor enclosed in the stent was able to continuously monitor real-time blood flow with high inductance and pressure resolution and transmit corresponding data to a cardiologist outside the body. In addition to superior moisture barrier property, the high thermal conductivity of graphene (which has a negative thermal expansion coefficient [-8.0×10^{-6}/K] between 0 and 700 K) guaranteed dimensional stability upon exposure to body temperature and continuous blood flow.

The pressure sensors and the microchip were mounted on the rectangular areas of the stent structure, as shown in Figure 3C. The pressure sensors bound to the steel stent [137,138] were molded into graphene-embedded silicone rings, and the pattern was cut on a thin stainless-steel foil. These digital transmission techniques reduced the power radiated by the external reader, thus minimizing the patient's exposure to electromagnetic fields.

In electromagnetic coupling, a continuous electromagnetic wave with relatively large power is radiated by the reader, and the microchip modulates the impedance of the antenna by connecting or disconnecting a load to it according to the data to be transmitted [130].

Figure 3. (1) stent and rings, (2) dimension and size of rings (A: Ring inner diameter (i.e., same as stent outer diameter); B: Stent inner diameter), (3) 3-Dimentional Stent (C: Stent length; D: Ring outer diameter) and (4) application of external electromagnetic stimuli.



Conclusion

Biotechnologies play an important role in cardiovascular repair and regeneration. Genetic variables in CVD, currently available technology, and biomaterials for organ-independent cardiovascular repair systems were updated in this article in a timely manner. Advanced biotechnologies aimed at target-specific therapeutics are designed for customized and personalized cardiac treatment strategies with one or multiple administration routes whose methods should be further improved to enhance targeting and treatment efficacy.

The goal of gene therapy for cardiac repair and regeneration is to achieve cardiac transfection outcomes via the selection of proper gene vectors and modifying a gene or genetic pathways. Moreover, 3D bioprinting technology has been widely used in cardiac repair by integrating biomaterials with various manufacturing processes to customize cardiac conditions. 3D scaffolds with varying cell types have demonstrated better biocompatibility, delivery efficiency, and low immunogenicity. In the future, screening and designing of viral vectors through structure evolution mediated by 3D printing would enhance cardiac gene therapy.

To overcome the current obstacles in cardiac repair and regeneration and achieve successful therapeutic applications, future interdisciplinary collaborative work should be integral. Advanced new material and cell biology, along with AI-based telehealth, will be essential to create efficient implantable biomedical devices, including cardiovascular stents. Advanced innovative bioengineering, gene delivery, and cell biology technologies will continuously revolutionize medical devices for cardiovascular repair and regeneration in the future.

Acknowledgments

This study was supported by a Funding for Excellence award from the Office of Research Development, University of Missouri-Kansas City.

Conflicts of Interest

None declared.

References

1. Martin S, Aday A, Almarzooq Z, Anderson C, Arora P, Avery C, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. 2024 heart disease and stroke statistics: a report of US and global data from the American Heart Association. Circulation 2024 Feb 20;149(8):e347-e913 [FREE Full text] [doi: 10.1161/CIR.0000000000001209] [Medline: 38264914]
2. Braunwald E. Cell-based therapy in cardiac regeneration. Circulation Research 2018 Jul 06;123(2):132-137. [doi: 10.1161/circresaha.118.313484]
3. Zhu Z, Ng DWH, Park HS, McAlpine MC. 3D-printed multifunctional materials enabled by artificial-intelligence-assisted fabrication technologies. Nat Rev Mater 2020 Oct 12;6(1):27-47. [doi: 10.1038/s41578-020-00235-2]
4. Tsao C, Aday A, Almarzooq Z, Alonso A, Beaton A, Bittencourt M, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
5. Kolwicz SC, Purohit S, Tian R. Cardiac metabolism and its interactions with contraction, growth, and survival of cardiomyocytes. Circulation Research 2013 Aug 16;113(5):603-616. [doi: 10.1161/circresaha.113.302095]
6. O'Neal WT, Griffin WF, Kent SD, Virag JAI. Cellular pathways of death and survival in acute myocardial infarction. J Clin Exp Cardiolog 2012;S6(003):1-9. [doi: 10.4172/2155-9880.S6-003]
7. Broughton KM, Wang BJ, Firouzi F, Khalafalla F, Dimmeler S, Fernandez-Aviles F, et al. Mechanisms of cardiac repair and regeneration. Circ Res 2018 Apr 13;122(8):1151-1163. [doi: 10.1161/circresaha.117.312586]
8. Anzai T, Yoshikawa T, Shiraki H, Asakura Y, Akaishi M, Mitamura H, et al. C-reactive protein as a predictor of infarct expansion and cardiac rupture after a first Q-wave acute myocardial infarction. Circulation 1997 Aug 05;96(3):778-784. [doi: 10.1161/01.cir.96.3.778] [Medline: 9264482]
9. Chingale M, Zhu D, Cheng K, Huang K. Bioengineering technologies for cardiac regenerative medicine. Front Bioeng Biotechnol 2021 Jun 3;9:681705 [FREE Full text] [doi: 10.3389/fbioe.2021.681705] [Medline: 34150737]
10. Bolli R. Oxygen-derived free radicals and postischemic myocardial dysfunction ("stunned myocardium"). J Am Coll Cardiol 1988 Jul;12(1):239-249 [FREE Full text] [doi: 10.1016/0735-1097(88)90381-6] [Medline: 3288676]
11. Hori M, Nishida K. Oxidative stress and left ventricular remodelling after myocardial infarction. Cardiovasc Res 2009 Feb 15;81(3):457-464. [doi: 10.1093/cvr/cvn335] [Medline: 19047340]
12. Sies H. Hydrogen peroxide as a central redox signaling molecule in physiological oxidative stress: Oxidative eustress. Redox Biol 2017 Apr;11:613-619 [FREE Full text] [doi: 10.1016/j.redox.2016.12.035] [Medline: 28110218]
13. Weng X, Tan H, Huang Z, Chen J, Zhang N, Wang Q, et al. Targeted delivery and ROS-responsive release of Resolvin D1 by platelet chimeric liposome ameliorates myocardial ischemia-reperfusion injury. J Nanobiotechnology 2022 Oct 20;20(1):454 [FREE Full text] [doi: 10.1186/s12951-022-01652-x] [Medline: 36266658]
14. Forman HJ, Bernardo A, Davies KJ. What is the concentration of hydrogen peroxide in blood and plasma? Arch Biochem Biophys 2016 Aug 01;603:48-53. [doi: 10.1016/j.abb.2016.05.005] [Medline: 27173735]
15. Heusch G, Gersh BJ. The pathophysiology of acute myocardial infarction and strategies of protection beyond reperfusion: a continual challenge. Eur Heart J 2017 Mar 14;38(11):774-784. [doi: 10.1093/eurheartj/ehw224] [Medline: 27354052]
16. D'Amore A, Yoshizumi T, Luketich SK, Wolf MT, Gu X, Cammarata M, et al. Bi-layered polyurethane - extracellular matrix cardiac patch improves ischemic ventricular wall remodeling in a rat model. Biomaterials 2016 Nov;107:1-14. [doi: 10.1016/j.biomaterials.2016.07.039] [Medline: 27579776]
17. Sebastião MJ, Gomes-Alves P, Reis I, Sanchez B, Palacios I, Serra M, et al. Bioreactor-based 3D human myocardial ischemia/reperfusion in vitro model: a novel tool to unveil key paracrine factors upon acute myocardial infarction. Transl Res 2020 Jan;215:57-74 [FREE Full text] [doi: 10.1016/j.trsl.2019.09.001] [Medline: 31541616]
18. Kim M, Choe Y, Kim G. Injectable hierarchical micro/nanofibrous collagen-based scaffolds. Chemical Engineering Journal 2019 Jun;365:220-230. [doi: 10.1016/j.cej.2019.02.044]

19. Hou S, Niu X, Li L, Zhou J, Qian Z, Yao D, et al. Simultaneous nano- and microscale structural control of injectable hydrogels via the assembly of nanofibrous protein microparticles for tissue regeneration. Biomaterials 2019 Dec;223:119458. [doi: 10.1016/j.biomaterials.2019.119458] [Medline: 31491598]
20. Bian T, Zhao K, Meng Q, Tang Y, Jiao H, Luo J. The construction and performance of multi-level hierarchical hydroxyapatite (HA)/collagen composite implant based on biomimetic bone Haversian motif. Materials Design 2019 Jan;162:60-69. [doi: 10.1016/j.matdes.2018.11.040]
21. Veerubhotla K, Lee CH. Design of biodegradable 3D-printed cardiovascular stent. Bioprinting 2022 Jun;26:e00204. [doi: 10.1016/j.bprint.2022.e00204]
22. Ravindran S. Materials for Advanced Packaging (Lu, D. and Wong, C.P.) [Book reviews]. IEEE Nanotechnology Mag 2009 Sep;3(3):27-28. [doi: 10.1109/mnano.2009.934216]
23. Veerubhotla K, McGraw H, Lee CH. Assessment of biocompatibility of 3D‐printed cardiovascular stent. Adv Ther 2023 Mar 15;6(5):1. [doi: 10.1002/adtp.202200292]
24. Sugarman B, Young EJ. Infections associated with prosthetic devices: magnitude of the problem. Infect Dis Clin North Am 1989 Jun;3(2):187-198 [FREE Full text] [doi: 10.1016/S0891-5520(20)30257-9]
25. Hamon M, Champ-Rigot L, Morello R, Riddell JW, Hamon M. Diagnostic accuracy of in-stent coronary restenosis detection with multislice spiral computed tomography: a meta-analysis. Eur Radiol 2008 Feb 1;18(2):217-225. [doi: 10.1007/s00330-007-0743-6] [Medline: 17763854]
26. Dai T, Wang J, Hu P. Diagnostic performance of computed tomography angiography in the detection of coronary artery in-stent restenosis: evidence from an updated meta-analysis. Eur Radiol 2018 Apr 9;28(4):1373-1382. [doi: 10.1007/s00330-017-5097-0] [Medline: 29124384]
27. Zimmerman SL, Kral BG, Fishman EK. Diagnostic Quality of Dual-Source Coronary CT Examinations Performed Without Heart Rate Control. J Computer Assisted Tomogr 2014;38(6):949-955. [doi: 10.1097/rct.0000000000000135]
28. Flohr TG, McCollough CH, Bruder H, Petersilka M, Gruber K, Süss C, et al. First performance evaluation of a dual-source CT (DSCT) system. Eur Radiol 2006 Feb 10;16(2):256-268. [doi: 10.1007/s00330-005-2919-2] [Medline: 16341833]
29. Liu W, Li G, Liu H, Lei J. Diagnostic accuracy of dual-source computed tomography angiography for the detection of coronary in-stent restenosis: A systematic review and meta-analysis. Echocardiography 2018 Apr 23;35(4):541-550. [doi: 10.1111/echo.13863] [Medline: 29569751]
30. Carrabba N, Schuijf JD, de Graaf FR, Parodi G, Maffei E, Valenti R, et al. Diagnostic accuracy of 64-slice computed tomography coronary angiography for the detection of in-stent restenosis: a meta-analysis. J Nucl Cardiol 2010 Jun;17(3):470-478 [FREE Full text] [doi: 10.1007/s12350-010-9218-2] [Medline: 20379863]
31. Byrne RA, Serruys PW, Baumbach A, Escaned J, Fajadet J, James S, et al. Report of a European Society of Cardiology-European Association of Percutaneous Cardiovascular Interventions task force on the evaluation of coronary stents in Europe: executive summary. Eur Heart J 2015 Oct 07;36(38):2608-2620 [FREE Full text] [doi: 10.1093/eurheartj/ehv203] [Medline: 26071600]
32. Tang X. The role of artificial intelligence in medical imaging research. BJR Open 2020 Nov;2(1):20190031 [FREE Full text] [doi: 10.1259/bjro.20190031] [Medline: 33178962]
33. Valente J, António J, Mora C, Jardim S. Developments in image processing using deep learning and reinforcement learning. J Imaging 2023 Sep 30;9(10):207 [FREE Full text] [doi: 10.3390/jimaging9100207] [Medline: 37888314]
34. Kolber MR, Scrimshaw C. Family history of cardiovascular disease. Can Fam Physician 2014 Nov;60(11):1016 [FREE Full text] [Medline: 25392442]
35. Tahir UA, Gerszten RE. Molecular biomarkers for cardiometabolic disease: risk assessment in young individuals. Circulation Research 2023 Jun 09;132(12):1663-1673. [doi: 10.1161/circresaha.123.322000]
36. Inherited Cardiac Conditions (Genetic Disorders). University of Ottawa Heart Institute. URL: https://www.ottawaheart.ca/heart-condition/inherited-cardiac-conditions-genetic-disorders [accessed 2025-02-21]
37. Lloyd-Jones DM, Nam B, D'Agostino RB, Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA 2004 May 12;291(18):2204-2211. [doi: 10.1001/jama.291.18.2204] [Medline: 15138242]
38. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Jun 18;139(25):e1082-e1143 [FREE Full text] [doi: 10.1161/CIR.0000000000000625] [Medline: 30586774]
39. Márquez-Luna C, Loh P, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol 2017 Dec 07;41(8):811-823 [FREE Full text] [doi: 10.1002/gepi.22083] [Medline: 29110330]
40. Widen E, Raben TG, Lello L, Hsu SDH. Machine learning prediction of biomarkers from SNPs and of disease risk from biomarkers in the UK Biobank. Genes (Basel) 2021 Jun 29;12(7):991 [FREE Full text] [doi: 10.3390/genes12070991] [Medline: 34209487]

41. Collister JA, Liu X, Clifton L. Calculating polygenic risk scores (PRS) in UK Biobank: a practical guide for epidemiologists. Front Genet 2022 Feb 18;13:818574 [FREE Full text] [doi: 10.3389/fgene.2022.818574] [Medline: 35251129]
42. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, UK Biobank CardioMetabolic Consortium CHD Working Group. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. J Am Coll Cardiol 2018 Oct 16;72(16):1883-1893 [FREE Full text] [doi: 10.1016/j.jacc.2018.07.079] [Medline: 30309464]
43. Aung N, Vargas JD, Yang C, Cabrera CP, Warren HR, Fung K, et al. Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. Circulation 2019 Oct 15;140(16):1318-1330. [doi: 10.1161/circulationaha.119.041161]
44. Axelrud LK, Santoro ML, Pine DS, Talarico F, Gadelha A, Manfro GG, et al. Polygenic risk score for Alzheimer's disease: implications for memory performance and hippocampal volumes in early life. Am J Psychiatry 2018 Jun 01;175(6):555-563 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17050529] [Medline: 29495896]
45. Gola D, Erdmann J, Müller-Myhsok B, Schunkert H, König IR. Polygenic risk scores outperform machine learning methods in predicting coronary artery disease status. Genet Epidemiol 2020 Mar 10;44(2):125-138. [doi: 10.1002/gepi.22279] [Medline: 31922285]
46. Hajar R. Genetics in cardiovascular disease. Heart Views 2020;21(1):55. [doi: 10.4103/heartviews.heartviews_140_19]
47. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019 Apr 29;51(4):584-591 [FREE Full text] [doi: 10.1038/s41588-019-0379-x] [Medline: 30926966]
48. Efovi D, Xiao Q. Noncoding RNAs in vascular cell biology and restenosis. Biology (Basel) 2022 Dec 22;12(1):24 [FREE Full text] [doi: 10.3390/biology12010024] [Medline: 36671717]
49. Padua L, Cuccagna C, Pazzaglia C. Novel sensory paradigms for neuromodulation in disorders of consciousness in traumatic brain injury. Curr Opin Neurol 2019 Dec;32(6):844-849. [doi: 10.1097/WCO.0000000000000747] [Medline: 31567499]
50. Ference B, Ginsberg H, Graham I, Ray K, Packard C, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J 2017 Aug 21;38(32):2459-2472 [FREE Full text] [doi: 10.1093/eurheartj/ehx144] [Medline: 28444290]
51. Sato K, Kaikita K, Nakayama N, Horio E, Yoshimura H, Ono T, et al. Coronary vasomotor response to intracoronary acetylcholine injection, clinical features, and long‐term prognosis in 873 consecutive patients with coronary spasm: analysis of a single‐center study over 20 years. JAHA 2013 Aug 22;2(4):1. [doi: 10.1161/jaha.113.000227]
52. Eshlaghi SN, Syedmoradi L, Amini A, Omidfar K, Omidfar K. A label-free electrochemical aptasensor based on screen printed carbon electrodes with gold nanoparticles-polypyrrole composite for detection of cardiac troponin I. IEEE Sensors J 2023 Feb 15;23(4):3439-3445. [doi: 10.1109/jsen.2023.3235740]
53. Montone R, Niccoli G, Fracassi F, Russo M, Gurgoglione F, Cammà G, et al. Patients with acute myocardial infarction and non-obstructive coronary arteries: safety and prognostic relevance of invasive coronary provocative tests. Eur Heart J 2018 Jan 07;39(2):91-98. [doi: 10.1093/eurheartj/ehx667] [Medline: 29228159]
54. Musunuru K, Hershberger RE, Day SM, Klinedinst NJ, Landstrom AP, Parikh VN, et al. Genetic testing for inherited cardiovascular diseases: a scientific statement from the American Heart Association. Circ: Genomic and Precision Medicine 2020 Aug;13(4):1. [doi: 10.1161/hcg.0000000000000067]
55. Wiegman A, de Groot E, Hutten BA, Rodenburg J, Gort J, Bakker HD, et al. Arterial intima-media thickness in children heterozygous for familial hypercholesterolaemia. The Lancet 2004 Jan;363(9406):369-370. [doi: 10.1016/s0140-6736(04)15467-6]
56. Kusters DM, Wiegman A, Kastelein JJ, Hutten BA. Carotid intima-media thickness in children with familial hypercholesterolemia. Circulation Research 2014 Jan 17;114(2):307-310. [doi: 10.1161/circresaha.114.301430]
57. Vlahos AP, Naka KK, Bechlioulis A, Theoharis P, Vakalis K, Moutzouri E, et al. Endothelial dysfunction, but not structural atherosclerosis, is evident early in children with heterozygous familial hypercholesterolemia. Pediatr Cardiol 2014 Jan 3;35(1):63-70. [doi: 10.1007/s00246-013-0742-0] [Medline: 23821294]
58. de Jongh S, Lilien MR, Bakker HD, Hutten BA, Kastelein JJ, Stroes ES. Family history of cardiovascular events and endothelial dysfunction in children with familial hypercholesterolemia. Atherosclerosis 2002 Jul;163(1):193-197. [doi: 10.1016/s0021-9150(02)00003-5] [Medline: 12048139]
59. Lozano P, Henrikson NB, Dunn J, Morrison CC, Nguyen M, Blasi PR, et al. Lipid screening in childhood and adolescence for detection of familial hypercholesterolemia: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2016 Aug 09;316(6):645-655. [doi: 10.1001/jama.2016.6176] [Medline: 27532919]
60. Schrodi SJ, Mukherjee S, Shan Y, Tromp G, Sninsky JJ, Callear AP, et al. Genetic-based prediction of disease traits: prediction is very difficult, especially about the future. Front Genet 2014 Jun 02;5:162 [FREE Full text] [doi: 10.3389/fgene.2014.00162] [Medline: 24917882]
61. Ajufo EC, Aragam KG. Improving polygenic risk scores for coronary artery disease: what helps and by how much? JACC Basic Transl Sci 2023 Dec;8(12):1500-1502 [FREE Full text] [doi: 10.1016/j.jacbts.2023.10.005] [Medline: 38205353]

62. Sharif F, Daly K, Crowley J, O'Brien T. Current status of catheter- and stent-based gene therapy. Cardiovasc Res 2004 Nov 01;64(2):208-216. [doi: 10.1016/j.cardiores.2004.07.003] [Medline: 15485679]
63. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr 2020 Jul;14(4):337-339 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.012] [Medline: 32305024]
64. van der Schaar M, Alaa AM, Floto A, Gimson A, Scholtes S, Wood A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Mach Learn 2021 Dec 09;110(1):1-14 [FREE Full text] [doi: 10.1007/s10994-020-05928-x] [Medline: 33318723]
65. Krittanawong C, Johnson KW, Choi E, Kaplin S, Venner E, Murugan M, et al. Artificial intelligence and cardiovascular genetics. Life (Basel) 2022 Feb 14;12(2):279 [FREE Full text] [doi: 10.3390/life12020279] [Medline: 35207566]
66. Zreik M, van Hamersvelt RW, Wolterink JM, Leiner T, Viergever MA, Isgum I. A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. IEEE Trans. Med. Imaging 2019 Jul;38(7):1588-1598. [doi: 10.1109/tmi.2018.2883807]
67. Shelke S, Veerubhotla K, Lee Y, Lee CH. Telehealth of cardiac devices for CVD treatment. Biotechnol Bioeng 2024 Mar 27;121(3):823-834. [doi: 10.1002/bit.28637] [Medline: 38151894]
68. James S, Rao SV, Granger CB. Registry-based randomized clinical trials--a new clinical trial paradigm. Nat Rev Cardiol 2015 May 17;12(5):312-316. [doi: 10.1038/nrcardio.2015.33] [Medline: 25781411]
69. Lauschensky A, Hayn D, Eggerth A, Modre-Osprian R, Pfeifer B, Egelseer-Bründl T, et al. Concept for visualisation of guideline adherence of medication prescriptions in a heart failure telehealth system. Stud Health Technol Inform 2020 Jun 23;271:49-56. [doi: 10.3233/SHTI200073] [Medline: 32578540]
70. Bradbury A, Patrick-Miller L, Harris D, Stevens E, Egleston B, Smith K, et al. Utilizing remote real-time videoconferencing to expand access to cancer genetic services in community practices: a multicenter feasibility study. J Med Internet Res 2016 Feb 01;18(2):e23 [FREE Full text] [doi: 10.2196/jmir.4564] [Medline: 26831751]
71. Boothe E, Greenberg S, Delaney CL, Cohen SA. Genetic counseling service delivery models: a study of genetic counselors' interests, needs, and barriers to implementation. J Genet Couns 2021 Feb 03;30(1):283-292. [doi: 10.1002/jgc4.1319] [Medline: 32885542]
72. Peshkin BN, Kelly S, Nusbaum RH, Similuk M, DeMarco TA, Hooker GW, et al. Patient perceptions of telephone vs. in-person BRCA1/BRCA2 genetic counseling. J Genet Couns 2016 Jun 12;25(3):472-482 [FREE Full text] [doi: 10.1007/s10897-015-9897-6] [Medline: 26455498]
73. Athens BA, Caldwell SL, Umstead KL, Connors PD, Brenna E, Biesecker BB. A systematic review of randomized controlled trials to assess outcomes of genetic counseling. J Genet Couns 2017 Oct 02;26(5):902-933 [FREE Full text] [doi: 10.1007/s10897-017-0082-y] [Medline: 28255928]
74. Gorrie A, Gold J, Cameron C, Krause M, Kincaid H. Benefits and limitations of telegenetics: a literature review. J Genet Couns 2021 Aug 04;30(4):924-937. [doi: 10.1002/jgc4.1418] [Medline: 33817891]
75. Hilgart JS, Hayward JA, Coles B, Iredale R. Telegenetics: a systematic review of telemedicine in genetics services. Genet Med 2012 Sep;14(9):765-776 [FREE Full text] [doi: 10.1038/gim.2012.40] [Medline: 22498847]
76. Cohen AJ, Shur N, Starin D, MacLeod E, Roshan Lal T, Leon E, et al. Pediatric medical genetics house call: telemedicine for the next generation of patients and providers. Am J Med Genet C Semin Med Genet 2021 Mar 11;187(1):55-63. [doi: 10.1002/ajmg.c.31882] [Medline: 33427371]
77. Mills R, MacFarlane IM, Caleshu C, Ringler MA, Zierhut HA. Genetic counselor experiences with telehealth before and after COVID-19. J Genet Couns 2021 Aug 07;30(4):999-1009 [FREE Full text] [doi: 10.1002/jgc4.1465] [Medline: 34231953]
78. Vandenberk B, Raj SR. Remote patient monitoring: what have we learned and where are we going? Curr Cardiovasc Risk Rep 2023 Apr 22;17(6):103-115 [FREE Full text] [doi: 10.1007/s12170-023-00720-7] [Medline: 37305214]
79. Tayal M, Mukherjee A, Chauhan U, Uniyal M, Garg S, Singh A, et al. Evaluation of remote monitoring device for monitoring vital parameters against reference standard: a diagnostic validation study for COVID-19 preparedness. Indian J Community Med 2020;45(2):235. [doi: 10.4103/ijcm.ijcm_317_20]
80. Ding X, Clifton D, Ji N, Lovell NH, Bonato P, Chen W, et al. Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic. IEEE Rev. Biomed. Eng 2021;14:48-70. [doi: 10.1109/rbme.2020.2992838]
81. Watson AR, Wah R, Thamman R. The value of remote monitoring for the COVID-19 pandemic. Telemed J E Health 2020 Sep 01;26(9):1110-1112. [doi: 10.1089/tmj.2020.0134] [Medline: 32384251]
82. Mishra N, Duke J, Lenert L, Karki S. Public health reporting and outbreak response: synergies with evolving clinical standards for interoperability. J Am Med Inform Assoc 2020 Jul 01;27(7):1136-1138 [FREE Full text] [doi: 10.1093/jamia/ocaa059] [Medline: 32692844]
83. Henkel AG, Spinner C. IT-STRATEGIE: Digitalumbau bei laufendem Betrieb. kma 2020 Apr 15;25(04):51-54. [doi: 10.1055/s-0040-1709870]
84. Munos B, Baker P, Bot B, Crouthamel M, de Vries G, Ferguson I, et al. Mobile health: the power of wearables, sensors, and apps to transform clinical trials. Ann N Y Acad Sci 2016 Jul;1375(1):3-18. [doi: 10.1111/nyas.13117] [Medline: 27384501]

85. Song X, Wang X, Zhang J, Shen S, Yin W, Ye G, et al. A tunable self-healing ionic hydrogel with microscopic homogeneous conductivity as a cardiac patch for myocardial infarction repair. Biomaterials 2021 Jun;273:120811. [doi: 10.1016/j.biomaterials.2021.120811] [Medline: 33882404]
86. Pecha S, Eschenhagen T, Reichenspurner H. Myocardial tissue engineering for cardiac repair. J Heart Lung Transplant 2016 Mar;35(3):294-298. [doi: 10.1016/j.healun.2015.12.007] [Medline: 26856673]
87. Lee M, Kim MC, Lee JY. Nanomaterial-based electrically conductive hydrogels for cardiac tissue repair. IJN 2022 Dec;Volume 17:6181-6200. [doi: 10.2147/ijn.s386763]
88. Lin X, Liu Y, Bai A, Cai H, Bai Y, Jiang W, et al. A viscoelastic adhesive epicardial patch for treating myocardial infarction. Nat Biomed Eng 2019 Aug 15;3(8):632-643. [doi: 10.1038/s41551-019-0380-9] [Medline: 30988471]
89. Wu T, Liu W. Functional hydrogels for the treatment of myocardial infarction. NPG Asia Mater 2022 Feb 18;14(1):1. [doi: 10.1038/s41427-021-00330-y]
90. Toyoda Y, Guy TS, Kashem A. Present status and future perspectives of heart transplantation. Circ J 2013;77(5):1097-1110 [FREE Full text] [doi: 10.1253/circj.cj-13-0296] [Medline: 23614963]
91. Zhang Y, Zhu D, Wei Y, Wu Y, Cui W, Liuqin L, et al. A collagen hydrogel loaded with HDAC7-derived peptide promotes the regeneration of infarcted myocardium with functional improvement in a rodent model. Acta Biomater 2019 Mar 01;86:223-234. [doi: 10.1016/j.actbio.2019.01.022] [Medline: 30660010]
92. Wu T, Cui C, Huang Y, Liu Y, Fan C, Han X, et al. Coadministration of an adhesive conductive hydrogel patch and an injectable hydrogel to treat myocardial infarction. ACS Appl Mater Interfaces 2020 Jan 15;12(2):2039-2048. [doi: 10.1021/acsami.9b17907] [Medline: 31859471]
93. Nguyen-Truong M, Li Y, Wang Z. Mechanical considerations of electrospun scaffolds for myocardial tissue and regenerative engineering. Bioengineering (Basel) 2020 Oct 03;7(4):122 [FREE Full text] [doi: 10.3390/bioengineering7040122] [Medline: 33022929]
94. Li Z, Yi N, Chen R, Meng Y, Wang Y, Liu H, et al. miR-29b-3p protects cardiomyocytes against endotoxin-induced apoptosis and inflammatory response through targeting FOXO3A. Cell Signal 2020 Oct;74:109716 [FREE Full text] [doi: 10.1016/j.cellsig.2020.109716] [Medline: 32707074]
95. Hodgkinson CP, Bareja A, Gomez JA, Dzau VJ. Emerging concepts in paracrine mechanisms in regenerative cardiovascular medicine and biology. Circulation Research 2016 Jan 08;118(1):95-107. [doi: 10.1161/circresaha.115.305373]
96. Williams DF. On the mechanisms of biocompatibility. Biomaterials 2008 Jul;29(20):2941-2953. [doi: 10.1016/j.biomaterials.2008.04.023] [Medline: 18440630]
97. Chan BP, Leong KW. Scaffolding in tissue engineering: general approaches and tissue-specific considerations. Eur Spine J 2008 Dec 13;17 Suppl 4(Suppl 4):467-479 [FREE Full text] [doi: 10.1007/s00586-008-0745-3] [Medline: 19005702]
98. Shapira A, Feiner R, Dvir T. Composite biomaterial scaffolds for cardiac tissue engineering. International Materials Reviews 2024 Jul 29;61(1):1-19. [doi: 10.1179/1743280415y.0000000012]
99. Cocarta A, Hobzova R, Trchova M, Svojgr K, Kodetova M, Pochop P, et al. 2 - hydroxyethyl methacrylate hydrogels for local drug delivery: study of topotecan and vincristine sorption/desorption kinetics and polymer-drug interaction by ATR - FTIR spectroscopy. Macro Chemistry & Physics 2021 May 04;222(13):1. [doi: 10.1002/macp.202100086]
100. Ghamkhari A, Abbaspour-Ravasjani S, Talebi M, Hamishehkar H, Hamblin MR. Development of a graphene oxide-poly lactide nanocomposite as a Smart Drug Delivery System. Int J Biol Macromol 2021 Feb 01;169:521-531. [doi: 10.1016/j.ijbiomac.2020.12.084] [Medline: 33340628]
101. Kazemi Ashtiani M, Zandi M, Shokrollahi P, Ehsani M, Baharvand H. Surface modification of poly(2 - hydroxyethyl methacrylate) hydrogel for contact lens application. Polymers for Advanced Techs 2018 Jan 12;29(4):1227-1233. [doi: 10.1002/pat.4233]
102. Rossos A, Banti C, Kalampounias A, Papachristodoulou C, Kordatos K, Zoumpoulakis P, et al. pHEMA@AGMNA-1: a novel material for the development of antibacterial contact lens. Mater Sci Eng C Mater Biol Appl 2020 Jun;111:110770. [doi: 10.1016/j.msec.2020.110770] [Medline: 32279741]
103. Macková H, Plichta Z, Hlídková H, Sedláček O, Konefal R, Sadakbayeva Z, et al. Reductively degradable poly(2-hydroxyethyl methacrylate) hydrogels with oriented porosity for tissue engineering applications. ACS Appl Mater Interfaces 2017 Mar 29;9(12):10544-10553. [doi: 10.1021/acsami.7b01513] [Medline: 28287694]
104. Passos MF, Carvalho NMS, Rodrigues AA, Bavaresco VP, Jardini AL, Maciel MRW, et al. PHEMA hydrogels obtained by infrared radiation for cartilage tissue engineering. International Journal of Chemical Engineering 2019 Jan 31;2019:1-9. [doi: 10.1155/2019/4249581]
105. Kalinke C, Muñoz RAA. 3D-printed microdevices: from design to applications. Micromachines (Basel) 2024 Jun 15;15(6):791 [FREE Full text] [doi: 10.3390/mi15060791] [Medline: 38930761]
106. Li M, Wu H, Yuan Y, Hu B, Gu N. Recent fabrications and applications of cardiac patch in myocardial infarction treatment. VIEW 2021 Sep 16;3(2):1. [doi: 10.1002/viw.20200153]
107. Mei X, Cheng K. Recent development in therapeutic cardiac patches. Front Cardiovasc Med 2020 Nov 27;7:610364 [FREE Full text] [doi: 10.3389/fcvm.2020.610364] [Medline: 33330673]

108. Rinaldi A, Caraffi R, Grazioli MV, Oddone N, Giardino L, Tosi G, et al. Applications of the ROS-responsive thioketal linker for the production of smart nanomedicines. Polymers (Basel) 2022 Feb 11;14(4):687 [FREE Full text] [doi: 10.3390/polym14040687] [Medline: 35215600]
109. Wu S, Yang C, Hsu W, Lin L. 3D-printed microelectronics for integrated circuitry and passive wireless sensors. Microsyst Nanoeng 2015 Jul 20;1(1):1. [doi: 10.1038/micronano.2015.13]
110. Joe Lopes A, MacDonald E, Wicker R. Integrating stereolithography and direct print technologies for 3D structural electronics fabrication. Rapid Prototyping Journal 2012;18(2):43 [FREE Full text] [doi: 10.1108/13552541211212113]
111. Guo S, Qiu K, Meng F, Park SH, McAlpine MC. 3D printed stretchable tactile sensors. Adv Mater 2017 Jul 05;29(27):1 [FREE Full text] [doi: 10.1002/adma.201701218] [Medline: 28474793]
112. Roda A, Guardigli M, Calabria D, Calabretta MM, Cevenini L, Michelini E. A 3D-printed device for a smartphone-based chemiluminescence biosensor for lactate in oral fluid and sweat. Analyst 2014 Dec 21;139(24):6494-6501. [doi: 10.1039/c4an01612b] [Medline: 25343380]
113. Sun P. How AI Helps Physicians Improve Telehealth Patient Care in Real-Time. Arizona Telemedicine Program. 2022 Jun 23. URL: https://telemedicine.arizona.edu/blog/how-ai-helps-physicians-improve-telehealth-patient-care-real-time [accessed 2025-02-21]
114. Lafont A, Guzman LA, Whitlow PL, Goormastic M, Cornhill JF, Chisolm GM. Restenosis after experimental angioplasty. Intimal, medial, and adventitial changes associated with constrictive remodeling. Circ Res 1995 Jun;76(6):996-1002. [doi: 10.1161/01.res.76.6.996] [Medline: 7758171]
115. Thiele H, Oettel S, Jacobs S, Hambrecht R, Sick P, Gummert JF, et al. Comparison of bare-metal stenting with minimally invasive bypass surgery for stenosis of the left anterior descending coronary artery. Circulation 2005 Nov 29;112(22):3445-3450. [doi: 10.1161/circulationaha.105.578492]
116. Komatsu R, Ueda M, Naruko T, Kojima A, Becker AE. Neointimal tissue response at sites of coronary stenting in humans: macroscopic, histological, and immunohistochemical analyses. Circulation 1998 Jul 21;98(3):224-233. [doi: 10.1161/01.cir.98.3.224] [Medline: 9697822]
117. Liu MW, Hearn JA, Luo JF, Anderson PG, Roubin GS, Iyer S, et al. Reduction of thrombus formation without inhibiting coagulation factors does not inhibit intimal hyperplasia after balloon injury in pig coronary arteries. Coron Artery Dis 1996 Sep;7(9):667-671. [doi: 10.1097/00019501-199609000-00008] [Medline: 8950497]
118. Smith-Norowitz TA, Shani J, Weiser W, Schulhoff N, Norowitz K, Lichstein E, et al. Lymphocyte activation in angina pectoris. Clin Immunol 1999 Nov;93(2):168-175. [doi: 10.1006/clim.1999.4776] [Medline: 10527693]
119. Picard FJ, Bergeron MG. Rapid molecular theranostics in infectious diseases. Drug Discov Today 2002 Nov 01;7(21):1092-1101. [doi: 10.1016/s1359-6446(02)02497-2] [Medline: 12546841]
120. Libby P. Atherosclerosis: the new view. Sci Am 2002 May;286(5):46-55. [doi: 10.1038/scientificamerican0502-46] [Medline: 11951331]
121. Lekshmi KM, Che H, Cho C, Park I. Drug- and gene-eluting stents for preventing coronary restenosis. Chonnam Med J 2017 Jan;53(1):14-27 [FREE Full text] [doi: 10.4068/cmj.2017.53.1.14] [Medline: 28184335]
122. Klugherz BD, Jones PL, Cui X, Chen W, Meneveau NF, DeFelice S, et al. Gene delivery from a DNA controlled-release stent in porcine coronary arteries. Nat Biotechnol 2000 Nov;18(11):1181-1184. [doi: 10.1038/81176] [Medline: 11062438]
123. Yang J, Zeng Y, Zhang C, Chen Y, Yang Z, Li Y, et al. The prevention of restenosis in vivo with a VEGF gene and paclitaxel co-eluting stent. Biomaterials 2013 Feb;34(6):1635-1643 [FREE Full text] [doi: 10.1016/j.biomaterials.2012.11.006] [Medline: 23199742]
124. Paul A, Shao W, Shum-Tim D, Prakash S. The attenuation of restenosis following arterial gene transfer using carbon nanotube coated stent incorporating TAT/DNA(Ang1+Vegf) nanoparticles. Biomaterials 2012 Oct;33(30):7655-7664. [doi: 10.1016/j.biomaterials.2012.06.096] [Medline: 22818986]
125. Ji R, Cheng Y, Yue J, Yang J, Liu X, Chen H, et al. MicroRNA expression signature and antisense-mediated depletion reveal an essential role of MicroRNA in vascular neointimal lesion formation. Circulation Research 2007 Jun 08;100(11):1579-1588. [doi: 10.1161/circresaha.106.141986]
126. Che H, Bae I, Lim KS, Song IT, Lee H, Muthiah M, et al. Suppression of post-angioplasty restenosis with an Akt1 siRNA-embedded coronary stent in a rabbit model. Biomaterials 2012 Nov;33(33):8548-8556. [doi: 10.1016/j.biomaterials.2012.07.045] [Medline: 22940215]
127. Fishbein I, Alferiev I, Bakay M, Stachelek SJ, Sobolewski P, Lai M, et al. Local delivery of gene vectors from bare-metal stents by use of a biodegradable synthetic complex inhibits in-stent restenosis in rat carotid arteries. Circulation 2008 Apr 22;117(16):2096-2103. [doi: 10.1161/circulationaha.107.746412]
128. Kim TG, Lee Y, Park TG. Controlled gene-eluting metal stent fabricated by bio-inspired surface modification with hyaluronic acid and deposition of DNA/PEI polyplexes. Int J Pharm 2010 Jan 15;384(1-2):181-188. [doi: 10.1016/j.ijpharm.2009.09.042] [Medline: 19799974]
129. Hannan MA, Mutashar S, Samad SA, Hussain A. Energy harvesting for the implantable biomedical devices: issues and challenges. BioMed Eng OnLine 2014 Jun 20;13(1):1. [doi: 10.1186/1475-925x-13-79]
130. Takahata K, Gianchandani Y. A planar approach for manufacturing cardiac stents: design, fabrication, and mechanical evaluation. J. Microelectromech. Syst 2004 Dec;13(6):933-939. [doi: 10.1109/jmems.2004.838357]

131. Want R. An introduction to RFID technology. IEEE Pervasive Comput 2006 Jan;5(1):25-33. [doi: 10.1109/mprv.2006.2]
132. Chow EY, Beier BL, Francino A, Chappell WJ, Irazoqui PP. Toward an implantable wireless cardiac monitoring platform integrated with an FDA-approved cardiovascular stent. J Interv Cardiol 2009 Oct 05;22(5):479-487 [FREE Full text] [doi: 10.1111/j.1540-8183.2009.00483.x] [Medline: 19807844]
133. Kanda Y. Piezoresistance effect of silicon. Sensors and Actuators A: Physical 1991 Jul;28(2):83-91. [doi: 10.1016/0924-4247(91)85017-I]
134. Zeng S, Baillargeat D, Ho H, Yong K. Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. Chem Soc Rev 2014 May 21;43(10):3426-3452. [doi: 10.1039/c3cs60479a] [Medline: 24549396]
135. Akar O, Akin T, Najafi K. A wireless batch sealed absolute capacitive pressure sensor. Sensors and Actuators A: Physical 2001 Dec;95(1):29-38. [doi: 10.1016/s0924-4247(01)00753-1]
136. Chow EY, Chlebowski AL, Chakraborty S, Chappell WJ, Irazoqui PP. Fully wireless implantable cardiovascular pressure monitor integrated with a medical stent. IEEE Trans. Biomed. Eng 2010 Jun;57(6):1487-1496. [doi: 10.1109/tbme.2010.2041058]
137. Pan CT, Yang H, Shen SC, Chou MC, Chou HP. A low-temperature wafer bonding technique using patternable materials. Journal of Micromechanics and Microengineering 2002 Jun 21;12(5):611-615 [FREE Full text] [doi: 10.1088/0960-1317/12/5/315]
138. Svasek P, Svasek E, Lendl B, Vellekoop M. Fabrication of miniaturized fluidic devices using SU-8 based lithography and low temperature wafer bonding. Sensors and Actuators A: Physical 2004 Sep;115(2-3):591-599. [doi: 10.1016/j.sna.2004.03.055]

Abbreviations

AI: artificial intelligence
CAD: coronary artery disease
CTA: computed tomography angiography
CVD: cardiovascular disease
H2O2: hydrogen peroxide
MI: myocardial infarction
pHEMA: poly(hydroxyethyl) methacrylate
PLGA: polylactic-polyglycolic acid
PRS: polygenic risk score
ROS: reactive oxygen species

Edited by A Teles; submitted 13.08.24; peer-reviewed by Y Xu, W Cui, D Sun; comments to author 13.12.24; revised version received 22.12.24; accepted 08.01.25; published 08.03.25.

Please cite as:
Lee Y, Shelke S, Lee C
Cardiac Repair and Regeneration via Advanced Technology: Narrative Literature Review
JMIR Biomed Eng 2025;10:e65366
URL: https://biomedeng.jmir.org/2025/1/e65366
doi:10.2196/65366
PMID:

Optimizing Voice Sample Quantity and Recording Settings for the Prediction of Type 2 Diabetes Mellitus: Retrospective Study

Atousa Assadi, MSc, MEng; Jessica Oreskovic, MAS; Jaycee Kaufman, MSc; Yan Fossat, MSc

Klick Applied Sciences, 175 Bloor St East, North Tower, 3rd floor, Toronto, ON, Canada

Corresponding Author:
Yan Fossat, MSc
Klick Applied Sciences, 175 Bloor St East, North Tower, 3rd floor, Toronto, ON, Canada

Abstract

Background: The use of acoustic biomarkers derived from speech signals is a promising non-invasive technique for diagnosing type 2 diabetes mellitus (T2DM). Despite its potential, there remains a critical gap in knowledge regarding the optimal number of voice recordings and recording schedule necessary to achieve effective diagnostic accuracy.

Objective: This study aimed to determine the optimal number of voice samples and the ideal recording schedule (frequency and timing), required to maintain the T2DM diagnostic efficacy while reducing patient burden.

Methods: We analyzed voice recordings from 78 adults (22 women), including 39 individuals diagnosed with T2DM. Participants had a mean (SD) age of 45.26 (10.63) years and mean (SD) BMI of 28.07 (4.59) kg/m². In total, 5035 voice recordings were collected, with a mean (SD) of 4.91 (1.45) recordings per day; higher adherence was observed among women (5.13 [1.38] vs 4.82 [1.46] in men). We evaluated the diagnostic accuracy of a previously developed voice-based model under different recording conditions. Segmented linear regression analysis was used to assess model accuracy across varying numbers of voice recordings, and the Kendall tau correlation was used to measure the relationship between recording settings and accuracy. A significance threshold of $P<.05$ was applied.

Results: Our results showed that including up to 6 voice recordings notably improved the model accuracy for T2DM compared to using only one recording, with accuracy increasing from 59.61 to 65.02 for men and from 65.55 to 69.43 for women. Additionally, the day on which voice recordings were collected did not significantly affect model accuracy ($P>.05$). However, adhering to recording within a single day demonstrated higher accuracy, with accuracy of 73.95% for women and 85.48% for men when all recordings were from the first and second days.

Conclusions: This study underscores the optimal voice recording settings to reduce patient burden while maintaining diagnostic efficacy.

(JMIR Biomed Eng 2025;10:e64357) doi:10.2196/64357

KEYWORDS

vocal biomarker; acoustic biomarker; voice analysis; type 2 diabetes; diagnostics; digital phenotyping; voice data

Introduction

Diabetes mellitus is a chronic metabolic disorder characterized by persistent elevated blood glucose levels due to inadequate or impaired insulin production or utilization. It affects 10.5% of the worldwide population (536.6 million people) [1], with type 2 diabetes mellitus (T2DM) accounting for 90% of cases [2]. Uncontrolled diabetes is a major contributors to cardiovascular disorders, blindness, renal failure, and lower limb amputation [2].

Traditional diagnostic methods of fasting plasma glucose and oral glucose tolerance tests involve blood sampling, which can cause inconvenience or discomfort to patients owing to frequent monitoring. Moreover, the lack of a glucometer and the time spent for self-testing are barriers in the self-management of diabetes [3,4]. In response to these challenges, acoustic biomarkers from speech signals have emerged as promising non-invasive alternatives, offering a convenient solution for diagnosing and monitoring diabetes, especially for individuals in remote areas with restricted health care accessibility.

Sustained periods of high blood glucose and the detrimental effects of peripheral neuropathy and myopathy in individuals with T2DM impact the elastic properties of the vocal folds [5], weaken the laryngeal muscles, and induce respiratory changes [6]. These physiological changes can affect voice parameters, leading to voice disorders like hoarseness [7] and dysphagia [8]. Consequently, compared to those without T2DM, individuals with the condition exhibit significant vocal differences, quantified by phonation time, fundamental frequency, jitters, and shimmers [6], highlighting the importance of investigating vocal variations as potential markers of T2DM [9-13].

Our group previously assessed the feasibility of using voice recordings from mobile applications to detect T2DM [14]. Our results demonstrated that voice biomarkers—specifically pitch, jitters, and shimmers—combined with age and BMI could predict T2DM with an accuracy of 0.89 for women and 0.86 for men [14]. However, requiring participants to record their voices at least 6 times daily over a 2-week period posed challenges related to participant burden and recording consistency.

Therefore, this study aims to optimize the voice sampling process by determining (1) the minimum number of voice samples required, and (2) the optimal recording schedule (frequency and timing) needed to maintain diagnostic accuracy while reducing participant burden. We hypothesize that a more streamlined voice sampling protocol can achieve comparable predictive performance to prior studies while improving feasibility for long-term diabetes monitoring.

Methods

Study Design

To address the objectives of this project, we designed a retrospective study based on our previously developed model and the same dataset that yielded the highest balanced accuracy [14]. The original data collection took place between August 30, 2021, and June 30, 2022 in India [14]. In total, 505 participants (mean [SD] age 41.03 [13.29] years, 336 male participants) were recruited and instructed to record a short English phrase ("Hello. How are you? What is my glucose level right now?") up to 6 times daily using their smartphone for 14 consecutive days.

Participants and Measurements

A balanced subset of the original dataset was used for this analysis and included 78 participants (aged >18 years old, 22 women), with 39 diagnosed with T2DM [14]. Participants in the T2DM and non-T2DM groups were matched for age and BMI to minimize the demographic impact on voice recordings (Table 1). A T2DM diagnosis was confirmed by a physician according to the American Diabetes Association guidelines [15]. All participants were nonsmokers, had no diagnosed neurological or speech impairments, and signed the consent forms.

Table . Patient demographic characteristics.

Variable	Total	Non-T2DM[a] group	T2DM group
N (%)	78 (100.0)	39 (50.0)	39 (50.0)
Women	22 (28.21)	11 (50.0)	11 (50.0)
Men	56 (71.79)	28 (50.0)	28 (50.0)
Age (years), mean (SD)	45.26 (10.63)	45.49 (10.8)	45.03 (10.58)
Women	45.82 (10.4)	45.91 (10.85)	45.73 (10.47)
Men	45.04 (10.8)	45.32 (10.98)	44.75 (10.8)
BMI (kg/m^2), mean (SD)	28.07 (4.59)	28.77 (5.01)	27.36 (4.06)
Women	30.25 (5.35)	31.41 (5.4)	29.09 (5.29)
Men	27.21 (3.98)	27.74 (4.53)	26.68 (3.34)

[a]T2DM: type 2 diabetes mellitus.

As part of the study protocol, participants recorded their voice at least 6 times per day over a 2-week period using a custom mobile application installed on their personal cell phones. These recordings took place either at home or in a quiet environment with minimal background noise [12]. To establish a consistent starting point, a participant's first day (d01) was defined as the day they recorded at least 2 voice samples. Voice samples recorded prior to d01 were excluded from the analysis.

Optimizing Voice Recording Quantity and Settings for Enhanced Model Accuracy

To analyze the collected voice recordings, 14 acoustic features were extracted to characterize key parameters related to pitch, intensity, harmonic noise ratio, shimmers, and jitters [14]. Features that were significantly different between the T2DM and non-T2DM groups (P<.05, Cohen d >0.02) were included in he model development pipeline, with separate models for women and men. For women, the key features were pitch SD, mean pitch, RAP jitter, and apq3 shimmer, while for men, mean intensity, apq11 shimmer, intensity SD, and ppq5 jitter were used. A 5-fold cross-validation was performed for feature selection, threshold determination, and model optimization based on the best predictive balanced accuracy [14]. The optimal model for women was a logistic regression model (threshold of 0.3) with BMI and 3 voice features: mean pitch, pitch SD, and RAP jitter. For men, the optimal model was a naive Bayes classifier (threshold of 0.215) with age, BMI, and 2 voice features: mean intensity and apq11 shimmer.

The analysis pipeline included (1) indicating the optimal number of voice recordings for effective T2DM classification based on changes in model accuracy across varying quantities of voice samples, and (2) studying the effect of voice recording configurations on predictive performance (Multimedia Appendix 1).

To study the changes in the model's accuracy trend across varying number of voice samples, we employed segmented linear regression by fitting two distinct linear models to the data before and after the n voice samples breakpoint. The Kendall tau measure of correlation was used to investigate the strength and direction of the relationship between ordinal variables (such as days) and model accuracy. *P* values of .05 were considered statistically significant.

Ethical Considerations

The protocol (ID MGCTS107) received ethics approval by Saanvi Ethical Research LLP (No. MGCTS/20/107 V01), all participants provided informed consent, and data were stored in a secure cloud database with no identifying information. Participants were compensated for their time; each participant received US $56.

Results

Participants and Measurements

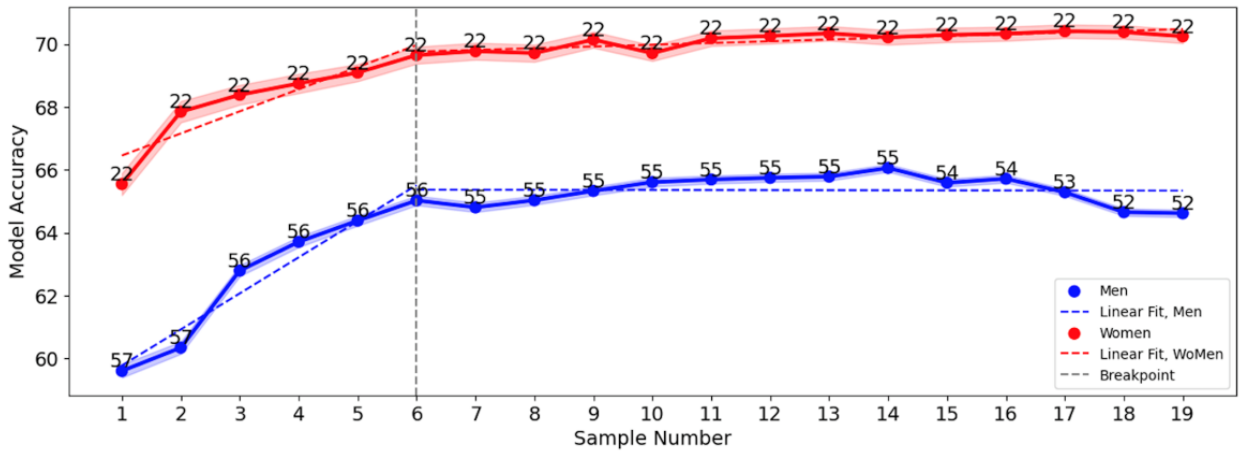
The mean (SD) age and mean (SD) BMI of participants were 45.26 (10.63) years and 28.07 (4.59) kg/m^2, respectively (Table 1). In total, 5035 recordings were included in the analysis, and 2620 from participants with T2DM (Table 2). The mean (SD) number of daily recordings for all participants was 4.91 (1.45) with women more adherent to the protocol than men (5.13 [1.38] vs 4.82 [1.46], Multimedia Appendix 2, Figure 1).

Table . Voice recording data.

Variable	Total	Non-T2DM[a] group	T2DM group
N (%)	5035 (100)	2415 (48)	2620 (52)
Women, n (%)	1539 (30.6)	713 (46.3)	826 (53.7)
Men, n (%)	3496 (69.4)	1702 (48.7)	1794 (51.3)
Number of recordings per participant, mean (SD)	64.55 (19.53)	61.92 (21.67)	67.18 (17)
Women	69.95 (16.26)	64.82 (19.12)	75.09 (11.45)
Men	62.43 (20.42)	60.79 (22.82)	64.07 (17.97)

[a]T2DM: type 2 diabetes mellitus.

Figure 1. The accuracy of the model using different number of voice recordings. The lines present the average accuracy for men (blue) and women (red). The shaded area shows the confidence interval. The numbers in the figure show the number of participants whose data were included in the analysis per day.

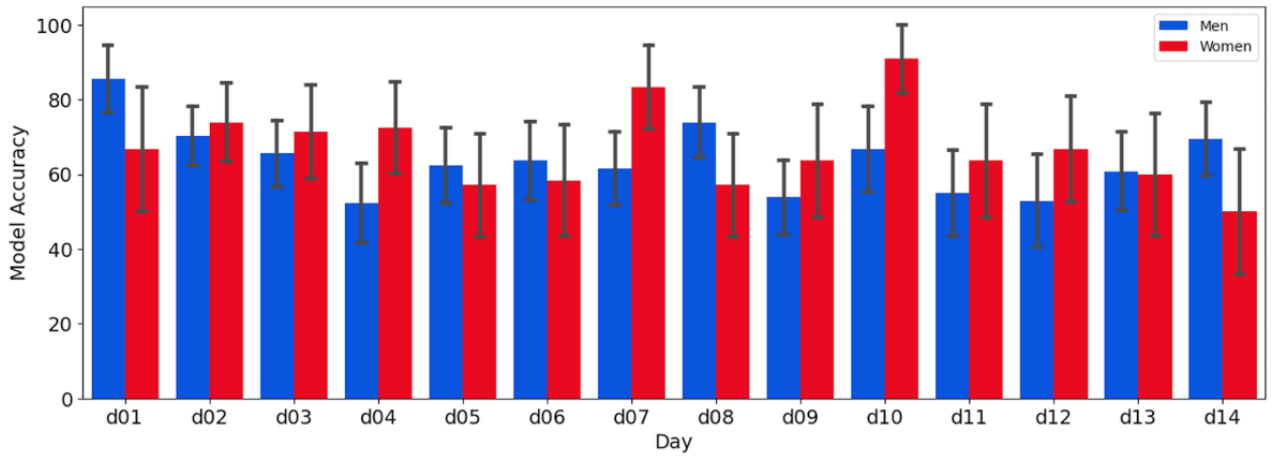


Optimizing Voice Recording Quantity and Settings for Enhanced Model Accuracy

Both in men and women, the model accuracy improved with the inclusion of up to 6 voice samples, after which it stabilized with no significant improvement (Figure 1). The changes in the slope of the linear fit were −1.15 for men and −0.65 for women, indicating a faster accuracy improvement in men than in women with the addition of initial voice samples.

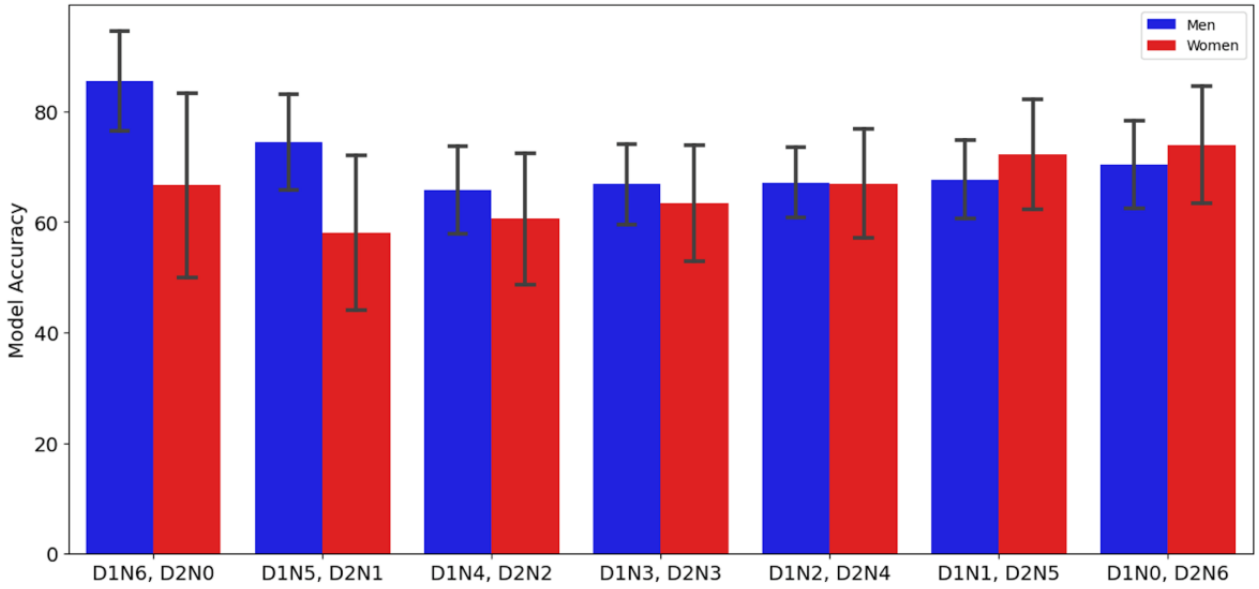
Considering 6 voice samples for effective T2DM diagnosis, the highest model accuracy was achieved with recordings from day 1 in men, while for women, the peak accuracy was observed with recordings from day 10 (Figure 2). However, the variations in model accuracy across different days were not significant, and no statistically discernible trend was observed (*P*=.23 for men, *P*=.27 for women). The model accuracy was generally higher for women than for men on most days, although the difference was not statistically significant, as indicated by the overlapping confidence intervals.

Figure 2. The accuracy of the model using 6 voice recordings per day. d: day.



Moreover, we observed that the model accuracy was higher in men than in women when the majority of recordings were from the first day (Figure 3). As the distribution of recordings were balanced between the first and second days, the accuracy gap between men and women narrowed. Finally, when the majority of recordings were from the second day, the model accuracy was slightly higher for women than for men, with the differences in accuracy levels becoming less pronounced. Our statistical test indicated no significant trend in the model accuracy for men when using 6 recordings across 2 days (P>.99). For women, there was a significant increasing trend in the model accuracy when the majority of recordings were taken on the second day (P=.03), suggesting that consistent participation in women can improve the model performance.

Figure 3. The accuracy of the model using total 6 voice recordings from day 1 and day 2. D1: day 1, D2: day 2, N: number of samples from the day.



Discussion

Principal Findings

This research, to our knowledge, is the first to investigate the optimal balance between the number and settings of voice recordings for effective T2DM diagnosis, with the goal of reducing patient burden. Our findings indicated that 6 voice recordings are sufficient to maintain diagnostic accuracy, improving patient compliance and accessibility for T2DM screening. No significant differences in model accuracy were observed across different days while adherence to recording in a single day showed higher accuracy. This study lays the groundwork for future research and clinical applications focused on optimizing health care delivery for T2DM.

Comparison to Prior Works

Previous studies have shown that an increased burden from the treatment and self-management of chronic health conditions such as T2DM is associated with higher levels of distress, lower adherence to self-care routines, decreased satisfaction with medications, reduced quality of life, poorer physical and mental health, and greater risk of complications and deaths [4,16-18]. In alignment with these, our study demonstrated that up to 6 voice recordings are sufficient to effectively diagnose T2DM,

thereby reducing patient burden while maintaining diagnostic accuracy.

There are conflicting reports on self-management among men and women. While Zhou et al [19] observed that women exhibited greater compliance in self-care than men did, Mogre et al [20] reported higher self-monitoring of blood glucose in men. In our study, despite the lower number of women participants than men, women showed a higher adherence to the voice recording protocol. This higher adherence among women may explain the observed increase in model accuracy, as recordings were distributed across 2 days, suggesting that consistent participation enhances the model performance.

Prior research has reported no significant day-to-day variability in voice recordings while there exists a significant time-of-day influence on acoustics with voice quality enhanced with increased voice use [21,22]. In alignment with these findings, our results showed that both in men and in women, the model accuracy was not significantly different between days while there was an increase in accuracy when the majority of the recordings were from a single day. Due to the limited distribution of samples across different times of the day, we were unable to assess the time-of-day variability.

Strengths and Limitations

This study provides important insights into optimizing voice-based T2DM diagnostics while minimizing participant burden. However, several limitations should be considered. First, there was a limited sample size of women. The smaller number of women participants may reduce the generalizability of our findings, particularly regarding sex-specific effects. Future studies with larger, more balanced datasets are needed to validate these observations. Second, our relatively small dataset limited the use of more advanced machine learning techniques, such as neural networks. While these models may offer further improvements in the diagnostic accuracy and insight into optimal data collection strategies, future studies with larger datasets are required to fully explore their potential. Third, due to uneven distribution of recordings across different times of the day, we could not assess how the time-of-day influences voice-based diagnostics. Future studies should implement controlled recording schedules to systematically examine these effects.

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

All authors contributed to preparing this manuscript. AA selected the analysis pipeline, performed data analysis, and prepared the manuscript. JO contributed to data analysis and revised and edited the manuscript. JK developed the voice classification model and provided feedback in methodology, data analysis, and reviewing the manuscript. YF supervised all aspects of the project, provided funding resources, and was involved in reviewing and submitting the manuscript.

Conflicts of Interest

JO, JK, and YF are employees of Klick Inc.

Multimedia Appendix 1
Methods for optimizing voice recording quantity and settings for enhanced model accuracy.
[DOCX File, 15 KB - biomedeng_v10i1e64357_app1.docx]

Multimedia Appendix 2
Number of daily voice recordings per participant.
[DOCX File, 135 KB - biomedeng_v10i1e64357_app2.docx]

References

1. Sun H, Saeedi P, Karuranga S, et al. IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
2. Goyal R, Singhal M, Jialal I. Type 2 Diabetes: StatPearls; 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK513253/ [accessed 2025-06-25]
3. Mogre V, Johnson NA, Tzelepis F, Paul C. Barriers to diabetic self-care: a qualitative study of patients' and healthcare providers' perspectives. J Clin Nurs 2019 Jun;28(11-12):2296-2308. [doi: 10.1111/jocn.14835] [Medline: 30791160]
4. Gonzalez JS, Tanenbaum ML, Commissariat PV. Psychosocial factors in medication adherence and diabetes self-management: Implications for research and practice. Am Psychol 2016 Oct;71(7):539-551. [doi: 10.1037/a0040388] [Medline: 27690483]

5. Sidorova J, Carbonell P, Čukić M. Blood glucose estimation from voice: first review of successes and challenges. J Voice 2022 Sep;36(5):737. [doi: 10.1016/j.jvoice.2020.08.034] [Medline: 33041176]
6. Saghiri MA, Vakhnovetsky A, Vakhnovetsky J. Scoping review of the relationship between diabetes and voice quality. Diabetes Res Clin Pract 2022 Mar;185:109782. [doi: 10.1016/j.diabres.2022.109782] [Medline: 35176400]
7. Hamdan AL, Kurban Z, Azar ST. Prevalence of phonatory symptoms in patients with type 2 diabetes mellitus. Acta Diabetol 2013 Oct;50(5):731-736. [doi: 10.1007/s00592-012-0392-3] [Medline: 22527095]
8. Weinreb SF, Piersiala K, Hillel AT, Akst LM, Best SR. Dysphonia and dysphagia as early manifestations of autoimmune inflammatory myopathy. Am J Otolaryngol 2021;42(1):102747. [doi: 10.1016/j.amjoto.2020.102747] [Medline: 33038783]
9. Pinyopodjanard S, Suppakitjanusant P, Lomprew P, Kasemkosin N, Chailurkit L, Ongphiphadhanakul B. Instrumental acoustic voice characteristics in adults with type 2 diabetes. J Voice 2021 Jan;35(1):116-121. [doi: 10.1016/j.jvoice.2019.07.003] [Medline: 31427120]
10. Gölaç H, Atalik G, Türkcan AK, Yilmaz M. Disease related changes in vocal parameters of patients with type 2 diabetes mellitus. Logoped Phoniatr Vocol 2022 Oct;47(3):202-208. [doi: 10.1080/14015439.2021.1917653] [Medline: 33970753]
11. Chitkara D, Sharma RK. Voice based detection of type 2 diabetes mellitus. Presented at: 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB); Feb 27-28, 2016; Chennai, India p. 83-87. [doi: 10.1109/AEEICB.2016.7538402]
12. Kaufman J, Jeon J, Oreskovic J, Fossat Y. Linear effects of glucose levels on voice fundamental frequency in type 2 diabetes and individuals with normoglycemia. Sci Rep 2024 Aug 28;14(1):19012. [doi: 10.1038/s41598-024-69620-z] [Medline: 39198592]
13. Oreskovic J, Kaufman J, Fossat Y. Impact of audio data compression on feature extraction for vocal biomarker detection: validation study. JMIR Biomed Eng 2024 Apr 15;9:e56246. [doi: 10.2196/56246] [Medline: 38875677]
14. 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 Dec;1(4):534-544. [doi: 10.1016/j.mcpdig.2023.08.005] [Medline: 40206319]
15. American Diabetes Association Professional Practice Committee. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2022. Diabetes Care 2022 Jan 1;45(Suppl 1):S17-S38. [doi: 10.2337/dc22-S002] [Medline: 34964875]
16. Eton DT, Yost KJ, Lai JS, et al. Development and validation of the patient experience with treatment and self-management (PETS): a patient-reported measure of treatment burden. Qual Life Res 2017 Feb;26(2):489-503. [doi: 10.1007/s11136-016-1397-0] [Medline: 27566732]
17. Jannoo Z, Mamode Khan N. Medication adherence and diabetes self-care activities among patients with type 2 diabetes mellitus. Value Health Reg Issues 2019 May;18:30-35. [doi: 10.1016/j.vhri.2018.06.003] [Medline: 30419448]
18. Chen Y, Sloan FA, Yashkin AP. Adherence to diabetes guidelines for screening, physical activity and medication and onset of complications and death. J Diabetes Complications 2015;29(8):1228-1233. [doi: 10.1016/j.jdiacomp.2015.07.005] [Medline: 26316423]
19. Zhou C, Chen J, Tan F, et al. Relationship between self-care compliance, trust, and satisfaction among hypertensive patients in China. Front Public Health 2022;10:1085047. [doi: 10.3389/fpubh.2022.1085047] [Medline: 36743158]
20. Mogre V, Abanga ZO, Tzelepis F, Johnson NA, Paul C. Adherence to and factors associated with self-care behaviours in type 2 diabetes patients in Ghana. BMC Endocr Disord 2017 Mar 24;17(1):20. [doi: 10.1186/s12902-017-0169-3] [Medline: 28340613]
21. Heald SLM, Nusbaum HC. Variability in vowel production within and between days. PLoS ONE 2015;10(9):e0136791. [doi: 10.1371/journal.pone.0136791] [Medline: 26331478]
22. Pierce JL, Tanner K, Merrill RM, Shnowske L, Roy N. Acoustic variability in the healthy female voice within and across days: how much and why? J Speech Lang Hear Res 2021 Aug 9;64(8):3015-3031. [doi: 10.1044/2021_JSLHR-21-00018] [Medline: 34269598]

Abbreviations

T2DM: type 2 diabetes mellitus

Edited by S Rizvi, T Leung; submitted 15.07.24; peer-reviewed by V Martin, Z Wang; revised version received 03.06.25; accepted 04.06.25; published 26.06.25.

Please cite as:
Assadi A, Oreskovic J, Kaufman J, Fossat Y
Optimizing Voice Sample Quantity and Recording Settings for the Prediction of Type 2 Diabetes Mellitus: Retrospective Study
JMIR Biomed Eng 2025;10:e64357
URL: https://biomedeng.jmir.org/2025/1/e64357
doi:10.2196/64357

Using Vibration for Secure Pairing With Implantable Medical Devices: Development and Usability Study

Mo Zhang[1,2], PhD; Chaofan Wang[3], PhD; Weiwei Jiang[4], PhD; David Oswald[1], PhD; Toby Murray[2], PhD; Eduard Marin[5], PhD; Jing Wei[2], PhD; Mark Ryan[1], PhD; Vassilis Kostakos[2], PhD

[1]School of Computer Science, University of Birmingham, Birmingham, United Kingdom

[2]School of Computing and Information Systems, The University of Melbourne, Melbourne Connnect, 700 Swanston Street, Carlton, Melbourne, Australia

[3]College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, China

[4]School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, China

[5]Telefonica Research Spain, Barcelona, Spain

Corresponding Author:
Mo Zhang, PhD
School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Abstract

Background: Implantable medical devices (IMDs), such as pacemakers, increasingly communicate wirelessly with external devices. To secure this wireless communication channel, a pairing process is needed to bootstrap a secret key between the devices. Previous work has proposed pairing approaches that often adopt a “seamless” design and render the pairing process imperceptible to patients. This lack of user perception can significantly compromise security and pose threats to patients.

Objective: The study aimed to explore the use of highly perceptible vibrations for pairing with IMDs and aim to propose a novel technique that leverages the natural randomness in human motor behavior as a shared source of entropy for pairing, potentially deployable to current IMD products.

Methods: A proof of concept was developed to demonstrate the proposed technique. A wearable prototype was built to simulate an individual acting as an IMD patient (real patients were not involved to avoid potential risks), and signal processing algorithms were devised to use accelerometer readings for facilitating secure pairing with an IMD. The technique was thoroughly evaluated in terms of accuracy, security, and usability through a lab study involving 24 participants.

Results: Our proposed pairing technique achieves high pairing accuracy, with a zero false acceptance rate (indicating low risks from adversaries) and a false rejection rate of only 0.6% (1/192; suggesting that legitimate users will likely experience very few failures). Our approach also offers robust security, which passes the National Institute of Standards and Technology statistical tests (with all P values >.01). Moreover, our technique has high usability, evidenced by an average System Usability Scale questionnaire score of 73.6 (surpassing the standard benchmark of 68 for “good usability”) and insights gathered from the interviews. Furthermore, the entire pairing process can be efficiently completed within 5 seconds.

Conclusions: Vibration can be used to realize secure, usable, and deployable pairing in the context of IMDs. Our method also exhibits advantages over previous approaches, for example, lenient requirements on the sensing capabilities of IMDs and the synchronization between the IMD and the external device.

(JMIR Biomed Eng 2025;10:e57091) doi:10.2196/57091

KEYWORDS

implantable medical device; pairing; vibration; security; usability

Introduction

Background

Implantable medical devices (IMDs), such as pacemakers, implantable cardioverter defibrillators, or insulin pumps are widely deployed and evolving at a rapid pace [1]. Modern IMDs typically rely on a wireless interface to communicate with external devices. For instance, doctors use programmer devices to reprogram the patient’s IMD (eg, to change the patient’s therapy) and gather telemetry data. Such wireless connectivity can bring about much convenience to patients and doctors. However, it also poses new security and privacy threats, such as eavesdropping on sensitive medical data or hijacking life-critical functions. The consequences of such attacks can be severe because they can cause serious injuries or even death. However, these risks have often been overlooked. While no real-world attack against an IMD has been confirmed to date, previous research has demonstrated that many IMDs available

on the market today severely lack effective security mechanisms, and that attacks on patients would be practically possible [2-6].

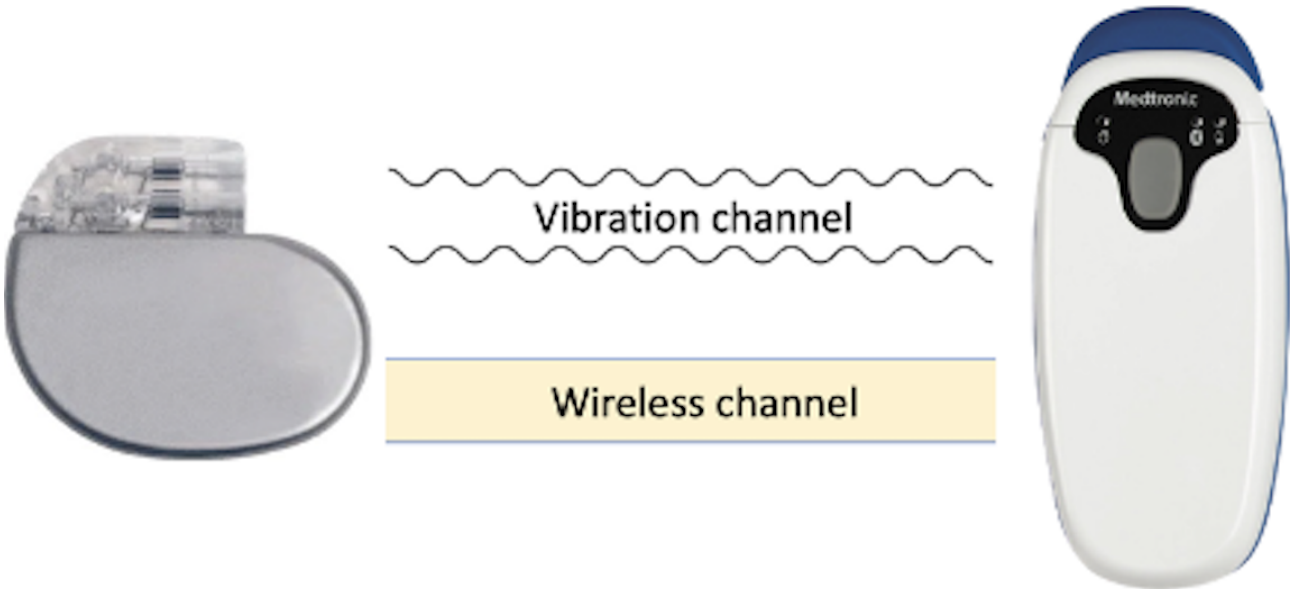
To protect wireless communication links, it is essential for the IMD and external device to undergo a pairing process. This process aims to exchange a cryptographic key between them, which can then be used to secure the wireless channel using standard protocols [7]. However, implementing such a key exchange in a secure manner is challenging because IMDs are resource-constrained with limited memory, computational power, and nonrechargeable and nonreplaceable batteries. Moreover, IMDs do not have physically accessible input or output interfaces, such as a keyboard or a screen once they are implanted. This obstructs traditional pairing methods used in technologies like Bluetooth, where manually typing a 4-digit PIN code on the devices is a standard procedure [8]. Furthermore, network connections with these devices can be ad-hoc. For instance, in an emergency (eg, patients with cardiac implants can experience syncope symptoms and become unconscious [9]), a doctor may quickly have to use a new programmer device to connect to the patient's IMD. Due to these limitations of IMDs, conventional pairing techniques (such as the ones based on symmetric or public keys [10]) are often not a viable option [5,11].

Previous work has proposed a variety of pairing techniques to overcome this challenge [12]. Rasmussen et al [13] propose an approach where the IMD and external device send ultrasound to each other to verify each other's legitimacy and exchange a key. Marin et al [5] and Tomlinson et al [14] propose a pairing method by transmitting a low alternating current through the patient's skin and tissue. Denning et al [15] and Gollakota et al [16] propose to delegate security to a proxy device that the patient can carry around (such as a bracelet). [17-20] propose a pairing process by the IMD and external device synchronously and simultaneously measuring a human physiological signal (such as heartbeats).

Across those previous approaches, a crucial aspect has been systematically overlooked: user perception. We observe that previous work has attempted to follow a "seamless" design approach that makes the IMD pairing as unobtrusive as possible to the patient, rendering the pairing process almost imperceptible at the same time. This can prevent patients from detecting unexpected pairing attempts made by adversaries in proximity, thereby hindering their ability to appropriately respond to such security threats, for example, by seeking assistance or fleeing the scene. Although the "seamless" design principle is common in everyday security systems [21], we question its suitability in the IMD context, where the device is part of the patient, and its security is life-critical.

To address this issue, a pairing protocol needs to incorporate a perceivable and robust (ie, cannot be hidden or canceled by an adversary) signal. This leads us to consider vibration as an out-of-band (OOB) channel (ie, a communication channel other than a wireless channel) for pairing (Figure 1). Vibrations are highly perceivable and have been widely used in smart consumer devices for notification services [22]. In addition, accelerometers, the primary type of vibration receiver used in previous approaches, are already present in state-of-the-art IMDs for medical purposes [23-25]. Another advantage of using vibration is its limited range of reliable reception. In the IMD context, this implies that if an external device intends to transmit a vibration to an IMD, it must be physically attached to the patient's skin for a while [26]. If an adversary overpowers the signal with a very strong vibration from a distance, the patient can easily notice this.

Figure 1. IMD and external device. The vibration channel is used to exchange a key that subsequently secures the wireless channel. IMD: implantable medical device.



Related Work

Vibration-Based Secret Transmission in Ad-Hoc Networks

Previous work has proposed vibration as an OOB channel for transmitting secrets between 2 devices that physically contact each other [26-32]. Table 1 summarizes their application scenarios and hardware setups. Most are designed for wearables and Internet of Things (IoT) devices that are not implanted in the human body. As a common setup, the transmitter (such as a smartphone) is equipped with a vibration motor and the receiver contains a sensor to detect the vibrations, such as an accelerometer [26,27,29-31], gyroscope [32] or microphone [28].

Table . Setup for previous vibration-based secret transmission. The application context refers to the intended receiver device.

Technique	Application context	Receiver sensor type	Sampling rate (Hz)
Vibrate-to-Unlock [30]	RFID tag	Accelerometer	Not reported
SYNCVIBE [27]	Wearable	Accelerometer	1600
SecureVibe [26]	IMD[a]	Accelerometer	3200
VibroComm [32]	IoT[b] device	Gyroscope	32000
Ripple [29]	Mobile device	Accelerometer	1600
Ripple II [28]	Mobile device	Microphone	48000
Touch-And-Guard [31]	Wristband device	Accelerometer	250

[a]IMD: implantable medical device.

[b]IoT: Internet of Things.

Previous work predominantly directly embeds the secret within the vibration signal itself [26-30,32]: the transmitter encodes the secret into vibration using specific modulation methods (eg, on-off keying [26,27,33]), and the receiver picks up this vibration with a sensor and decodes the secret. Another strategy leverages vibration to "amplify" the secret from humans: Wei et al [31] propose an approach that pairs an IoT device with a wristband device. When the user (who wears the wristband) touches the IoT device, the IoT device emits a vibration that sweeps through a range of frequencies. Contrary to the above methods, the vibration here does not carry the secret and remains consistent across different sessions. Instead, the secret comes from the (to some extent random) resonant properties of the user's hand-arm area, which can be derived from the devices' accelerometer readings.

However, we argue that most work (in their current form) is not deployable in existing IMD products because they have stringent requirements on the receiver sensor. Microphones do not exist in IMDs, while inertial sensors (ie, accelerometer and gyroscope) often require sampling rates in several thousands of Hz or higher. Such high-performance sensors are rare in IMDs [34-36] and are too energy-consuming for IMDs' limited battery capacity [37]. Future studies could certainly explore if previous work remains effective at reduced sensor sampling rates such as a few hundred Hz. Nevertheless, this is likely to significantly impact the performance because vibration signal demodulation often requires sensor data with high resolution [29].

Overall, we find that only [31] demands a lower sampling rate of 250 Hz. This is because the secret relies on the resonant frequencies of the user's hand-arm region, which are situated in the low-frequency domain ranging from several to a few hundred Hz [38,39]. Nonetheless, its practicality was only validated for wristbands but has not been tested in other deployment environments or with different hardware setups.

Suitable Protocols for OOB Channel-Based Pairing

Previous work has extensively proposed using an OOB channel for pairing with resource-constrained devices, including IMDs [5,19,26,40,41]. Typically, the ultimate objective of such pairing is to establish a 128-bit cryptographic key between 2 devices for data encryption [7]. However, these works commonly propose to directly exchange the entire key through the OOB channel, which raises several concerns.

First, OOB channels often have much lower data throughput compared to conventional wireless channels. For instance, the data throughput of the aforementioned vibration-based method [31] is only 7.15 bits per second. As a result, a 128-bit key bootstrap would require at least 18 seconds, potentially posing issues of usability and safety in emergencies. Second, OOB channels face threats from advanced side-channel eavesdropping attacks. For example, a vibration channel might be compromised using microphones in proximity due to acoustic leakage, leading to severe consequences.

To mitigate these concerns, prior work has suggested using a password-authenticated key agreement (PAKE) method [19,42,43], such as Diffie-Hellman Encrypted Key Exchange [44]. PAKE is a cryptographic protocol aiming at exchanging a high-entropy cryptographic key between parties who have previously shared a short and low-entropy secret. This approach allows 2 devices to initially exchange a short bitstring, after which they execute a PAKE to further exchange a 128-bit key. The latter step can be fast and thus largely reduce the impact of the low data rate of OOB channels. In addition, PAKE provides forward secrecy and rules out offline brute-force attacks. This is the approach that we adopt in our work, and therefore we consider that vibration is only to be used to exchange an ephemeral and low-entropy secret between the IMD and the external device.

Objectives

The objective of this paper is to explore the potential of using vibration for pairing with resource-constrained IMDs. This study aimed to (1) propose a novel technique that leverages vibration to extract secrets from the naturally random human motor behavior for pairing, (2) develop a prototype as a proof-of-concept to demonstrate our technique, and (3) evaluate our prototype's accuracy, security, and usability in a lab study involving 24 participants.

Methods

Pairing Technique

The pairing process requires the user (patient or doctor) to repeatedly attach the external device to the patient's body (near the IMD's location) for a few times. In this work, each repetition was referred to as a cycle, and the complete pairing process (including several cycles) was defined as a run. Each cycle comprises three main steps:

1. Device attachment: the user attaches the external device to the body and holds it steadily.
2. Vibration broadcast*:* the external device emits a vibration signal for a short period. The signal is always the same and does not serve as the secret. Both the IMD and external device take a measurement of the acceleration. The user releases the external device when the vibration stops.
3. Randomness extraction: both devices process the sensed acceleration signal and derive a shared secret from it.

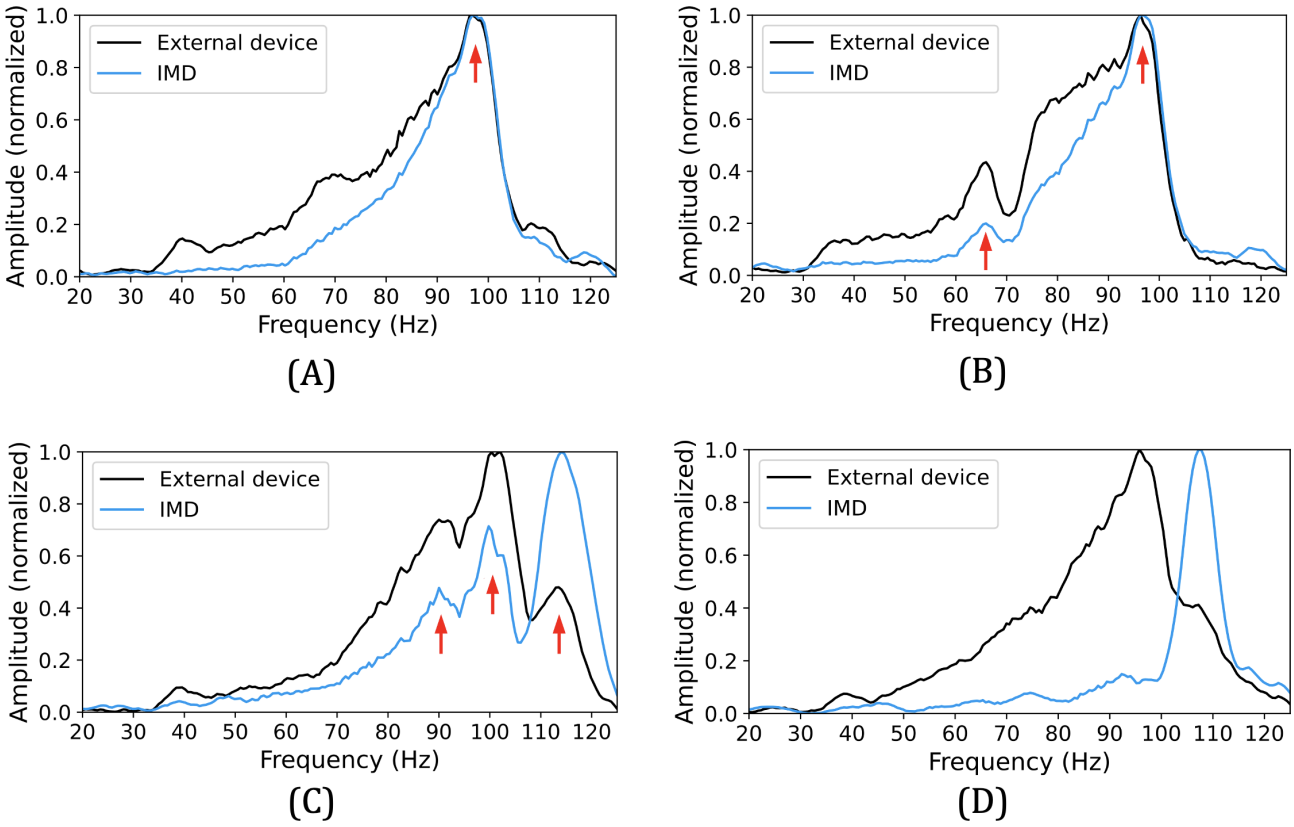
The security of pairing relies on the randomness of the shared secret, which originates from the diverse physiological characteristics of the human body as well as the inherent variability of human behavior (eg, the varying attachment position and the grip strength) [45]. The vibration signal itself remains constant in each cycle and is not a source of randomness. Instead, it serves as a "catalyst" that allows the randomness of body and motion to be reflected in the accelerometer measurements.

Obtaining a Shared Secret From Humans

The design of vibration strategy in each cycle—namely, the control of the motor to vibrate at a certain frequency for a certain time frame—is crucial. The feasibility of the aforementioned work [31] was first explored in the context of IMDs. The exact same experimental settings were replicated using our prototype that simulates the human body environment (elaborated in the following sections): the accelerometer sampling rates of the external device and IMD are set as 250 Hz. In each cycle, the motor is programmed to sweep between 20 Hz to 125 Hz within 1.75 seconds. During this period, 2 devices measure the z-axis acceleration data (aligning with the user's sagittal plane) and subsequently generate the frequency spectrum by doing fast Fourier transform (FFT) [46].

One researcher of the team performs 100 cycles as a preliminary test. The results are shown in Figure 2 (the locations of the resonant frequency peaks shared by both devices were regarded as secrets in [31]). Among all, 72 cycles show one stable peak; 17 cycles have 2 common peaks; 2 possess 3 peaks; in 9 cycles, the data is too noisy to capture any shared peaks. The results differ significantly from [31] where an average of 4 - 8 peaks can be obtained per cycle. In addition, the peaks in [31] are somehow uniformly distributed over the whole 20 - 125 Hz range, while ours are almost always in the range of 80 - 110 Hz. Our interpretation for the discrepancy in the performance of this strategy is the presence of the plastic board and shell in our prototype setup, which "masks" the resonant frequencies of the human body. Unfortunately, in the context of IMD pairing, the existence of such components (eg, a plastic or metal device housing) is inevitable.

Figure 2. Performance of the preliminary test (A) with 1 peak (72%), (B) with 2 peaks (17%), (C) with 3 peaks (2%), and (D) the noisy data (9%). IMD: implantable medical device.



Nevertheless, the above test implies the natural randomness inherent in the user attachment motions. Intuitively, we want to test if a constant-frequency vibration is a viable option. We program the motor to emit a 50 Hz vibration for 1 s per cycle, and the same researcher executes 100 cycles using our prototype. For each cycle, we collect z-axis acceleration data from both devices and generate the frequency spectrum using FFT. Figure 3 shows an example of the frequency spectrum in one cycle. It was observed that 2 devices can obtain very similar data, especially for a prominent amplitude peak. Figure 4 illustrates the spectrum change of the IMD over ten consecutive cycles. Each row in this figure corresponds to a frequency spectrum obtained in 1 cycle, and the bright spots indicate the prominent peaks on the curve. We observe that the peak locations vary around 50 Hz, suggesting the presence of a degree of randomness. These findings indicate that providing an excitement of a constant-frequency vibration, the prominent peak location in the frequency domain is a potentially qualified shared entropy source between the IMD and the external device, which can be used for pairing purposes.

Figure 3. Frequency spectrum given a constant vibration (50 Hz, 1 s) in one cycle. IMD: implantable medical device.

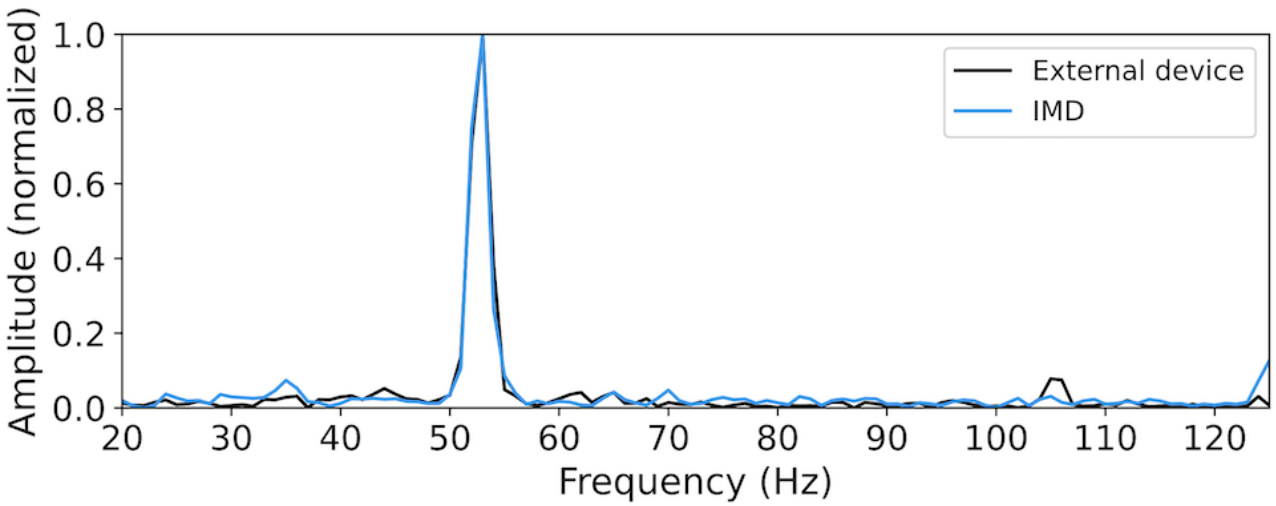
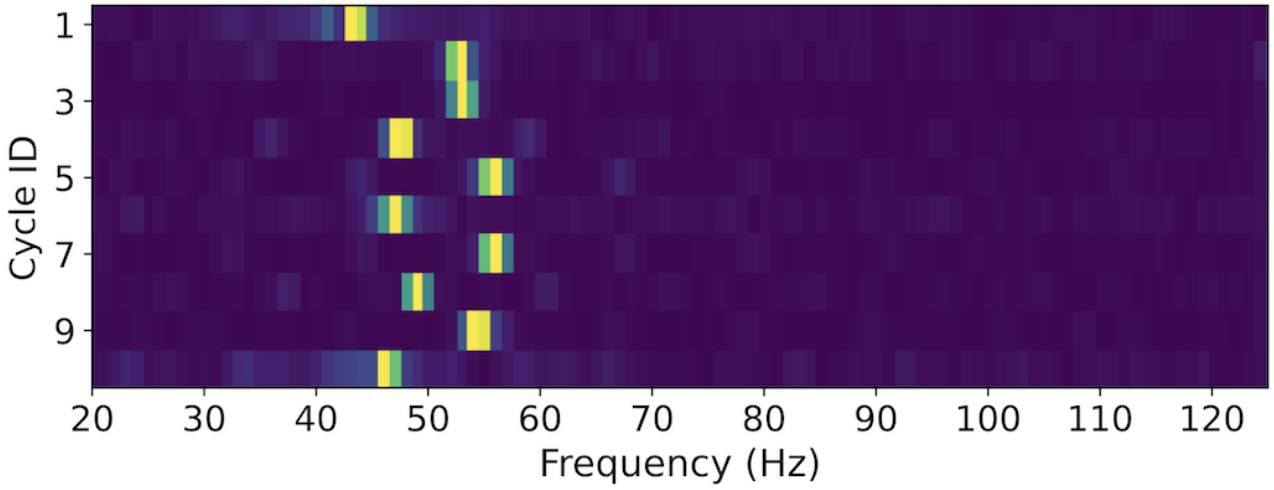


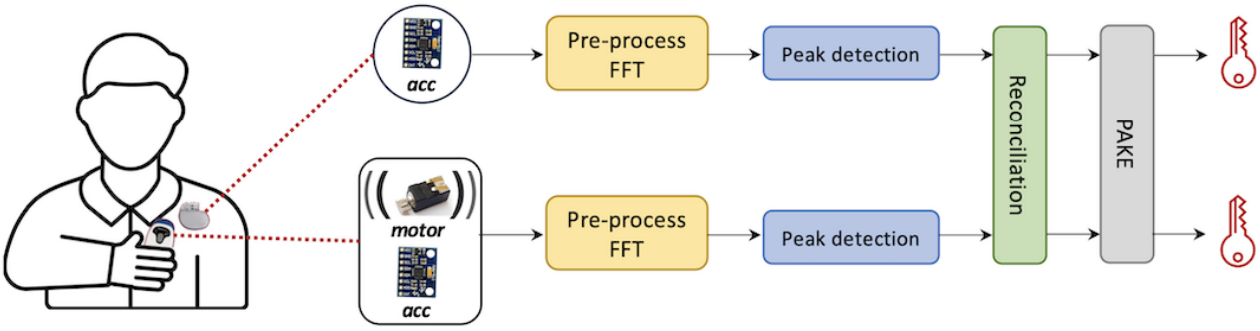
Figure 4. Frequency spectrum of IMD, given a constant vibration (50 Hz, 1 s) in 10 consecutive cycles. IMD: implantable medical device.



Signal Processing Workflow

Figure 5 shows the workflow of our pairing technique (assuming the IMD is a pacemaker). In each cycle, the patient holds the external device and attaches it on their chest. During the attachment, the motor vibrates, and both the IMD and the external device measure a pair of z-axis acceleration data.

Figure 5. An overview of our pairing technique. acc: accelerometer; FFT: fast Fourier transform; PAKE: password-authenticated key agreement.



To remove the noise of the direct current component, each device subtracts the acceleration data with its mean value. In addition, when the vibration motor is switched on from standstill or switched off, the generated vibration signal is not amplified or attenuated immediately but with a slow and damped response [26,27]. This means that the transition parts (ie, 2 ends) of a vibration signal segment are often noisy. This is addressed by applying a Hanning window on the data.

Subsequently, each device applies FFT on the acceleration signal to obtain the frequency spectrum. The frequency range of 0 to 20 Hz is then excluded to avoid the effects of noisy motion artifacts like human breathing movements, as well as ambient vibrations present in the patient's environment [47]. As mentioned, there is a prominent amplitude peak in the frequency spectrum. In order to detect the location of this peak, each device simply traverses the frequency domain to find the frequency value corresponding to the maximum amplitude.

Based on the above procedure, after the user completes a pairing (ie, a run) by repeating the attachment for several times, each of the 2 devices will possess a sequence of peak locations. However, these sequences may not be exactly the same due to the measurement noise and human error (eg, hand wobbles). To resolve this, the peak locations are encoded into binary format using Gray code [48]. This coding method ensures minimal bit mismatches if the discrepant peak locations are very close on 2 devices, which is the case of our technique. Then, we use a cryptographic algorithm known as a fuzzy extractor [11,49] to reconcile any remaining bit differences between the 2 bitstrings without revealing the secret itself. If the rate of bit mismatches falls within the error-correcting capability of the fuzzy extractor, the IMD and the external device agree on an identical bitstring as a shared secret.

Adversary Model

Given our review of relevant literature about IMD pairing techniques [3,12,13,16,19,40,50], we assume a sophisticated adversary following the Dolev-Yao model [51] who has full knowledge of our pairing protocol, has full control over the wireless communication channels, and can be a man-in-the-middle (MITM) attacker by intercepting legitimate devices' signals and sending their own messages instead. In particular, the adversary can launch the following attacks relevant in the context of our pairing technique: (1) impersonation attack: the adversary uses a sequence of peak locations in an attempt to impersonate a legitimate device. They could succeed if their sequence closely matches the one measured by the IMD or external device. (2) Brute-Force attack: the adversary brute-forces possible peak location sequences and launches MITM attacks to decipher and manipulate the

communication between legitimate devices. The brute force can be done online, that is, during the pairing process, the adversary tries every possible sequence until they hit a correct one. Alternatively, this can be done offline, where the adversary records the pairing traffic and performs offline analysis to crack the secret after pairing. (3) Acoustic eavesdropping: The adversary may also attempt to eavesdrop on the vibration signals using a microphone near the patient to reveal the secret.

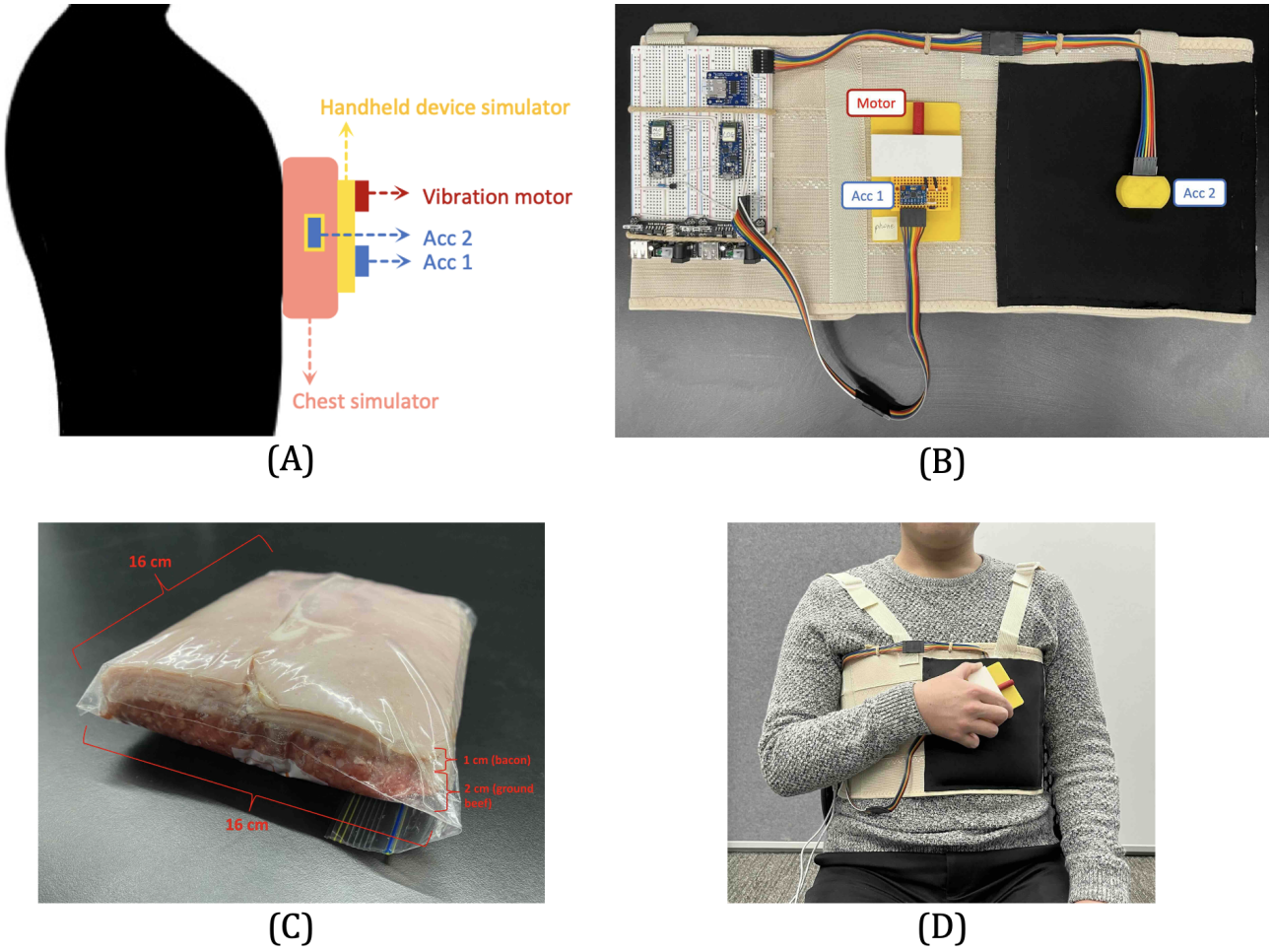
Experimental Setup

The proposed pairing technique was validated through the design and testing of the prototype in a user study. It was assumed that the IMD is a pacemaker implanted beneath the chest and considered the external handheld device to resemble a smartphone with a plastic casing. Moreover, both devices contain an accelerometer, and the external device is equipped with a vibration motor.

Prototype Implementation

We show an overview of our prototype in Figure 6A. The prototype consists of three main parts:

Figure 6. Experiment setup. (A) Prototype overview. (B) Hardware setup. (C) Chest simulator. (D) Participant in user study.



IMD

We use an InvenSense triaxial MPU-6050 accelerometer [52] to simulate a pacemaker and house it inside a 3D-printed case (Acc2 in Figure 6B). An Arduino Nano 33 BLE board interfaces with the sensor, which contains a 32-bit Cortex-M microcontroller and closely resembles the capabilities of an IMD [53]. The sampling rate of the accelerometer is set at 250 Hz, the same as in previous work [31].

External Device (the Vibration Transmitter)

We do not directly use a smartphone as the external device because the most common operating systems on mobile devices—Android and iOS systems do not provide an API interface for direct control of the vibration motor frequency. Instead, we use an eccentric rotating mass type vibration motor [54], along with another MPU-6050 sensor (Acc1 in Figure 6B) to simulate an external device. These components are mounted on an 11 cm × 7 cm × 0.5 cm plastic cuboid board, replicating the size and shape of a typical smartphone.

We use a separate Arduino Nano 33 BLE board to control both the vibration motor and the accelerometer. Particularly, this Arduino board connects to the vibration motor and supplies voltage to it. By using the pulse width modulation technique [31], the board can adjust the driving voltage, allowing the vibration motor's frequency to be altered accordingly. In addition, the accelerometer is set to a sampling rate of 250 Hz.

Chest Environment

Given that pacemakers are embedded inside the body, it is important for our experiments to mimic an environment that resembles the human chest. We adopt the design in previous research [5,26,40] and use 1 cm layer of bacon and 2 cm layer of lean ground beef to replicate the chest's physical properties (see Figure 6C). The 1 cm depth is a standard depth for pacemaker implantation [55]. In our study, we embed our pacemaker simulator within the meat layers, which are kept inside a food storage bag at room temperature. This bag of meat is subsequently placed in a pocket stitched onto an elastic chest band, positioned around an area corresponding to the human heart's location (see Figure 6B). Participants were asked to wear the chest strap throughout the user study to mimic the conditions of pacemaker users.

Participant Recruitment

We first conducted a pilot study with 6 individuals (ages 22 to 32 years, 4 females and 2 males) to identify and resolve any problems with our experimental setup. Subsequently, we recruited 24 participants for the main study, including 11 males and 13 females of ages ranging from 18 to 52.

Moreover, given that patients who carry IMDs are often seniors [56], we also conducted a co-design workshop with 2 senior individuals who had intimate knowledge and experience with pacemakers: (1) a 74-year-old female cardiology doctor and (2) a 79-year-old male pacemaker patient.

Experiment Procedure

In total, 2 essential vibration settings, frequency, and duration, were manipulated to measure the effect on pairing performance. Based on experiences gained from our pilot study, we set vibration motor frequencies to 50 Hz, 75 Hz, and 100 Hz, and vibration durations to 400 ms, 700 ms, and 1000 ms. The 9 frequency–duration combinations enabled successful pairing and avoided excessive participant workload.

During the user study, participants were instructed to wear our prototype and sit on a chair. Then they need to grasp the external device simulator and repeatedly attach it to the black pocket area of the chest strap, as shown in Figure 6D. They were advised to attach the device in a random manner (such as to random positions), and (in each cycle) stay attached until the vibration had completely ceased. Before starting the data collection, participants were asked to acquaint themselves with the prototype to understand the pairing process. This introductory process took under a minute for all participants. Subsequently, for each of the 9 vibration conditions, participants were asked to conduct the attachment for 5 consecutive cycles as one run and complete 4 such runs in total. The order in which participants used different vibration frequencies was counterbalanced.

At the end of the user study, participants were requested to fill out a standard system usability scale (SUS) questionnaire [57] to assess the usability of the pairing method. We then conducted an interview with them to gather further insights. Full details of the questionnaire and interview are given in Multimedia Appendix 1.

During the co-design workshop, we asked the two senior participants to try our prototype for only 6 runs (considering their physical conditions) and provide their opinions and advice.

Evaluation Metrics

Our study focuses on certain metrics to evaluate the pairing performance.

Accuracy

The accuracy of a pairing system is typically measured by false rejection rate (FRR) and false acceptance rate (FAR) [43,58,59]. FRR is the frequency at which the pairing between legitimate devices is incorrectly rejected. FAR indicates the frequency that a pair of illegitimate devices (such as the IMD and a malicious external device) is mistakenly authorized and gauges the resilience of pairing against impersonation attacks. A high FRR and FAR could lead to poor usability and security, respectively. These 2 metrics are calculated as follows:

$$\text{FRR} = \frac{\text{Number of rejected pairings between legitimate devices}}{\text{All pairing attempts between legitimate devices}}$$

$$\text{FAR} = \frac{\text{Number of successful pairings between illegitimate devices}}{\text{All pairing attempts between illegitimate devices}}$$

During the pairing process, there is often a mismatch (denoted by d) between the readings of the IMD and the external device due to inherent noises. As aforementioned, we use a fuzzy extractor scheme to correct the mismatch. At the core of this method is the selection of a threshold (denoted by Thr): the mismatch can be rectified (and thus the pairing is accepted) if $d \leq \text{Thr}$; otherwise, the pairing is rejected. As such, one can balance FRR and FAR by adjusting Thr. Because security is of utmost importance for the IMDs, we set a smaller Thr to ensure FAR=0 and use the corresponding lowest FRR to represent accuracy.

Security

The FAR metric evaluates the system's security against impersonation attacks. The resilience against brute-force attacks is determined by the randomness level of the attachment motions, which can be measured in two primary ways: (1) By the National Institute of Standards and Technology (NIST) statistical test suite [60] that provides a comprehensive randomness assessment of a random number generator, a method widely recognized within the cybersecurity community [19,31,43]. (2) By measuring Shannon entropy, which quantifies the amount of information contained in each motion event [17,31,61,62].

Usability

Usability is assessed based on the results from our SUS questionnaires and interviews.

Ethical Considerations

This study involved human participants and underwent thorough ethical review, particularly given the potential involvement of older participants. Ethics approval was obtained from the relevant institutions prior to participant recruitment and user study, in accordance with institutional regulations (the University of Melbourne: approved by the Human Ethics Committee, application number: 2022-24851-31088-3; the University of Birmingham: approved by the Science,

Technology, Engineering and Mathematics Ethics Committee, application number: ERN_2022 - 0255).

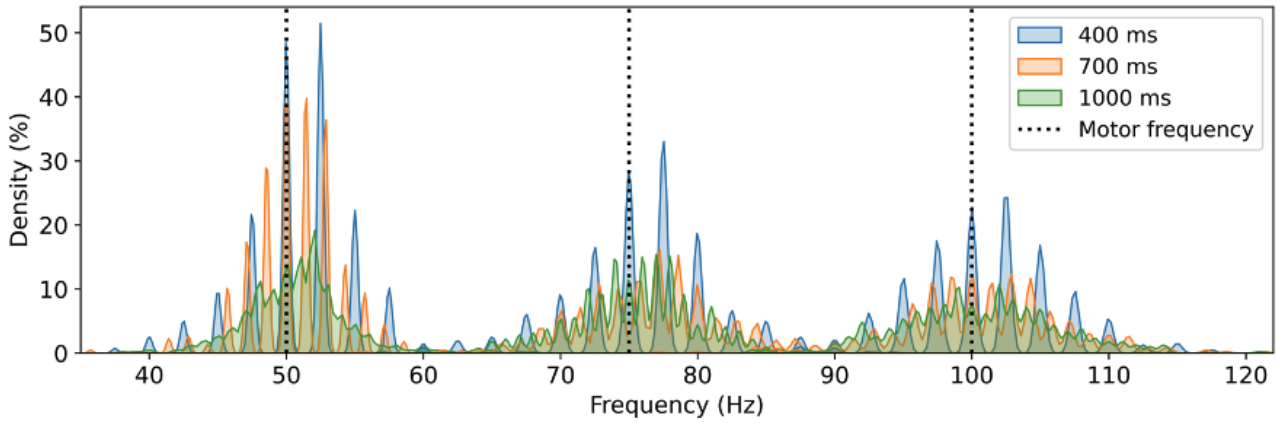
Participants were recruited via online advertisements and were offered US $30 for their time. All participants provided informed consent prior to participation. The data collected during the user study were specifically processed to ensure anonymity and untraceability of identity and were securely stored in the University of Melbourne's data center. All participant data were anonymized by removing personally identifiable information before analysis, and participants were assigned unique identification codes to ensure confidentiality. The entire user study process was overseen by a departmental delegate of the university's ethics committee, with all study details reported to them on a weekly basis.

Results

Performance of the Pairing Technique

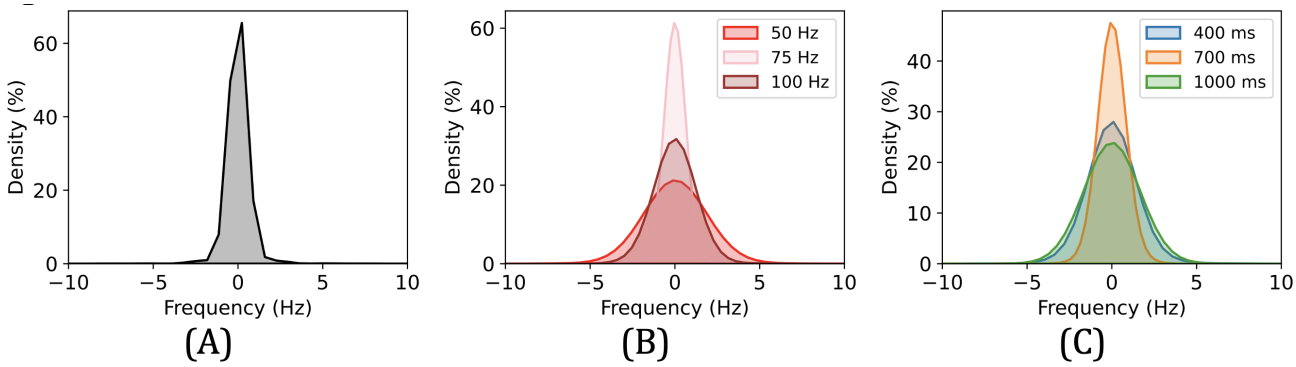
Figure 7 shows the distribution of all peak locations (ie, the secret) collected by the IMD from 24 participants. We observe that for a specific vibration frequency, such as 50 Hz, the peak locations range between 30 and 70 Hz and generally approximate a normal distribution centered by the motor's frequency, suggesting a certain degree of randomness from the user. Additionally, it appears that the distribution is slightly flatter (thereby the level of randomness increases) with an increase in vibration frequency. Notably, the possible options for peak locations in the frequency domain are not continuous due to the sample-based nature of the time domain acceleration data.

Figure 7. Distribution of peak location measured by the IMD among 24 participants. The black dashed lines indicate frequencies of the vibration motor. IMD: implantable medical device.



Mismatch is calculated by subtracting peak location values between the IMD and the external device and represents the level of noise and error. The mismatch distribution for our prototype, as illustrated in Figure 8A, resembles a normal distribution centered around a mean near zero and with a standard deviation of 2.8 Hz. This implies that user-induced errors and sensor noise are limited. Note that this result considers situations where participants did not strictly follow our pairing norms during the study. For instance, there were a number of occasions when participants released the external device while it was still vibrating. Such cases were not excluded from our dataset as they present a more realistic use scenario; otherwise, we expect that the mismatch levels would be even lower. On the other hand, Figure 8B and C show that the degree of mismatch does not have a straightforward correlation with either the vibration frequency or the duration.

Figure 8. Mismatch between the IMD and the external device. (A) Mismatch of all data among participants. (B) Mismatch with vibration frequency. (C) Mismatch with vibration time. IMD: implantable medical device.



Experimental Evaluation

Accuracy Assessment

For each of the 9 vibration conditions, we build two sets to measure accuracy:

Set I comprises 96 (=24 × 4) pairs of peak locations, each with a length of 5 (since we collected 5 cycles per run). All the pairs in Set I come from legitimate pairings of an IMD and an external device. This set calculates the FRR metric as aforementioned.

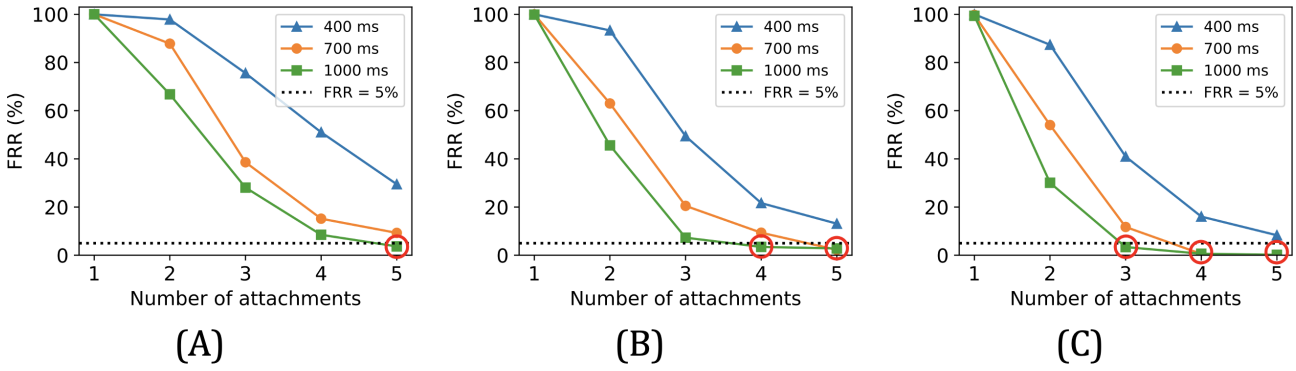
Set II has 96 pairs of peak locations (the same size as Set I), where each pair is created by randomly mixing data from illegitimate device pairings. This set calculates the FAR metric as aforementioned.

An effective pairing technique should maximize the acceptance of pairs from Set I (ie, low FRR), while minimizing the acceptance of pairs from Set II (ie, low FAR). Note that not all five motions are necessarily needed, ie, we can vary the length of runs ranging from 1 to 5, by truncating the initial elements.

The following figures show the accuracy of our prototype across various numbers of attachment motions performed. FAR is 0 in all cases, and we consider that an FRR below 5% signifies good usability [43,61]. As expected, increasing the number of motions consistently improves the accuracy of the pairing. Moreover, given a specific vibration frequency, longer vibration duration leads to higher accuracy, which will be further discussed in the coming sections. An additional observation is that with a fixed vibration duration and number of motions, the FRR tends to drop as the vibration frequency rises.

Overall, the red circles in the above figures indicate the 5 out of 9 vibration conditions that offer acceptable accuracy levels (with FAR=0 and FRR <5%). For example, a vibration condition of 50 Hz for 1000 ms per cycle requires the user to execute five attachments to achieve pairing with FAR=0 and FRR=3.7% (see Figure 9A). Note that for other vibration conditions, more than five motions are likely to also yield satisfactory accuracy. However, this would demand more effort from the user, which could harm usability and even safety in emergencies.

Figure 9. FRR versus number of attachments under different vibration frequencies: (A) 50 Hz, (B) 75 Hz, and (C) 100 Hz. FRR: false rejection rate.



Security Assessment

We refer to previous work [19,31,43] to assess the randomness of the secret generated by our technique: For each of the vibration conditions, we take the floor of the (fractional) entropy value for that specific setting (refer to Table 2) and extract that number of least significant bits from each peak location value. Subsequently, we combine these bitstrings from all vibration conditions (following the order in our user study) as a single 8.6 kbits string and evaluate its randomness using the NIST statistical test suite [60]. The full results are given in Table 3. The outputs of the NIST tests are *P* values that represent the probability the data is generated by an eligible random number generator. If a *P* value is smaller than a threshold (usually .01 [19,31,43]), the randomness hypothesis is rejected. Table 3 shows that all *P* values are larger than .01 and hence pass the NIST tests.

Table . Entropy of each attachment motion (unit is bit).

	50 Hz	75 Hz	100 Hz
400 ms	2.61	3.12	3.87
700 ms	3.03	3.86	4.15
1000 ms	3.15	4.01	4.48

Table . NIST statistical test results for attachment motions.

Test	*P* value	Test	*P* value
Frequency	.88	Block frequency	.14
Runs	.11	Longest runs	.10
Binary matrix rank	.17	FFT[a]	.62
Non-overlap template	.16	Overlapping template	.38
Serial (*P* value$_1$)	.64	Linear complexity	.72
Serial (*P* value$_2$)	.68	Approximate entropy	.18
Cumulative sums (forward)	.19	Random excursions	.22
Cumulative sums (reverse)	.26	Random excursions var.	.40

[a]FFT: fast Fourier transform.

Table 2 shows the entropy value contained in each motion across different vibration conditions. Overall, a single motion in our study carries an entropy from 2.61 to 4.48 bits. For a certain vibration frequency, the entropy grows with higher vibration durations. This is because extended measurements yield larger sample size and frequency resolution, enabling more possible peak locations and thus higher entropy. Furthermore, for a given vibration duration, the entropy value rises with an increase in vibration frequency. We leave the study of this phenomenon to future work. Nevertheless, the choice of vibration frequency is often limited by the capability of the motor and accelerometer in practice.

It is noteworthy that some entropy is sacrificed when rectifying mismatches between the 2 devices. Here, we make a preliminary estimation of the entropy loss: Using the encoding method in [43] on our dataset, the maximum bit mismatch rates (ie, percentage of different bits between two devices) for our prototype vary between 0.7% and 3.0% for different vibration conditions. This can be addressed by a fuzzy extractor with (31, 29) Reed-Solomon code that has a 3.23% error tolerance [11,63], potentially leading to an entropy loss of 6.5%.

Usability Assessment

The average SUS score for our pairing technique is 73.6 (SD 18.14), which generally passes the typical benchmark value of 68 for "good usability" [57]. It is important to note that the SUS questionnaires were completed after an extensive data collection process including a repetition of 180 attachment motions. We expect that users carrying out a more realistic task would report even higher usability scores.

We gained further insights into usability from the interviews. Over half of the participants (15 out of 24) explicitly indicated that our technique was easy to use. For example, one participant (p8) commented, "The attachment doesn't require me to think. This is an advantage. I don't know what is happening here, but I prefer it as it requires less effort," and another participant (p13) remarked, "It's easy. You don't really have to move that much, and you can do it while you're sitting as well." Some participants expressed their preference for the vibrational feedback. One participant (p1) said, "The vibration is good feedback, and I don't have to visually see anything," and another participant (p22) noted, "The process is like listen to my heart." In addition, some participants conveyed that they found the pairing process to be enjoyable and fun. For example, 3 participants described the vibration as a hand massage and 2 compared the pairing activity to using a stethoscope.

Most participants (18 out of 24) experienced no discomfort during the study. Nonetheless, the rest of the 6 people did report some discomfort at the end of the study. In total, 4 participants noted that the intensity of the vibrations was excessive; for example, one participant (p4) stated, "I feel like my entire chest is vibrating, and I don't like the feeling." This concern might be resolved by selecting a vibration motor with lower amplitude. In addition, 4 participants reported feeling fatigued after the data collection process, but also noted this was due to the repetition of 180 attachments and that less motions will alleviate this issue. Furthermore, one participant (p3) criticized the prototype design and mentioned that the external device simulator was too big. We leave the refinement of our prototype as future work.

Valuable insights were also gathered from the co-design workshop. Both participants initially found the vibration-based pairing technique interesting and somewhat surprising, but they quickly became accustomed to it and could easily complete the remaining required motions. They both explicitly noted that the pairing operations were easy to learn and perform. The cardiology doctor described the pairing operation as "using a stethoscope" and confirmed that the vibration signal in the experiment would pose minimal risks to patients with IMDs. Both participants also appreciated the tactile feedback from the vibrations. The doctor commented, "The vibration tells you if you're on track," while the pacemaker patient added, "The vibration encourages me towards the end of pairing." However, both participants pointed out that the prototype used in the study was bulky and heavy—an improvement we leave for future work. Overall, both participants found the pairing experience acceptable and expressed willingness to use it in real-world scenarios if required.

Optimal Setups

Based on our analysis so far, we summarize all pairing configurations that (1) exhibit high accuracy with zero FAR and FRR under 5%, and (2) generate a level of entropy surpassing a standard four-digit PIN code (with an entropy of

13.3 bits), which is commonly used in pairing of Bluetooth technologies and other security systems [8]. All viable settings that meet these requirements (with minimum required number of motions) are shown in Table 4. Note that the time values include both the vibration duration and an additional "preparation time" necessary for a user to detach and then reattach the external device to their body; in our study, this interval was 0.5 seconds.

In summary, we find that with a vibration configuration set at 100 Hz and 700 ms, a user can carry out 4 attachment motions to enable the exchange of a secret with (FAR, FRR)=(0, 0.6%) and entropy of 15.5 bits. This process can be completed in a mere 4.8 seconds.

Table . Summary of well-performing pairing configurations.

Vibration condition	Motion, n	FAR[a], FRR[b] (%)	Entropy	Time (s)
50 Hz, 1000 ms	5	0, 3.7	14.7	7.5
75 Hz, 700 ms	5	0, 2.2	18.0	6
75 Hz, 1000 ms	4	0, 3.5	15.0	6
100 Hz, 700 ms	4	0, 0.6	15.5	4.8
100 Hz, 1000 ms	4	0, 0.6	16.8	6

[a]FAR: false acceptance rate.

[b]FRR: false rejection rate.

Discussion

Principal Findings

Our work introduces a new and reliable vibration-based pairing approach for IMDs, which only requires a low sampling rate accelerometer and relies on the natural randomness inherent in human behavior. We empirically validate the feasibility of our technique through a user study. Overall, we find that the workload required to bootstrap a secure pairing is minimal, and we estimate that it requires the user to attach a device to the body only 4 times in roughly 5 s. With an FAR of 0 and an FRR of 0.6%, the risk posed by adversaries is low, and legitimate users will likely experience very few failures.

As mentioned in the related work section, the use of a PAKE eliminates offline brute-force attacks. In addition, it also restricts the number of online MITM attempts. Typically, the adversary has a very limited period to obtain the secret and usually only one chance for a MITM attack [44]. As an estimate, 4 motions with 15.5 bits entropy reduce the adversary's success probability on online brute-force attacks to 0.002% [42] (assuming the adversary is limited to guessing only). Therefore, we believe these motions serve as adequate input for a PAKE. If needed, higher entropy can be easily achieved by performing more motions.

Our user study confirmed the high usability of our pairing method. Participants found it straightforward to understand, learn, and perform. The process of attaching the device is very intuitive, like using a stethoscope as described by the participants. Our technique also brings about certain entertainment to users, being both relaxing and enjoyable (such as described as hand massage). This could be advantageous in certain therapeutic treatments, where physical interaction can enhance memory, concentration, and mental health [64]. Moreover, it is worth noting that for patients who are unable (eg, due to disabilities or unconsciousness in emergencies) or unwilling to execute the motions, our pairing allows medical practitioners or caregivers (who have received appropriate training) to execute the motions on the patient's body on their behalf.

Our proposed method only requires an accelerometer, a component already present in the latest generation IMDs [23-25]. The signal processing and other cryptographic algorithms for the IMD are computationally lightweight and work efficiently on 32-bit Cortex-M microcontrollers, which closely resemble IMDs' capabilities [11,43,53]. Our approach solely depends on vibration at a constant frequency, which can be easily implemented on readily available consumer devices such as smartphones and tablets. This is beneficial considering that medical device companies already equip the IMDs with the ability to connect to personal mobile devices [65]. Moreover, while our work assumes that the IMD is a pacemaker, we argue that the technique can be easily transferred to other types of IMDs or even external wearables. Furthermore, our proposed pairing technique incurs minimal costs. In our prototype implementation, the combined cost of the vibration motor and accelerometer was under $30, and this cost could be further reduced during mass production.

Comparison With Prior Vibration-Based Work

Our pairing technique significantly relaxes the demands on the IMD's sampling capability. We use an accelerometer operating at 250 Hz, in contrast to previous work that often relies on sampling rates of several thousand Hz or more. In particular, the sampling rate can be further decreased by using lower vibration frequencies. For example, with a 50 Hz vibration, the frequency domain peaks cluster between 30 and 70 Hz (see Figure 7), indicating that an accelerometer with a 140 Hz maximum is adequate.

Conventional approaches typically try to avoid user-generated noise. For instance, the user needs to ensure stable contact between devices during data transmission. Conversely, our method harnesses user noise and benefits from it as a source of entropy. Indeed, our dataset includes many instances with significant user error, like when a participant releases the external device before the vibration completely stops. In such

scenarios, the IMD only captures a portion of the vibration within its measurement window. Despite this, our technique maintains high reliability.

Furthermore, previous work that encodes secrets into vibrations often demands precise time synchronization in milliseconds between devices, which itself is a challenging task for resource-constrained devices [66]. In contrast, our approach allows for more lenient synchronization—as long as the two devices capture similar vibration signals within most of their measurement windows, the peak locations effectively match. This aspect greatly enhances the feasibility of our technique for IMDs.

Notably, our data throughput is significantly lower than [26,27,29,32] and is comparable with [31,33]. Considering the scenario of transmitting a 4-digit PIN code for use in a PAKE, previous work [26,27,29,32] only needs 0.0004 to 0.665 s, which is much faster than the 4.8 seconds required by our method. However, this rapid transmission, while advantageous in many daily applications, may not be suitable for IMD pairing contexts, where the vibration serves not only for secret exchange but also as a crucial cue for patients to be aware of the pairing process. In contrast, we argue that a duration of 4.8 s strikes a balance: it is long enough to be noticeable, yet short enough to maintain usability and safety in emergencies.

Considerations of Health Implications With Vibrations

Our proposed pairing technique incorporates vibration, a feature that naturally raises concerns regarding the potential long-term health impacts on patients. However, current research indicates that only long-term and excessive exposure to vibrations is linked to adverse effects on mental and physical health [67,68]. In contrast, our method involves brief vibrational interactions, which last less than 5 seconds and may not occur every day. This limited exposure could reduce the likelihood of the negative health consequences.

Resilience to Acoustic Eavesdropping Attacks

Vibration is essentially a low-frequency audio signal, which inevitably emits acoustic side-channel information that might be eavesdropped using a microphone. This is particularly threatening for methods that encode secrets within vibration signals. For example, Halevi and Saxena [69] found that secrets transmitted this way could be severely compromised using an off-the-shelf microphone from a few meters away. To mitigate this, Kim et al [26] and Anand and Saxena [70,71] proposed using Gaussian white noise or masking signals to obscure the acoustic leaks. These approaches have shown promise in reducing side-channel vulnerabilities against advanced eavesdropping attacks.

In comparison, as shown in [31], the risks associated with eavesdropping are significantly reduced when the vibration is not the carrier of the secret. Our research aligns with this guideline, using a constant vibration signal across sessions to minimize acoustic leakage. In addition, existing countermeasures [26,70,71] are also applicable to our method.

Limitations

Our work has certain limitations. Our experiments did not explicitly recruit participants who were IMD patients (mainly due to ethics constraints of the institutions where the user study was conducted). Further validation of our approach with these patient groups is necessary.

We designed our prototype in line with previous work in the IMD security community [5,26,40]. However, there is room for enhancement, particularly in its size and weight. Future research should develop more skin-conformable and miniaturized prototypes.

Another aspect of future work is to empirically evaluate the susceptibility of our pairing technique against microphone-based eavesdropping attacks at a distance.

Conclusion

In this paper, we explore the potential of leveraging vibration to pair with an IMD. We propose a novel technique that uses a straightforward constant-frequency vibration to extract secrets from natural and random human motor behavior for device pairing. We implement and validate our technique through a user study. Overall, we show that it is feasible to establish a cryptographic key in 5 s with high usability, based only on standard vibration motors and accelerometers with low sampling capabilities. The ubiquity of accelerometers in today's commercial smart devices and IMDs maximizes the chance of acceptance of our design. In general, we hope that our work will serve as a reference for pairing with resource-constrained devices using vibrations in body area networks.

Acknowledgments

MZ is funded by the Priestley PhD Scholarship program organized by the University of Melbourne and University of Birmingham.

Data Availability

The data collected from the user study and the software used in our prototype are publicly available at GitHub [72]. All data have been anonymized in accordance with the requirements of the institutional ethics committees to protect participant privacy. Additional supporting documentations are included in the repository to facilitate reuse.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire and interview design.
[DOCX File, 13 KB - biomedeng_v10i1e57091_app1.docx]

References

1. Global number of pacemakers in 2016 and a forecast for 2023. Statista. 2022. URL: https://www.statista.com/statistics/800794/pacemakers [accessed 2025-08-18]
2. Halperin D, Heydt-Benjamin TS, Fu K, Kohno T, Maisel WH. Security and privacy for implantable medical devices. IEEE Pervasive Comput 2008;7(1):30-39. [doi: 10.1109/MPRV.2008.16]
3. Marin E, Singelée D, Garcia FD, Chothia T, Willems R, Preneel B. On the (in)security of the latest generation implantable cardiac defibrillators and how to secure them. Presented at: ACSAC '16: 2016 Annual Computer Security Applications Conference; Dec 5-8, 2016; Los Angeles, California, USA. [doi: 10.1145/2991079.2991094]
4. Marin E, Singelée D, Yang B, Verbauwhede I, Preneel B. On the feasibility of cryptography for a wireless insulin pump system. Presented at: CODASPY'16: Sixth ACM Conference on Data and Application Security and Privacy; Mar 9-11, 2016; New Orleans, Louisiana, USA. [doi: 10.1145/2857705.2857746]
5. Marin E, Singelée D, Yang B, et al. Securing wireless neurostimulators. Presented at: CODASPY '18: Eighth ACM Conference on Data and Application Security and Privacy; Mar 19-21, 2018; Tempe, AZ, USA. [doi: 10.1145/3176258.3176310]
6. Reverberi L, Oswald D. Breaking (and fixing) a widely used continuous glucose monitoring system. Presented at: 11th USENIX Workshop on Offensive Technologies; Aug 14-15, 2017; Vancouver, Canada.
7. Daemen J, Rijmen V. AES proposal: Rijndael. CMU School of Computer Science. 1999. URL: https://www.cs.cmu.edu/afs/cs/project/pscico-guyb/realworld/www/docs/rijndael.pdf [accessed 2025-08-18]
8. Bisdikian C. An overview of the Bluetooth wireless technology. IEEE Commun Mag 2001;39(12):86-94. [doi: 10.1109/35.968817]
9. Rocha EA, Cunha GS, Tavares AB, et al. Syncope in patients with cardiac pacemakers. Braz J Cardiovasc Surg 2021 Feb 1;36(1):18-24. [doi: 10.21470/1678-9741-2020-0076] [Medline: 33594860]
10. Adams C, Lloyd S. Understanding PKI: Concepts, Standards, and Deployment Considerations: Addison-Wesley Professional; 2003.
11. Zhang M, Marin E, Oswald D, Singelée D. FuzzyKey: comparing fuzzy cryptographic primitives on resource-constrained devices. In: Grosso V, Pöppelmann T, editors. In Smart Card Research and Advanced Applications: Springer International Publishing; 2022:289-309.
12. Rushanan M, Rubin AD, Kune DF, Swanson CM. SoK: security and privacy in implantable medical devices and body area networks. In: 2014 IEEE Symposium on Security and Privacy (SP): IEEE; 2014:524-539. [doi: 10.1109/SP.2014.40]
13. Rasmussen KB, Castelluccia C, Heydt-Benjamin TS, Capkun S. Proximity-based access control for implantable medical devices. Presented at: CCS '09; Nov 9-13, 2009; Chicago, Illinois, USA. [doi: 10.1145/1653662.1653712]
14. Tomlinson WJ, Banou S, Yu C, Nogueira M, Chowdhury KR. Secure on-skin biometric signal transmission using galvanic coupling. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications: IEEE; 2019:1135-1143. [doi: 10.1109/INFOCOM.2019.8737540]
15. Denning T, Fu K, Kohno.2008 T. Absence makes the heart grow fonder: new directions for implantable medical device security. Presented at: HOTSEC'08: Proceedings of the 3rd conference on Hot topics in security; Jul 29, 2008; San Jose, CA, USA.
16. Gollakota S, Hassanieh H, Ransford B, Katabi D, Fu K. They can hear your heartbeats: non-invasive security for implantable medical devices. In: Proceedings of the ACM SIGCOMM 2011 Conference: Association for Computing Machinery; 2011:2-13. [doi: 10.1145/2018436.2018438]
17. Lin Q, Xu W, Liu J, et al. H2B: heartbeat-based secret key generation using piezo vibration sensors. In: Proceedings of the 18th International Conference on Information Processing in Sensor Networks: Association for Computing Machinery; 2019:265-276. [doi: 10.1145/3302506.3310406]
18. Poon CCY. A novel biometrics method to secure wireless body area sensor networks for telemedicine and m-health. IEEE Commun Mag 2006;44(4):73-81. [doi: 10.1109/MCOM.2006.1632652]
19. Rostami M, Juels A, Koushanfar F. Heart-to-heart (H2H): authentication for implanted medical devices. In: Proceedings of Conference on Computer and Communications Security: Association for Computing Machinery; 2013:1099-1112. [doi: 10.1145/2508859.2516658]
20. Marin E, Argones Rúa E, Singelée D, Preneel B. On the difficulty of using patient's physiological signals in cryptographic protocols. In: Proceedings of the 24th ACM Symposium on Access Control Models and Technologies: Association for Computing Machinery; 2019:113-122. [doi: 10.1145/3322431.3325099]
21. Dierks T, Allen C. The TLS protocol version 1.0. : The Internet Society; 1999.
22. Ryu J. Psychophysical model for vibrotactile rendering in mobile devices. Presence (Camb) 2010:364-387. [doi: 10.1162/PRES_a_00011]

23. Rate response feature. Medtronic. 2016. URL: https://www.medtronicacademy.com/features/rate-response-rr-feature [accessed 2025-08-18]
24. Pacing & defibrillation. Cardiocases. URL: https://www.cardiocases.com/en/pacingdefibrillation [accessed 2025-08-18]
25. Puppala VK, Hofeld BC, Anger A, et al. Pacemaker detected active minutes are superior to pedometer-based step counts in measuring the response to physical activity counseling in sedentary older adults. BMC Geriatr 2020 May 6;20(1):162. [doi: 10.1186/s12877-020-01559-y] [Medline: 32375676]
26. Kim Y, Lee WS, Raghunathan V, Jha NK, Raghunathan A. Vibration-based secure side channel for medical devices. In: Proceedings of the 52nd Annual Design Automation Conference: Association for Computing Machinery; 2015:1-6. [doi: 10.1145/2744769.2744928]
27. Lee K, Raghunathan V, Raghunathan A, Kim Y. SYNCVIBE: fast and secure device pairing through physical vibration on commodity smartphones. In: 2018 IEEE 36th International Conference on Computer Design (ICCD: IEEE; 2018:234-241. [doi: 10.1109/ICCD.2018.00043]
28. Roy N, Choudhury RR. Ripple {II}: faster communication through physical vibration. Presented at: 13th USENIX Symposium on Networked Systems Design and Implementation; Apr 2-5, 2013; Lombard, IL, USA.
29. Roy N, Gowda M, Choudhury RR. Ripple: communicating through physical vibration. Presented at: 12th USENIX Symposium on Networked Systems Design and Implementation; Apr 25-27, 2012; San Jose.
30. Saxena N, Uddin MB, Voris J. Treat'em like other devices: user authentication of multiple personal RFID tags. In: SOUPS 2009, Vol. 9:1-1. [doi: 10.1145/1572532.1572573]
31. Wang W, Yang L, Zhang Q. Resonance-Based Secure Pairing for Wearables. IEEE Trans on Mobile Comput 2018;17(11):2607-2618. [doi: 10.1109/TMC.2018.2809736]
32. Xiao R, Mayer S, Harrison C. VibroComm: using commodity gyroscopes for vibroacoustic data reception. Presented at: MobileHCI '20; Oct 5-8, 2020; Oldenburg, Germany. [doi: 10.1145/3379503.3403540]
33. Saxena N, Uddin MB, Voris J, Asokan N. Vibrate-to-unlock: mobile phone assisted user authentication to multiple personal RFID tags. In: 2011 IEEE International Conference on Pervasive Computing and Communications (PerCom): IEEE; 2011:181-188. [doi: 10.1109/PERCOM.2011.5767583]
34. de Vaal MH, Neville J, Scherman J, Zilla P, Litow M, Franz T. The in vivo assessment of mechanical loadings on pectoral pacemaker implants. J Biomech 2010 Jun 18;43(9):1717-1722. [doi: 10.1016/j.jbiomech.2010.02.028] [Medline: 20202638]
35. Stenerson M, Cameron F, Payne SR, et al. The impact of accelerometer use in exercise-associated hypoglycemia prevention in type 1 diabetes. J Diabetes Sci Technol 2015 Jan;9(1):80-85. [doi: 10.1177/1932296814551045] [Medline: 25231116]
36. Stenerson M, Cameron F, Wilson DM, et al. The impact of accelerometer and heart rate data on hypoglycemia mitigation in type 1 diabetes. J Diabetes Sci Technol 2014 Jan;8(1):64-69. [doi: 10.1177/1932296813516208]
37. Khan A, Hammerla N, Mellor S, Plötz T. Optimising sampling rates for accelerometer-based human activity recognition. Pattern Recognit Lett 2016 Apr;73:33-40. [doi: 10.1016/j.patrec.2016.01.001]
38. Adewusi SA, Rakheja S, Marcotte P, Boutin J. Vibration transmissibility characteristics of the human hand–arm system under different postures, hand forces and excitation levels. J Sound Vib 2010 Jul;329(14):2953-2971. [doi: 10.1016/j.jsv.2010.02.001]
39. Adewusi S, Thomas M, Vu VH, Li W. Modal parameters of the human hand-arm using finite element and operational modal analysis. Mechanics & Industry 2014;15(6):541-549. [doi: 10.1051/meca/2014060]
40. Halperin D, Heydt-Benjamin TS, Ransford B, et al. Pacemakers and implantable cardiac defibrillators: software radio attacks and zero-power defenses. 2008 Presented at: 2008 IEEE Symposium on Security and Privacy p. 129-142.
41. Schechter S. Security that is meant to be skin deep: using ultraviolet micropigmentation to store emergency-access keys for implantable medical devices. Presented at: 1st USENIX Workshop on Health Security and Privacy; Aug 10, 2010; Washington, DC.
42. Katz J, Ostrovsky R, Yung M. Forward secrecy in password-only key exchange protocols. In: International Conference on Security in Communication Networks: Springer; 2002:29-44. [doi: 10.1007/3-540-36413-7_3]
43. Li X, Zeng Q, Luo L, Luo T. T2Pair: secure and usable pairing for heterogeneous iot devices. In: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security: Association for Computing Machinery; 2020:309-323. [doi: 10.1145/3372297.3417286]
44. Bellovin SM, Merritt M. Encrypted key exchange: password-based protocols secure against dictionary attacks. Presented at: CCS93: 1st ACM Conference on Communications and Computing Security; Nov 3-5, 1993; Fairfax, Virginia, USA. [doi: 10.1145/168588.168618]
45. Han J, Chung AJ, Sinha MK, et al. Do you feel what i hear? Enabling autonomous iot device pairing using different sensor types. In: 2018 IEEE Symposium on Security and Privacy (SP): IEEE; 2018:836-852. [doi: 10.1109/SP.2018.00041]
46. Introduction. CMSIS. 2022. URL: https://www.keil.com/pack/doc/CMSIS/DSP/html/group RealFFT.html [accessed 2025-08-18]
47. Zhang L, Pathak PH, Wu M, Zhao Y, Mohapatra P. Accelword: energy efficient hotword detection through accelerometer. Presented at: MobiSys'15: The 13th Annual International Conference on Mobile Systems, Applications, and Services; May 18-22, 2015; Florence Italy. [doi: 10.1145/2742647.2742658]
48. Doran RW. The gray code. J Univers Comput Sci 2007 [FREE Full text]

49. Dodis Y, Ostrovsky R, Reyzin L, Smith A. Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. SIAM J Comput 2008 Jan;38(1):97-139. [doi: 10.1137/060651380]
50. Xu F, Qin Z, Tan CC, Wang B, Li Q. IMDGuard: securing implantable medical devices with the external wearable guardian. Presented at: IEEE Annual Joint Conference: INFOCOM, IEEE Computer and Communications Societies; Apr 10-15, 2011; Shanghai, China.
51. Dolev D, Yao A. On the security of public key protocols. IEEE Trans Inform Theory 1983 Mar;29(2):198-208. [doi: 10.1109/TIT.1983.1056650]
52. MPU-6000 and MPU-6050 product specification revision 34. Adafruit. 2013. URL: https://cdn-learn.adafruit.com/downloads/pdf/mpu6050-6-axis-accelerometer-and-gyro.pdf [accessed 2025-08-18]
53. Azure pacing system. Medtronic. 2022. URL: https://europe.medtronic.com/xd-en/healthcare-professionals/products/cardiac-rhythm/pacemakers/azure.html [accessed 2025-08-18]
54. Model no307-103 vibration motor. Precision Microdrives. 2021. URL: https://catalogue.precisionmicrodrives.com/product/datasheet/307-103-9mm-vibration-motor-25mm-type-datasheet.pdf [accessed 2025-08-18]
55. Petronio AS, Sinning JM, Van Mieghem N, et al. Optimal implantation depth and adherence to guidelines on permanent pacing to improve the results of transcatheter aortic valve replacement with the Medtronic CoreValve System: The CoreValve Prospective, International, Post-Market ADVANCE-II Study. JACC Cardiovasc Interv 2015 May;8(6):837-846. [doi: 10.1016/j.jcin.2015.02.005] [Medline: 25999108]
56. Lim WY, Prabhu S, Schilling RJ. Implantable cardiac electronic devices in the elderly population. Arrhythm Electrophysiol Rev 2019 May 2;8(2):143-146. [doi: 10.15420/aer.2019.3.4]
57. Lewis JR. The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
58. Ahmed I, Ye Y, Bhattacharya S, et al. Checksum gestures: continuous gestures as an out-of-band channel for secure pairing. Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2807521]
59. Li X, Yan F, Zuo F, Zeng Q, Luo L. Touch well before use: intuitive and secure authentication for iot devices. Presented at: MobiCom '19: The 25th Annual International Conference on Mobile Computing and Networking; Oct 21-25, 2019; Los Cabos, Mexico. [doi: 10.1145/3300061.3345434]
60. Rukhin A, Soto J, Nechvatal J, et al. A statistical test suite for random and pseudorandom number generators for cryptographic applications. : NIST; 2001.
61. Mayrhofer R, Gellersen H. Shake well before use: intuitive and secure pairing of mobile devices. IEEE Trans on Mobile Comput 2009;8(6):792-806. [doi: 10.1109/TMC.2009.51]
62. Thomas M, Joy AT. Elements of Information Theory: Wiley-Interscience; 2006.
63. Lin S, Costello DJ. Error Control Coding: Prentice Hall; 2001, Vol. 2.
64. Hedayati M, Sum S, Hosseini SR, Faramarzi M, Pourhadi S. Investigating the effect of physical games on the memory and attention of the elderly in adult day-care centers in Babol and Amol. Clin Interv Aging 2019;14:859-869. [doi: 10.2147/CIA.S196148] [Medline: 31190772]
65. MyCareLink heart mobile app. Medtronic. 2021. URL: https://global.medtronic.com/xg-en/mobileapps/patient-caregiver/cardiac-monitoring/mycarelink-heart-app.html [accessed 2025-08-18]
66. Wang C, Sarsenbayeva Z, Luo C, Goncalves J, Kostakos V. Improving wearable sensor data quality using context markers. Presented at: UbiComp '19: The 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 9-13, 2019; London United Kingdom. [doi: 10.1145/3341162.3349334]
67. Bovenzi M. Health effects of mechanical vibration. G Ital Med Lav Ergon 2005;27(1):58-64. [Medline: 15915675]
68. Seidel H. Selected health risks caused by long‐term, whole‐body vibration. American J Industrial Med 1993 Apr;23(4):589-604. [doi: 10.1002/ajim.4700230407]
69. Halevi T, Saxena N. On pairing constrained wireless devices based on secrecy of auxiliary channels: the case of acoustic eavesdropping. Presented at: CCS '10: 17th ACM Conference on Computer and Communications Security 2010; Oct 4-8, 2010; Chicago, Illinois, USA. [doi: 10.1145/1866307.1866319]
70. Anand SA, Saxena N. Coresident evil: noisy vibrational pairing in the face of co-located acoustic eavesdropping. Presented at: WiSec '17: 10th ACM Conference on Security & Privacy in Wireless and Mobile Networks; Jul 18-20, 2017; Boston, Massachusetts. [doi: 10.1145/3098243.3098256]
71. Anand SA, Saxena N. Noisy vibrational pairing of IoT devices. IEEE Trans Dependable Secure Comput 2018:530-545 [FREE Full text]
72. MrZMN/vibkey. GitHub. URL: https://github.com/MrZMN/VibKey [accessed 2025-08-18]

Abbreviations

FAR: false acceptance rate
FFT: fast Fourier transform
FRR: false rejection rate

IMD: implantable medical device
IoT: Internet of things
MITM: man in the middle
NIST: National Institute of Standards and Technology
OOB: out-of-band
PAKE: password-authenticated key agreement
SUS: system usability scale

Edited by S Rizvi, T Leung; submitted 05.02.24; peer-reviewed by K Daripa, Y Liu; revised version received 09.07.25; accepted 09.07.25; published 26.08.25.

Please cite as:
Zhang M, Wang C, Jiang W, Oswald D, Murray T, Marin E, Wei J, Ryan M, Kostakos V
Using Vibration for Secure Pairing With Implantable Medical Devices: Development and Usability Study
JMIR Biomed Eng 2025;10:e57091
URL: https://biomedeng.jmir.org/2025/1/e57091
doi:10.2196/57091

Estimation of Brachial-Ankle Pulse Wave Velocity With Hierarchical Regression Model From Wrist Photoplethysmography and Electrocardiographic Signals: Method Design

Chih-I Ho[1], MEng; Chia-Hsiang Yen[1], BEng; Yu-Chuan Li[1], MEng; Chiu-Hua Huang[1], MEng; Jia-Wei Guo[1], MEng; Pei-Yun Tsai[2], PhD; Hung-Ju Lin[3], MD, PhD; Tzung-Dau Wang[3], MD, PhD

[1]Department of Electrical Engineering, National Central University, Taoyuan, Taiwan

[2]Graduate School of Advanced Technology, National Taiwan University, Taipei, Taiwan

[3]Cardiovascular Center and Divisions of Cardiology and Hospital Medicine, Department of Internal Medicine, National Taiwan University Hospital, No.7, Chung Shan S Rd, Taipei, Taiwan

Corresponding Author:
Tzung-Dau Wang, MD, PhD
Cardiovascular Center and Divisions of Cardiology and Hospital Medicine, Department of Internal Medicine, National Taiwan University Hospital, No.7, Chung Shan S Rd, Taipei, Taiwan

Abstract

Background: Photoplethysmography (PPG) signals captured by wearable devices can provide vascular age information and support pervasive and long-term monitoring of personal health condition.

Objective: In this study, we aimed to estimate brachial-ankle pulse wave velocity (baPWV) from wrist PPG and electrocardiography (ECG) from smartwatch.

Methods: A total of 914 wrist PPG and ECG sequences and 278 baPWV measurements were collected via the smartwatch from 80 men and 82 women with average age of 63.4 (SD 13.4) and 64.3 (SD 11.6) years. Feature extraction and weighted pulse decomposition were applied to identify morphological characteristics regarding blood volume change and component waves in preprocessed PPG and ECG signals. A systematic strategy of feature combination was performed. The hierarchical regression method based on the random forest for classification and extreme gradient boosting (XGBoost) algorithms for regression was used, which first classified the data into subdivisions. The respective regression model for the subdivision was constructed with an overlapping zone.

Results: By using 914 sets of wrist PPG and ECG signals for baPWV estimation, the hierarchical regression model with 2 subdivisions and an overlapping zone of 400 cm per second achieved root-mean-square error of 145.0 cm per second and 141.4 cm per second for 24 men and 26 women, respectively, which is better than the general XGBoost regression model and the multivariable regression model (all $P<.001$).

Conclusions: We for the first time demonstrated that baPWV could be reliably estimated by the wrist PPG and ECG signals measured by the wearable device. Whether our algorithm could be applied clinically needs further verification.

(JMIR Biomed Eng 2025;10:e58756) doi:10.2196/58756

KEYWORDS

photoplethysmography; PPG; pulse wave velocity; brachial-ankle pulse wave velocity; XGBoost; electrocardiography; signal processing; random forest

Introduction

Cardiovascular disease (CVD) is a major cause of death and disability globally. Hemodynamic parameters are essential to the assessment of CVD risks. Arterial compliance is defined as the change of arterial blood volume for a given change in pressure and reflects the extent of arterial stiffness. Pulse wave velocity (PWV) describes the propagation of pulsatile activity due to ventricular ejection of blood and its interaction with arterial compliance [1]. Carotid-femoral PWV (cfPWV) and brachial-ankle PWV (baPWV) are associated with future CVD risk and commonly measured for clinic use. Compared with cfPWV, baPWV can be easily obtained by the oscillometric method with cuffs on the 4 limbs and is more widely used [2].

Owing to the advance of technology, wearable devices with automatic or self-assisted monitoring have been recognized as a promising tool to facilitate the assessment and management of CVD risks. Photoplethysmography (PPG) [3,4], ballistocardiography [5,6], electrical bioimpedance [7], or tonometry [8] has been widely studied for these purposes. Due

to the ease of implementation, the optical PPG module is more often integrated into the wearable devices. The potential of estimation of BP [9,10] and PWV [11-13] from PPG signals attracts much attention.

Various approaches have been investigated to estimate PWV from PPG signals of different measurement sites [14]. The contour of PPG and its associated time interval features have been used to estimate either baPWV or cfPWV by approaches including multiple regression, artificial neural network, and support vector machine [15,16]. Most of the prior works used finger PPG signals for PWV estimation because of its clear contour and ease of feature extraction, compared with wrist PPG [17,18]. However, with the growing popularity of smartwatches as wearable health care devices, the use of wrist-based PPG in biomedical applications has attracted considerable attention. In this study, we aimed to estimate baPWV from wrist PPG and electrocardiography (ECG).

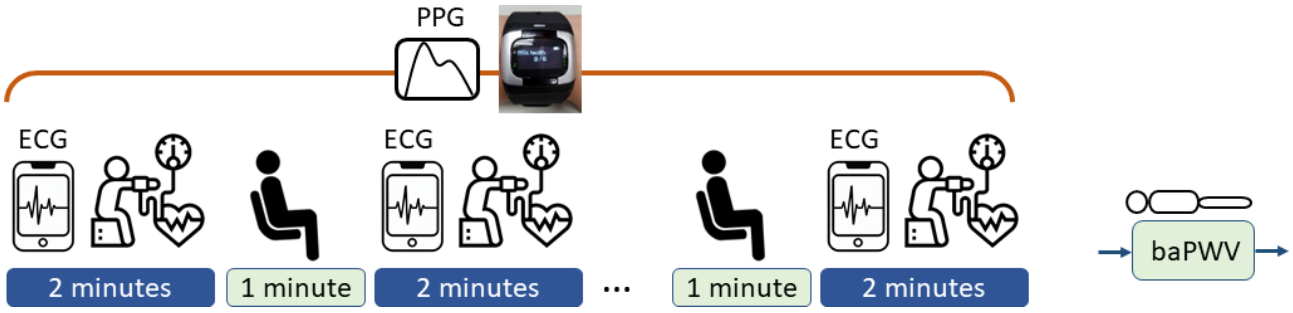
Methods

Methods and statistical analysis are briefly summarized in this section. Further details are provided in the Supplementary Section.

Data Collection

Figure 1 shows the measurement flow. Each volunteer wore a SENSIO smartwatch recording wrist PPG and ECG during the experimental period. For volunteers in the health management center, 3 rounds of measurements were conducted. For volunteers in the outpatient clinic, 5 rounds of measurements were made. In each round, the participants maintained the sitting position, and ECG was measured in the first minute. Blood pressures were then measured by the sphygmomanometer on the other arm (not wearing the smartwatch) with the cuff aligned at the heart level. A one-minute rest was reserved between 2 adjacent rounds. The wrist PPG signals were continuously recorded throughout the course. In the end, baPWV was measured by the OMRON noninvasive vascular screening device, with the cuffs on 4 limbs in the supine position.

Figure 1. Measurement flow. baPWV: brachial-ankle pulse wave velocity; ECG: electrocardiography; PPG: photoplethysmography.



Ethical Considerations

The experiment was approved by the research ethics committee of National Taiwan University Hospital (number 201902087RIPA). All data were collected in accordance with the approved protocol. Importantly, the dataset used in this study did not contain any personally identifiable information, and all records were fully anonymized prior to analysis. Informed consent was obtained from all participants, and the study was conducted in compliance with the ethical standards set forth in the Declaration of Helsinki and relevant national regulations.

Processing Flow

The signal-processing flow is indicated in Figure 2. The PPG and ECG, sampled at 256 Hz, were extracted from the first minute of each round in the synchronization phase (Figures S1 A and S1 B in Multimedia Appendix 1). In the preprocessing phase, baseline wandering of signals was corrected by the discrete wavelet transform, and the 60-Hz power interference was suppressed by the notch filter. The amplitude of the whole signal segment was then normalized to [−1, +1]. The R peak of ECG and the valley of PPG signals were detected to calculate cycle length (Figures S1 C and S1 D in Multimedia Appendix 1). The skewness and variation of ECG and PPG cycle lengths were adopted to establish the signal quality index to exclude suboptimal ECG or PPG cycles for feature extraction. The first-order derivative PPG (FDPPG) and the second-order derivative PPG (SDPPG) signals were calculated. The systolic peak, notch, and diastolic peak were marked by the algorithm [19] for each PPG cycle (Figure 3A). The maximal slope (max slope) of the ascending systolic pulse, corresponding to the maximal rate of blood volume change, was identified by the first local maximum in FDPPG (Figure 3B) [20]. The local extrema of the SDPPG in systole are defined as a, b, c, and d points, where points a and c are local maxima and points b and d are local minima (Figure 3C) [21]. Point e is the local maximum around the boundary of systole and diastole in SDPPG. Point f is the first local minimum after point e.

The PPG pulse is regarded as a summation of several component waves, including the forward waves by left ventricular contraction and the distally reflected waves due to aortic elasticity and reservoir property [22]. The pulse decomposition analysis helps segregate the component waves [23]. With proper weighting, the variation of component waves can be reduced [24]. Five Gaussian waves are used for synthesizing the PPG pulse. Given $\theta i=\alpha i,\beta i,\gamma i$ corresponding to pulse amplitude, pulse position, and pulse width of the component wave i, and $\Theta=\{\theta 1, \theta 2, \ldots,\theta 5\}$, the summation of the Gaussian waves takes the form of

$$(1)\quad G(t|\theta)=\sum i=15 g(t|\theta i)$$

with

(2)g(t|θi)=αie(t−βiTs)22(γiTs)2

Denote G_i as the component wave described by g(t|θi). Given the boundary constraints, Lαi≤αi≤Uαi, Lβi≤βi≤Uβi, and Lγi≤γi≤Uγi [24], the interior-point method is used to solve the following optimization problem,

(3)Θ^=arg minΘ1M∑n=1Mw(n)[s(n)−G(nTs|Θ)],

where w(n) is the weight to emphasize the informative portion of the PPG pulse sn with length M and is given by

(4)w(n)={ωna≤n≤nf1else

Variables na and nf refer to the position of points a and f. The weight ω is set to 80 for stabilizing the variation of component waves in the sequence with acceptable mean square error between the synthesized waveform and original waveform.

Once the component waves are acquired, the forward wave is generated by combining G_1 and G_2. The systolic wave and diastolic wave are derived by combining G_1 to G_3 and G_4 to G_5, respectively. The respective peaks of the synthesized forward wave, systolic wave, and diastolic wave are named as pf, ps, and pd. In the following, the amplitude and position of feature x in the PPG pulse are indicated by Ax and nx, respectively. The amplitude of feature x in the ith-order derivative PPG is represented byAx(i). The result of decomposed component waves by weighted pulse decomposition (WPD) is shown in Figure 4.

Figure 2. Signal-processing flow. ECG: electrocardiography; PPG: photoplethysmography; SQI: signal quality index; WPD: weighted pulse decomposition.

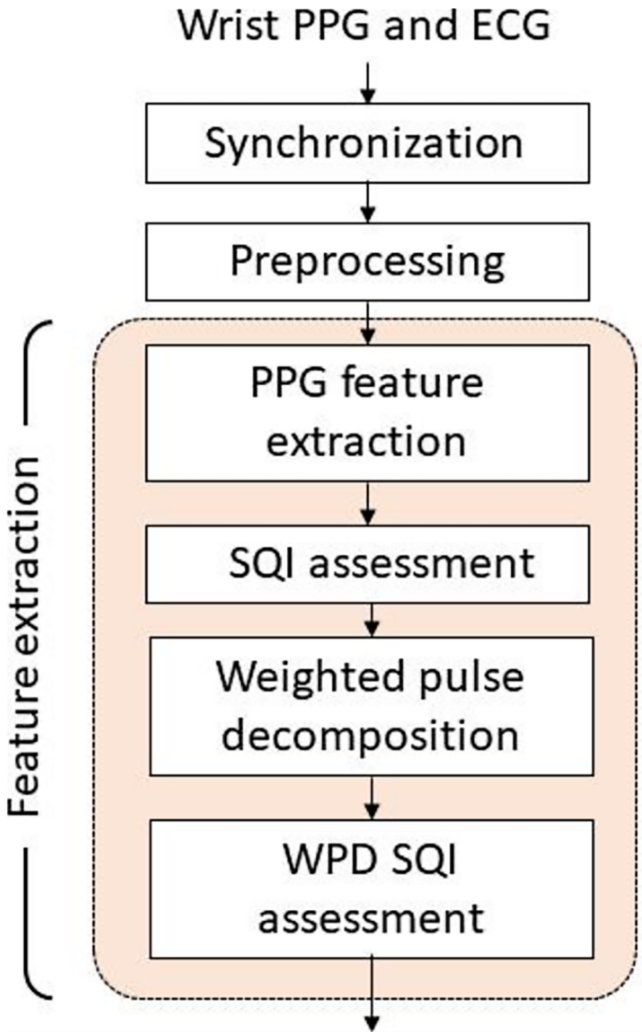


Figure 3. Photoplethysmography, first-order derivative photoplethysmography, and second-order derivative photoplethysmography waveforms and features in 1 cardiac cycle (from A to C). FDPPG: first-order derivative photoplethysmography; PPG: photoplethysmography; SDPPG: second-order derivative photoplethysmography.

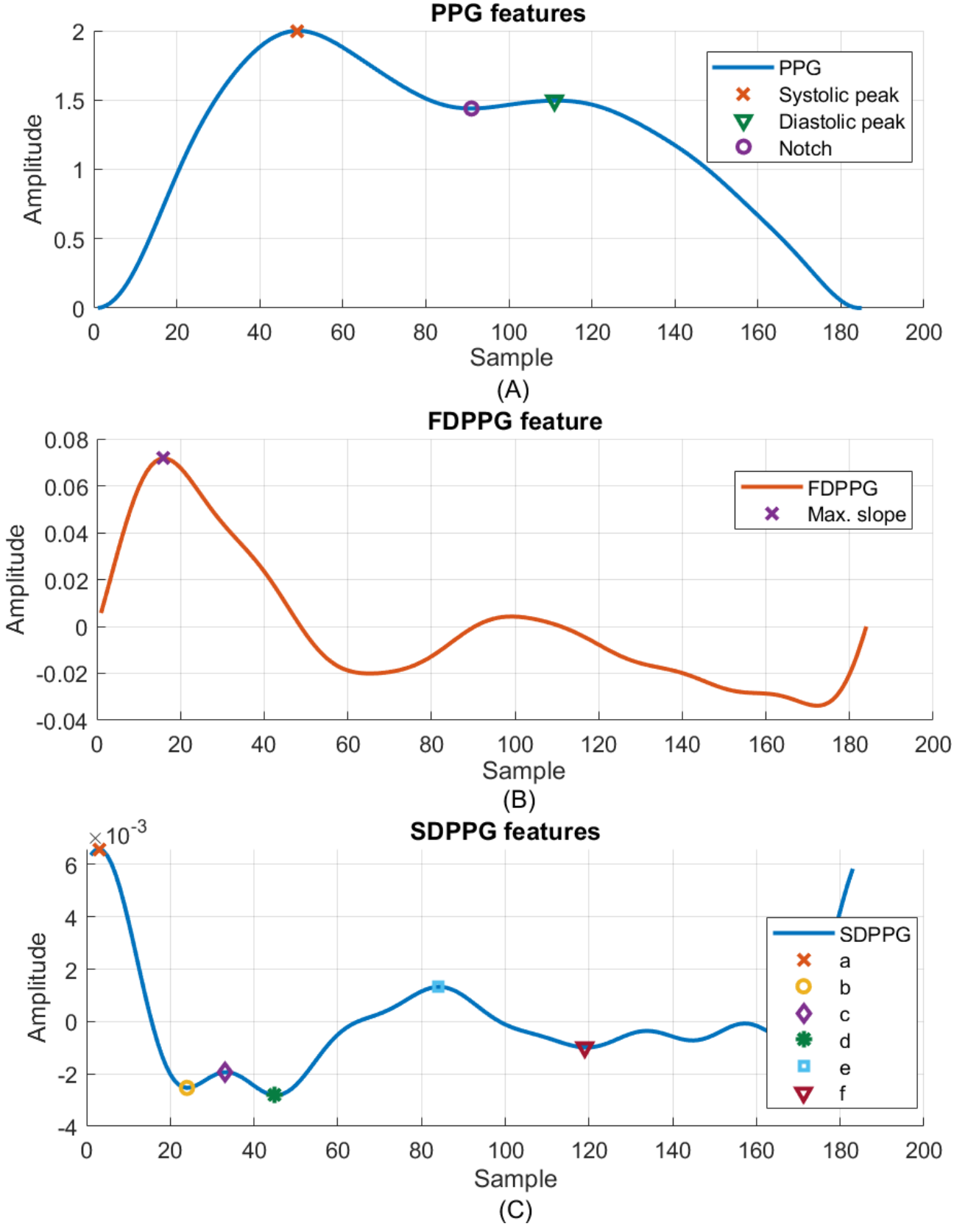
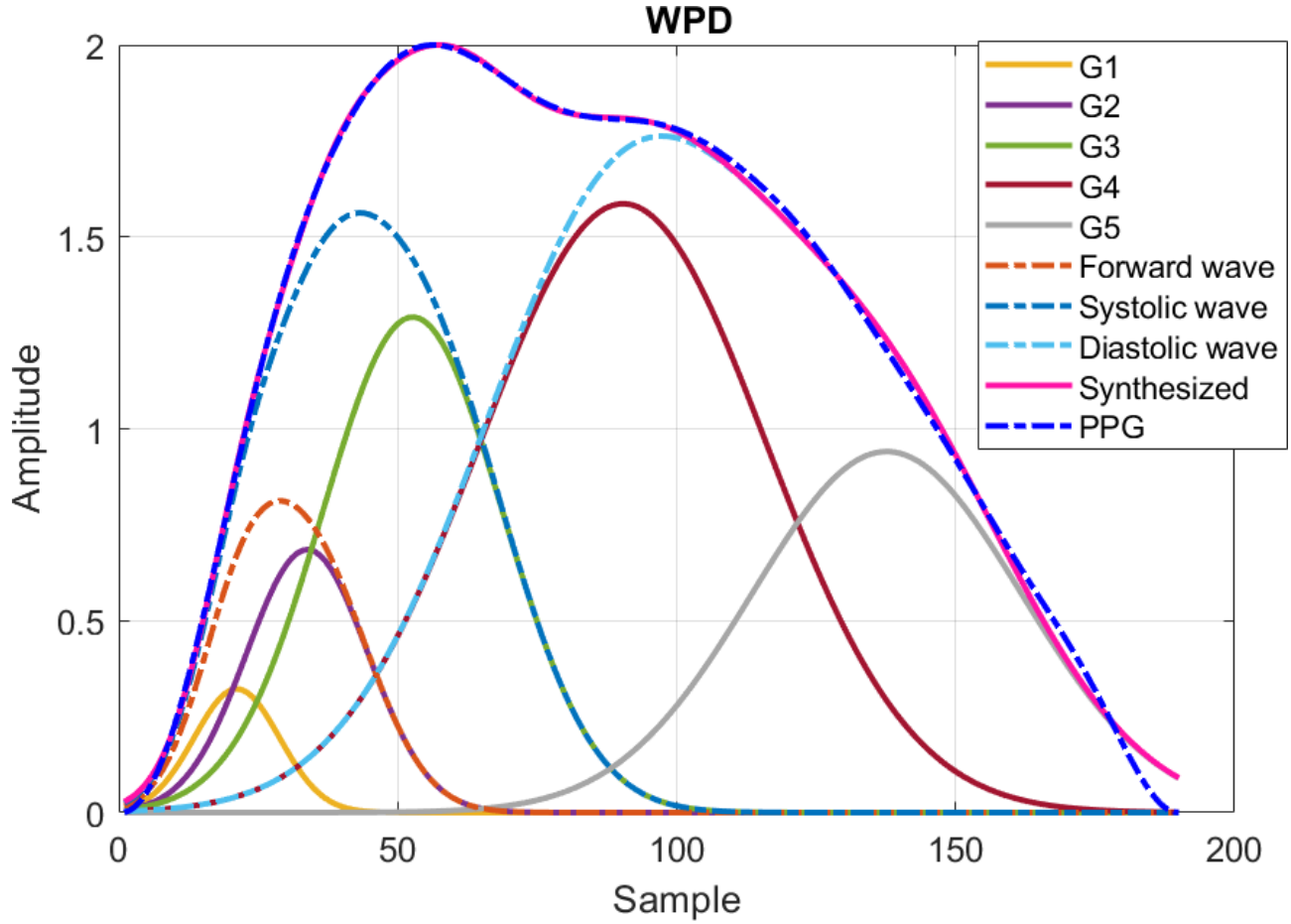


Figure 4. Component waves after weighted pulse decomposition. G1: Gaussian component wave 1; G2: Gaussian component wave 2; G3: Gaussian component wave 3; G4: Gaussian component wave 4; G5: Gaussian component wave 5; WPD: weighted pulse decomposition.



To assess the quality of WPD, WPD signal quality index, which was defined as mean square error between the PPG pulse, s(n), and the synthesized pulse, G(nTs|Θ), of >2×10-3, was implemented to remove disqualified pulses.

A total of 22 features were derived from the PPG pulse, FDPPG, and SDPPG (Table S1 A in Multimedia Appendix 2). The age index, which has been shown to be correlated with the augmentation index of aortic pressure [21,25],

(5)Ab(2)−Ac(2)−Ad(2)−Ae(2)Aa(2)

and its related variant combining only highly correlated components,

(6)Ab(2)−Ac(2)−Ad(2)Aa(2)

were also used. There were 27 features derived from WPD (Table S1 B in Multimedia Appendix 2). The stiffness index (SI) is defined as the time interval between the peaks of systolic and diastolic waves [23] and is denoted by npd-nps. The time intervals of the third or fourth component wave to the forward wave were also calculated. Note that nps and npd were obtained from synthesized systolic wave peak ps and diastolic wave peak pd of WPD as shown in Figure 4 while nsys and ndia were marked as the positions of systolic peak and diastolic peak in PPG as shown in Figure 3.

The ECG-related features were also adopted (Table S1 C in Multimedia Appendix 2). The R peak and T peak of the ECG waveform were identified and marked as nR and nT. Since the R peak occurs earlier than the PPG valley of the same heartbeat, nR is negative in number. The pulse arrival time (PAT) measures the time span between R peak and PPG valley, denoted by -nR. PAT^2 and $Height^2/PAT^2$ were included since either linear or nonlinear relationship between BP and pulse transit time has been shown [26]. The time span from R peak to maximum slope, peak of systolic wave, or component wave 2 was also considered.

Basic information (Table S1 D in Multimedia Appendix 2) contains age, height (H), weight, BMI, and lengths from arm to wrist (Law) and finger (Laf). The lengths from heart to brachium (Lb) and from heart to ankle (La) can be approximated by [27]

(7)La=0.8219H+12.328

(8)Lb=0.2195H−2.073.

The length difference between ankle and brachium could be expressed by La-Lb.

Feature normalization is often adopted since the relative change of 2 features could provide additional information than each feature alone. To systematically derive the normalization results, we generate combined features by dividing the value of feature u by value of feature v. The combined features contain not only magnitude-normalized or time-normalized features but also basic information features.

Estimation Approach

Multivariable Regression

Linear regression and multivariable regression had been applied for baPWV estimation [12,28]. The time difference between the systolic peak to diastolic peak has been used and normalized by the Fridericia formula [28] while the systolic peak to the next onset (P2O), M-nsys (feature 1 in Table S1 A in Multimedia Appendix 2), of the PPG signal normalized by the PPG pulse length was also examined for PWV estimation [12]. These 2 variables were selected from the finger PPG features by the authors due to their high correlation to baPWV reported in the literature. The wrist PPG was used in this study for baPWV estimation. Because diastolic peak often vanished in wrist PPG pulses, we used SI (feature 51 in Table S1 B in Multimedia Appendix 2), which denotes the time span between peaks of decomposed systolic wave and diastolic wave according to WPD, and its normalized form with the Friderician formula is given by SI/M1/3. The multivariable linear equations are described by [12,28]

(9)PWV=C1Age+C2SIM1/3+C3

and

(10)PWV=C1Age+C2P2OM+C3.

Hierarchical Regression

The linear estimation regarding the correlations between PPG features and PWV, as used in multivariable regression analysis, may oversimplify the vascular hemodynamic state. The machine learning algorithms have been prosperously developed and used for biomedical applications, such as neural network and decision tree regression for estimation of vascular age [29] and gradient boosting decision tree regression for estimation of blood pressure [30]. We herein developed the hierarchical regression model based on the random forest and extreme gradient boosting (XGBoost) algorithms. A general regression model by XGBoost was also implemented for comparison.

The random forest and XGBoost algorithms of high scalability have been shown to achieve excellent performance in many fields [31]. In the random forest algorithm, a large number of decision trees are constructed. A different subset of the data and a random selection of features are used for each decision tree to prevent overfitting in the training process. The final classification is often made by taking the majority vote. On the other hand, inherited from gradient boosting, XGBoost adds the new regression tree in each iteration to improve the previous prediction and to approach the target. The XGBoost introduces the regularization term that considers the complexity of the tree so as to avoid overfitting. In addition, the second-order gradient statistics are used for accelerating the computation.

The concept of hierarchical regression can be described as classification by random forest algorithm and then regression by XGBoost algorithm (Multimedia Appendix 3). The whole PWV range is partitioned into several subdivisions. Thus, a global classifier handles the entire PWV range, and several local regressors are in charge of the respective subdivisions. First, an outcome regarding the possible baPWV subdivision is generated by the global classifier. Then, the estimation result is calculated by the associated local regressor. Because it is possible that the data around the subdivision boundary are erroneously classified, the adjacent regressors are designed to have an overlapping zone to extend the respective coverages. Owing to the data quantity, 2 subdivisions were adopted and the boundary threshold was set at 1600 cm per second. The widths of the overlapping zone were set as 200 cm per second, 400 cm per second, and 600 cm per second.

Statistical Analysis

The differences between the estimated results v^j and the measured PWV vj of the jth measurement are shown by the mean absolute error, mean error, SD, and root-mean-square error (RMSE), which are defined as follows.

(11)ej=vj−vj^

(12)MAE=E{|ej|}

(13)ME=e−=E{ej}

(14)SD=1N−1∑j=1N(ej−e−)2

(15)RMSE=E{ej2}.

The correlation coefficients together with *P* values are also provided. Since some participants have more than 1 measurement, to avoid unbalanced weighting, averaged PWV estimation and averaged PWV measurement are used for the statistical results per participant.

Results

In this study, 80 male participants and 82 female participants were recruited. Their demographic characteristics are shown in Table 1. The averaged PWV value of left baPWV and right baPWV was used. The PWV values of male participants and female participants were 1591 (SD 266) cm per second and 1613 (SD 321) cm per second. Among total participants, 39 male participants and 23 female participants had more than 1 PWV values due to their multiple visits. A total of 914 PPG as well as ECG sequences were collected from the smartwatch, corresponding to 278 PWV values. On average, 1 male participant has 3.5 PPG and ECG sequences associated with 1 PWV measurement while 1 female participant has 3.1 PPG and ECG sequences for 1 PWV measurement. Among 278 PWV measurements, there are 123 PWV measurements from participants taking antihypertensive medications on the same day.

The medians of the respective combined features in the 528 and 386 sequences were used for computing correlation coefficients for men and women. The correlation coefficients of combined features defined by the X and Y indices are often higher than the original one (Multimedia Appendix 4). For example, the correlation coefficients of the age and maximum slope time (nms) to baPWV are 0.334 and −0.281, whereas the correlation coefficient of the combined feature Age/nms becomes 0.491 (Multimedia Appendix 5). The correlation coefficients of SI corrected by Friderician's formula and the time interval between systolic peak to the onset of next PPG (P2O) normalized by pulse length from the wrist PPG versus baPWV are −0.271 (*P*<.001), −0.036 (*P*=.413) and −.370 (*P*<.001), −0.070 (*P*=.171) for men and women, respectively.

The reproducibility of the measured baPWV was also checked. The PWVs of 31 participants were measured twice by the same OMRON noninvasive vascular screening device with 1-minute separation. The maximal differences of left baPWV and right baPWV of these participants were 276 cm per second and 210 cm per second, respectively. The maximal difference of averaged baPWV from left baPWV and right baPWV was 196.5 cm per second. The RMSEs of 2 consecutively measured left baPWV and right baPWV were 83.4 cm per second and 62.0 cm per second, respectively. The RMSE of consecutive averaged baPWV was 68.8 cm per second.

For multivariable regression, 39 and 34 PWV measurements from 24 male participants and 26 female participants, respectively, were reserved as the testing dataset. The medians of the respective features from the sequences associated with the same PWV measurement were averaged. The testing dataset was selected to approach uniform distribution in the range between 1000 cm per second and 2100 cm per second. The mean and SD of the male and female PWV values in the testing dataset were 1538 (SD 237) cm per second and 1638 (SD 283) cm per second. The training dataset for deriving the coefficients contained 114 PWV measurements with 391 PPG per ECG sequences from 56 male participants and 91 PWV measurements with 291 sequences from 56 female participants. The participant-split criterion is obeyed. The baPWV estimation results by multivariable regression are shown in Table 2 for men and women, respectively.

Table . Demographic summary.[a]

Characteristics	Male participants, mean (SD; n)	Female participants, mean (SD; n)
Age (years)	63.4 (13.4; 80)	64.3 (11.6; 82)
Heart rate (bps)	73.9 (12.7; 528)	71.0 (8.2; 386)
SBP[b] (mm Hg)	126.0 (15.7; 528)	125.9 (17.9; 386)
DBP[c] (mm Hg)	79.4 (10.6; 528)	77.0 (12.0; 386)
PWV[d] (cm per second)	1591 (266; 153)	1613 (321; 125)

[a]Among a total of 278 pulse wave velocity measurements, 123 measurements were obtained from participants taking antihypertensive medications on the same day.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]PWV: pulse wave velocity.

Table . Estimation results from multivariate regression[a].

Methods	N	MAE[b] (cm per second)	ME[c] (cm per second)	SD (cm per second)	RMSE[d] (cm per second)	Correlation coefficient (*P* value)
Men						
PWV=C1Age+C2SM2+C3 [28[e,f]]	39 rounds	179.1	−49.0	214.3	217.2	0.44 (.006)
	24 participants	160.4	−40.4	195.7	195.8	0.55 (.006)
PWV=C1Age+C2P2O+C3 [2]	39 rounds	189.0	−57.7	219.8	224.6	0.37 (.02)
	24 participants	176.1	−48.1	207.8	209.1	0.43 (.04)
Women						
PWV=C1Age+C2SM2+C3 [28]	34 rounds	165.2	1.8	211.7	208.6	0.66 (<.001)
	26 participants	157.4	−12.1	197.4	194.0	0.72 (<.001)
PWV=C1Age+C2P2O+C3 [2]	34 rounds	196.0	10.0	233.0	229.8	0.62 (<.001)
	26 participants	188.8	8.6	221.8	217.6	0.67 (<.001)

[a]The testing set contained 39 and 34 pulse wave velocity measurements from 24 male participants and 26 female participants, respectively.

[b]MAE: mean absolute error.

[c]ME: mean error.

[d]RMSE: root-mean-square error.

[e]PWV indicates pulse wave velocity.

[f]SI: stiffness index.

[g]P2O: systolic peak to the next onset.

For hierarchical regression, the same training and testing datasets as those in multivariable regression were used to keep participants split. The training dataset was oversampled to make the distribution balanced in each interval of 100 cm per second.

Several parameters, such as the shrinkage factor, tree depth, and column subsampling, are required for the random forest and XGBoost algorithms. Hence, a validation set split from the training dataset was used for parameter settings. Because the number of PWV measurements of extreme high and low values was not sufficiently large, leave-one-out validation was used to ensure that the model for validation is similar to that for training. For the general model, the male validation set contained 23 participants and 33 PWV measurements, while the female validation set had 22 participants and 39 PWV measurements. The validation set consisted of more than one-third of participants in the training dataset and kept uniformly distributed in the range from 1000 cm per second to 2100 cm per second. During leave-one-out validation, all the PPG or ECG sequences associated with the PWV measurements of 1 validation participant were removed from the training dataset to avoid data leak. For each submodel of the local regressor, the validation dataset in each subdivision includes those with the PWV measurements in the overlapping zone. Given the overlapping zone of 400 cm per second, there were 24 PWV measurements from 13 male participants and 26 PWV measurements from 12 female participants in the high submodel for validation from 1400 cm per second. On the other hand, 25 PWV measurements from 13 male participants and 25 PWV measurements from 16 female participants were used in the low submodel for validation up to 1800 cm per second.

Table 3 lists the estimation results from the general and hierarchical regression models by the random forest classification and XGBoost regression algorithms with different settings of the width of the overlapping zones. First, the RMSE results from the hierarchical regression models are better than those from the multivariable linear regression model. The hierarchical regression model also outperforms the general regression model. Figures 5 and 6 show the Bland-Altman and scatter plots of regression results by the hierarchical regression model with overlapping zone of 400 cm per second for men and women participants. Their participant numbers are indicated in the legend. Good estimation was obtained for this setting. The left subfigures indicate the Bland-Altman plot. The scatter plots in the right subfigures provide the final estimation results. The classification accuracies of total rounds from male participants and female participants are 76.9% and 91.2%, respectively. The estimation of erroneously classified data close to the boundary got improved with the introduction of an overlapping zone. The best estimation results achieve RMSE of 145.0 cm per second and 141.4 cm per second for men and women, respectively. In the random forest classifier for male participants, the number of estimators is 100 and the maximum tree depth is 20. As to the random forest for female participants, the number of estimators is 250 and the maximum tree depth is 9. In both cases, the minimum samples for tree split should be larger than 2 and the minimum number of samples in leaf nodes is 1. As to the XGBoost regressors, the number of estimators is 200; the fraction of features sampled for each tree is 0.7; and the minimum loss reduction for further partition is 0. The maximum depth of the low submodel for male participants is 5 and is set to 3 for the remaining submodels.

The XGBoost algorithm performs tree splitting by evaluating structure scores to accumulate gradient statistics according to the sorted feature values while the random forest algorithm can assess the impact on pureness of the leaves from a feature. Hence, both can report the feature importance. Given the overlapping zone of 400 cm per second in the hierarchical regression model, besides PAT (nR), PAT square (nR2), and age, PPG features and WPD features were also frequently used (Multimedia Appendix 6). Local regression models used features different from those used in global classification models. Features from component wave, points a, b, c, and d of SDPPG were often adopted.

Table . Hierarchical regression results for men and for women are listed.

Method	Overlapping zone (cm per second)	N	MAE[a] (cm per second)	ME[b] (cm per second)	SD (cm per second)	RMSE[c] (cm per second)	Correlation coefficient (*P* value)
Men							
General regression	—[d]	39 rounds	157.4	−16.5	187.0	185.3	0.61 (<.001)
General regression	—	24 participants	141.7	−8.4	173.1	169.7	0.66 (<.001)
Hierarchical regression	200	39 rounds	156.0	−19.4	185.3	183.9	0.64 (<.001)
Hierarchical regression	200	24 participants	152.1	−18.4	185.6	182.6	0.63 (.001)
Hierarchical regression	400	39 rounds	133.6	−8.1	160.1	158.3	0.74 (<.001)
Hierarchical regression	400	24 participants	126.5	−8.9	147.8	*145.0* [e]	*0.77*[e] (<.001)
Hierarchical regression	600	39 rounds	153.6	−2.3	182.9	180.5	0.63 (<.001)
Hierarchical regression	600	24 participants	143.6	13.7	165.0	162.1	0.70 (<.001)
Women							
General regression	—	34 rounds	174.3	−36.0	217.0	216.8	0.67 (<.001)
General regression	—	26 participants	177.7	−22.4	217.8	214.7	0.66 (<.001)
Hierarchical regression	200	34 rounds	141.5	−20.7	171.0	169.7	0.80 (<.001)
Hierarchical regression	200	26 participants	131.4	−29.2	157.4	157.0	0.83 (<.001)
Hierarchical regression	400	34 rounds	127.3	−3.5	156.7	154.5	0.83 (<.001)
Hierarchical regression	400	26 participants	116.7	−6.0	144.1	*141.4* [e]	*0.86*[e] (<.001)
Hierarchical regression	600	34 rounds	144.3	24.2	173.9	173.0	0.79 (<.001)
Hierarchical regression	600	26 participants	141.2	24.0	173.5	171.8	0.79 (<.001)

[a]MAE:mean absolute error.

[b]ME: mean error.

[c]RMSE: root-mean-square error.

[d]Not applicable.

[e]Values in italics indicate best estimation result with acceptable accuracy set by the ARTERY Society.

Figure 5. (A) Bland-Altman plot and (B) scatter plot of pulse wave velocity regression by the hierarchical regression model with 2 submodels and overlapping zone of 400 cm per second for 24 men. PWV: pulse wave velocity.

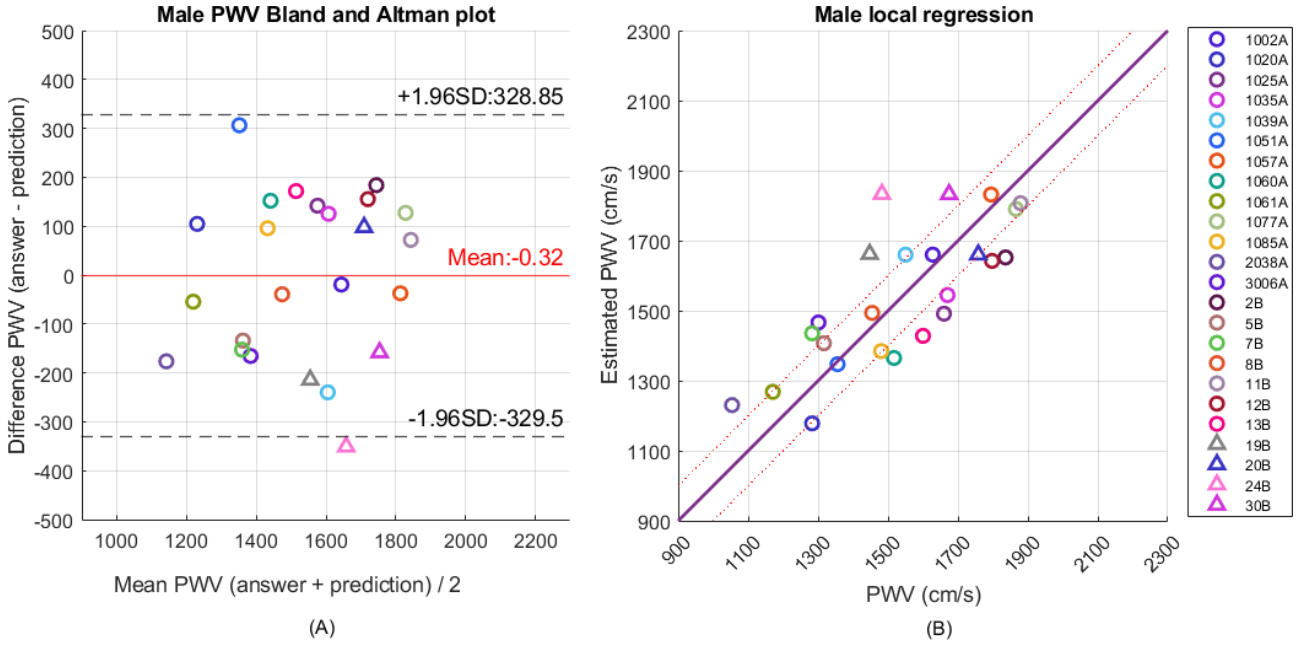
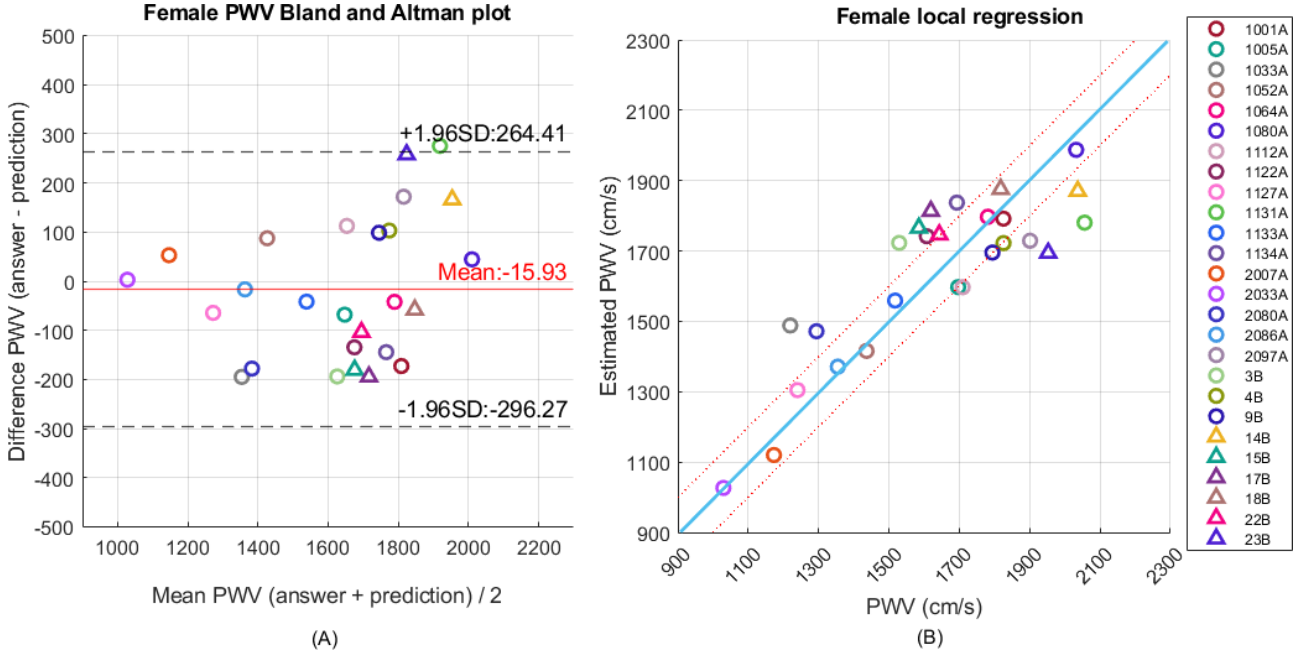


Figure 6. (A) Bland-Altman plot and (B) scatter plot of pulse wave velocity regression by the hierarchical regression model with 2 submodels and overlapping zone of 400 cm per second for 26 women. PWV: pulse wave velocity.



Discussion

Principal Findings

In this study, we used wrist PPG and ECG signals to estimate baPWV. The morphology of wrist PPG signals is quite different from that of finger PPG signals. The conventional approach that used finger PPG morphology features may encounter the problem of feature missing due to much fewer identifiable features of wrist PPG signals. In addition, the multivariable regression model used in prior works may be too simple to describe the complicated hemodynamic state in the vessels. Hence, we resorted to the machine learning algorithm to deal with the estimation. Although the wrist PPG and ECG signals were acquired before the baPWV measurement, they are still related to the vessel condition and stiffness. To further improve and refine the estimation results, hierarchical regression was adopted to shrink the range handled in the submodel. The achieved RMSE and SD by our hierarchical regression models for both men and women are lower than the threshold (150 cm per second) of acceptable accuracy for PWV estimation set by the ARTERY Society [32].

Comparison With Prior Work

With the WPD and feature imputation techniques developed by us, more than 98% of all ambiguous and missing features of wrist PPG can be identified [19]. From the correlation results (Multimedia Appendix 4), besides age (feature 23) and age

square (feature 63), correlation related to SDPPG amplitude of point c (feature 18), point d (feature 19), and point e (feature 20) are still obvious as what has been mentioned in finger PPG [25]. In addition, SI (npd-nps; feature 51), which are often missing in the original wrist PPG pulses, can be computed through the synthesized systolic and diastolic waves in decomposed wrist PPG. According to the feature importance (Multimedia Appendix 6), it still plays an important role for PWV estimation.

The multivariable regression uses only a few features. If significantly high correlations of those features to baPWV do not appear, the performance of estimation will be degraded. However, the machine learning algorithm can help exploit more linear or nonlinear information embedded in the PPG waveform or its component waves and thus is suitable for these applications. Furthermore, the combined features from PPG and ECG morphology, WPD, and basic information supplied more feature information sources that can be selected by the model.

Hierarchical Model Insights

The concept of hierarchical regression is to introduce different models to refine the estimation results. However, the global classifier or regressor must provide sufficiently correct classification to avoid model mismatch. From the hierarchical regression results, it is clear that the inclusion of overlapping zone in local regressors indeed improved the estimation results, as reflected in the improved correlation coefficients (Table 2). However, the determination of optimal range of overlapping zone is still controversial. If the overlapping zone is too wide, the hierarchical regression model would become similar to the general regression model. On the other hand, if the overlapping zone is too narrow, the misclassified data cannot be properly handled. In this study, we recommend the overlapping zone of 400 cm per second of 2 subdivision models because the misclassified data are near the boundary due to good capability of the global classifier and can be appropriately covered by the submodel. We conducted further analysis on the features that were misclassified for those samples not near the decision boundary. The results showed no significant outliers. Additionally, the vote counts for 2 classes across the entire forest were close, indicating low confidence among the trees. The latent properties beyond the observed features should be further studied. On the other hand, we also applied a Kernel Density Estimation–based mutual information analysis [33] to assess the relevance of individual features in male and female datasets. The mutual information values from male features were lower than those from female features, which can also explain the lower classification accuracy in male participants of our dataset.

Limitations and Future Directions

This study has limitations, which point to the directions for future research. First, the sample size remained small and more older adult people were recruited in the study, which might limit its applicability in younger populations. While the current dataset demonstrates feasibility in estimating PWV using wrist PPG in older individuals, the skewed dataset toward older individuals may have influenced the performance due to age-related vascular characteristics. In future work, we plan to expand the study population by actively recruiting more young participants. The inclusion of younger participants will help balance the age distribution and allow for more robust assessment of the model performance across different age groups. This extension will not only improve the generalizability of the model but also enable a more comprehensive evaluation of age-related vascular changes. Second, the current model adopts machine learning algorithms to exploit linear and nonlinear features within the scope of this dataset. As the dataset grows in size and diversity, other deep learning algorithms, such as Bayesian neural networks or multilayer perceptrons, can be applied, which may offer better uncertainty quantification or modeling capabilities. Third, the feature space used in the current model is relatively high-dimensional, which may hinder its practical deployment on wearable or edge devices with limited computational resources. Feature compression or dimensionality reduction techniques can be considered to decrease model complexity in the future. This optimization will help make the system more suitable for real-time, low-power applications in wearable health care settings. Together, these improvements aim to enhance both the robustness and the applicability of the proposed approach, facilitating its transition toward practical use in diverse and real-world scenarios.

Acknowledgments

The authors would like to acknowledge Mr Bowen Ku of Mediatek Inc for his design experience feedback about wearable devices in biomedical applications. This work is supported by Mediatek Inc (grant numbers MTKC-2021 - 0477 and MTKC-2023 - 1363).

Authors' Contributions

P-YT, C-HH, and Y-CL contributed to conceptualization; P-YT contributed to methodology; C-IH, C-HY, C-HH, Y-CL, and J-WG contributed to software; C-IH, C-HY, and Y-CL participated in validation; T-DW and H-JL contributed to resources; P-YT participated in writing—original draft preparation; T-DW participated in writing—review and editing; and T-DW, P-YT, and H-JL participated in supervision. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

PYT has received research grants (grant numbers MTKC-2021 - 0477 and MTKC-2023 - 1363) from Mediatek Company. All other authors have no relevant relationships to disclose.

Multimedia Appendix 1
(A) Electrocardiography before preprocessing, (B) photoplethysmography before preprocessing, (C) electrocardiography with R peak after preprocessing, and (D) photoplethysmography with valley after preprocessing.
[PNG File, 97 KB - biomedeng_v10i1e58756_app1.png]

Multimedia Appendix 2
List of extracted features: (A) photoplethysmography features, (B) weighted pulse decomposition features, (C) electrocardiography features, and (D) basic information.
[DOCX File, 23 KB - biomedeng_v10i1e58756_app2.docx]

Multimedia Appendix 3
Concept of hierarchical regression.
[PNG File, 251 KB - biomedeng_v10i1e58756_app3.png]

Multimedia Appendix 4
Heat map of correlation coefficients of combined features (defined in Tables S1 A, S1 B, S1 C, and S1 D in Multimedia Appendix 2) versus brachial-ankle pulse wave velocity for (A) 80 male participants and (B) 82 female participants with 528 and 386 data, respectively. The diagonal elements are the correlation coefficients of original features. The off-diagonal elements are the correlation coefficients of combined features.
[PNG File, 71 KB - biomedeng_v10i1e58756_app4.png]

Multimedia Appendix 5
The distributions of brachial-ankle pulse wave velocity versus (A) age, (B) n_{ms}, and (C) Age/n_{ms} for 386 female data.
[PNG File, 79 KB - biomedeng_v10i1e58756_app5.png]

Multimedia Appendix 6
Top 10 important features of classification and local regression for (A) men and (B) women.
[DOCX File, 24 KB - biomedeng_v10i1e58756_app6.docx]

References

1. Pereira T, Correia C, Cardoso J. Novel methods for pulse wave velocity measurement. J Med Biol Eng 2015;35(5):555-565. [doi: 10.1007/s40846-015-0086-8] [Medline: 26500469]
2. Ohkuma T, Ninomiya T, Tomiyama H, et al. Brachial-ankle pulse wave velocity and the risk prediction of cardiovascular disease. Hypertension 2017 Jun;69(6):1045-1052. [doi: 10.1161/HYPERTENSIONAHA.117.09097]
3. Kachuee M, Kiani MM, Mohammadzade H, Shabany M. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. IEEE Trans Biomed Eng 2017 Apr;64(4):859-869. [doi: 10.1109/TBME.2016.2580904] [Medline: 27323356]
4. Yan C, Li Z, Zhao W, et al. Novel deep convolutional neural network for cuff-less blood pressure measurement using ECG and PPG signals. Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 23-27, 2019; Berlin, Germany. [doi: 10.1109/EMBC.2019.8857108]
5. Wu Q, Yang J, Zheng G, et al. An ambulatory blood pressure monitoring system based on the uncalibrated steps of the wrist. Presented at: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); Oct 19-21, 2019; Suzhou, China p. 1-6. [doi: 10.1109/CISP-BMEI48845.2019.8966011]
6. Yousefian P, Shin S, Mousavi AS, et al. Pulse transit time-pulse wave analysis fusion based on wearable wrist ballistocardiogram for cuff-less blood pressure trend tracking. IEEE Access 2020 Jul;8:138077-138087. [doi: 10.1109/ACCESS.2020.3012384]
7. Krivoshei A, Min M, Uuetoa H, Lamp J, Annus P. Electrical bio-impedance based non-invasive method for the central aortic blood pressure waveform estimation. Presented at: 2014 14th Biennial Baltic Electronic Conference (BEC); Oct 6-8, 2014; Tallinn, Estonia. [doi: 10.1109/BEC.2014.7320586]
8. Meidert AS, Saugel B. Techniques for non-invasive monitoring of arterial blood pressure. Front Med (Lausanne) 2018;4:231. [doi: 10.3389/fmed.2017.00231] [Medline: 29359130]
9. Priyanka KNG, Chao PCP, Tu TY, et al. Estimating blood pressure via artificial neural networks based on measured photoplethysmography waveforms. Presented at: 2018 IEEE Sensors; Oct 28-31, 2018; New Delhi p. 1-4. [doi: 10.1109/ICSENS.2018.8589796]
10. Schlesinger O, Vigderhouse N, Eytan D, Moshe Y. Blood pressure estimation from PPG signals using convolutional neural networks and Siamese network. Presented at: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 4-8, 2020; Barcelona, Spain. [doi: 10.1109/ICASSP40776.2020.9053446]

11. Warren S. High resolution wireless body area network with statistically synchronized sensor data for tracking pulse wave velocity. 2012 Presented at: 2012 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 28 to Sep 1, 2012; San Diego, CA p. 2080-2083. [doi: 10.1109/EMBC.2012.6346369]
12. Jang DG, Park SH, Hahn M. Enhancing the pulse contour analysis-based arterial stiffness estimation using a novel photoplethysmographic parameter. IEEE J Biomed Health Inform 2015 Jan;19(1):256-262. [doi: 10.1109/JBHI.2014.2306679] [Medline: 25561448]
13. Padilla JM, Berjano EJ, Facila SJ, Diaz P L, Merce S. Assessment of relationships between blood pressure, pulse wave velocity and digital volume pulse. Presented at: 2006 Computers in Cardiology; Sep 17-20, 2006; New Delhi, India p. 893-896.
14. Nabeel PM, Jayaraj J, Mohanasankar S. Single-source PPG-based local pulse wave velocity measurement: a potential cuffless blood pressure estimation technique. Physiol Meas 2017 Nov 30;38(12):2122-2140. [doi: 10.1088/1361-6579/aa9550] [Medline: 29058686]
15. Salvi P, Magnani E, Valbusa F, et al. Comparative study of methodologies for pulse wave velocity estimation. J Hum Hypertens 2008 Oct;22(10):669-677. [doi: 10.1038/jhh.2008.42] [Medline: 18528411]
16. Alty SR, Angarita-Jaimes N, Millasseau SC, Chowienczyk PJ. Predicting arterial stiffness from the digital volume pulse waveform. IEEE Trans Biomed Eng 2007 Dec;54(12):2268-2275. [doi: 10.1109/tbme.2007.897805] [Medline: 18075043]
17. Hartmann V, Liu H, Chen F, Qiu Q, Hughes S, Zheng D. Quantitative comparison of photoplethysmographic waveform characteristics: effect of measurement site. Front Physiol 2019;10:198. [doi: 10.3389/fphys.2019.00198] [Medline: 30890959]
18. Rajala S, Lindholm H, Taipalus T. Comparison of photoplethysmogram measured from wrist and finger and the effect of measurement location on pulse arrival time. Physiol Meas 2018 Aug 1;39(7):075010. [doi: 10.1088/1361-6579/aac7ac] [Medline: 29794339]
19. Tsai PY, Huang CH, Guo JW, et al. Coherence between decomposed components of wrist and finger PPG signals by imputing missing features and resolving ambiguous features. Sensors (Basel) 2021 Jun 24;21(13):4315. [doi: 10.3390/s21134315] [Medline: 34202597]
20. Warren S, Li K. Initial study on pulse wave velocity acquired from one hand using two synchronized wireless reflectance pulse oximeters. Annu Int Conf IEEE Eng Med Biol Soc 2011:6907-6910. [doi: 10.1109/IEMBS.2011.6091739]
21. Hashimoto J, Watabe D, Kimura A, et al. Determinants of the second derivative of the finger photoplethysmogram and brachial-ankle pulse-wave velocity: the Ohasama study. Am J Hypertens 2005 Apr;18(4 Pt 1):477-485. [doi: 10.1016/j.amjhyper.2004.11.009] [Medline: 15831356]
22. Davies JE, Baksi J, Francis DP, et al. The arterial reservoir pressure increases with aging and is the major determinant of the aortic augmentation index. Am J Physiol Heart Circ Physiol 2010 Feb;298(2):H580-H586. [doi: 10.1152/ajpheart.00875.2009] [Medline: 20008272]
23. Couceiro R, Carvalho P, Paiva RP, et al. Assessment of cardiovascular function from multi-Gaussian fitting of a finger photoplethysmogram. Physiol Meas 2015 Sep;36(9):1801-1825. [doi: 10.1088/0967-3334/36/9/1801] [Medline: 26235798]
24. Huang CH, Guo JW, Yang YC, et al. Weighted pulse decomposition analysis of fingertip photoplethysmogram signals for blood pressure assessment. Presented at: 2020 IEEE International Symposium on Circuits and Systems (ISCAS); 2020; Seville, Spain. [doi: 10.1109/ISCAS45731.2020.9180616]
25. Takazawa K, Tanaka N, Fujita M, et al. Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform. Hypertension 1998 Aug;32(2):365-370. [doi: 10.1161/01.hyp.32.2.365] [Medline: 9719069]
26. Ding X, Zhang YT. Pulse transit time technique for cuffless unobtrusive blood pressure measurement: from theory to algorithm. Biomed Eng Lett 2019 Feb;9(1):37-52. [doi: 10.1007/s13534-019-00096-x] [Medline: 30956879]
27. Tomiyama H, Yamashina A, Arai T, et al. Influences of age and gender on results of noninvasive brachial-ankle pulse wave velocity measurement--a survey of 12517 subjects. Atherosclerosis 2003 Feb;166(2):303-309. [doi: 10.1016/s0021-9150(02)00332-5] [Medline: 12535743]
28. Jang DG, Farooq U, Park SH, Goh CW, Hahn M. A knowledge-based approach to arterial stiffness estimation using the digital volume pulse. IEEE Trans Biomed Circuits Syst 2012 Aug;6(4):366-374. [doi: 10.1109/TBCAS.2011.2177835] [Medline: 23853181]
29. Miao F, Wang X, Yin L, Li Y. A wearable sensor for arterial stiffness monitoring based on machine learning algorithms. IEEE Sensors J 2019 Feb;19(4):1426-1434. [doi: 10.1109/JSEN.2018.2880434]
30. Zhang B, Ren J, Cheng Y, Wang B, Wei Z. Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm. IEEE Access 2019;7:32423-32433. [doi: 10.1109/ACCESS.2019.2902217]
31. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: Proceedings of the 22nd ACM International Conference on Knowledge Discovery and Data Mining (KDD); Aug 13-17, 2016; San Francisco, CA p. 785-794.
32. Wilkinson IB, McEniery CM, Schillaci G, et al. ARTERY Society guidelines for validation of non-invasive haemodynamic measurement devices: part 1, arterial pulse wave velocity. ARTRES 2010;4(2):34. [doi: 10.1016/j.artres.2010.03.001]
33. Kwak N. Input feature selection by mutual information based on Parzen window. IEEE Trans Pattern Anal Machine Intell 2002 Dec;24(12):1667-1671. [doi: 10.1109/TPAMI.2002.1114861]

Abbreviations

baPWC: brachial-ankle pulse wave velocity
cfPWV: carotid-femoral pulse wave velocity
CVD: cardiovascular disease
ECG: electrocardiography
FDPPG: first-order derivative photoplethysmography
PAT: pulse arrival time
PPG: photoplethysmography
P2O: peak to the next onset
PWV: pulse wave velocity
RMSE: root-mean-square error
SDPPG: second-order derivative photoplethysmography
SI: stiffness index
WPD: weighted pulse decomposition
XGBoost: extreme gradient boosting

Edited by T Leung; submitted 06.06.24; peer-reviewed by R Guo, X Xing; revised version received 05.06.25; accepted 06.06.25; published 26.08.25.

Please cite as:
Ho CI, Yen CH, Li YC, Huang CH, Guo JW, Tsai PY, Lin HJ, Wang TD
Estimation of Brachial-Ankle Pulse Wave Velocity With Hierarchical Regression Model From Wrist Photoplethysmography and Electrocardiographic Signals: Method Design
JMIR Biomed Eng 2025;10:e58756
URL: https://biomedeng.jmir.org/2025/1/e58756
doi:10.2196/58756

Challenges and Solutions in Applying Large Language Models to Guideline-Based Management Planning and Automated Medical Coding in Health Care: Algorithm Development and Validation

Peter Sarvari[1], MBA, MEng, MS; Zaid Al-fagih[1], BSc, MBBS, MPP; Alexander Abou-Chedid[1], BS; Paul Jewell[2], BSc, MBBS, MBA; Rosie Taylor[2], BA, MBBS, MS; Arouba Imtiaz[3], MBChB

[1]Rhazes AI, 85 Great Portland Street, London, United Kingdom

[2]Assuric, London, United Kingdom

[3]Frimley Health NHS Foundation Trust, Camberley, United Kingdom

Corresponding Author:
Peter Sarvari, MBA, MEng, MS
Rhazes AI, 85 Great Portland Street, London, United Kingdom

Abstract

Background: Diagnostic errors and administrative burdens, including medical coding, remain major challenges in health care. Large language models (LLMs) have the potential to alleviate these problems, but their adoption has been limited by concerns regarding reliability, transparency, and clinical safety.

Objective: This study introduces and evaluates 2 LLM-based frameworks, implemented within the Rhazes Clinician platform, designed to address these challenges: generation-assisted retrieval-augmented generation (GARAG) for automated evidence-based treatment planning and generation-assisted vector search (GAVS) for automated medical coding.

Methods: GARAG was evaluated on 21 clinical test cases created by medically qualified authors. Each case was executed 3 times independently, and outputs were assessed using 4 criteria: correctness of references, absence of duplication, adherence to formatting, and clinical appropriateness of the generated management plan. GAVS was evaluated on 958 randomly selected admissions from the Medical Information Mart for Intensive Care (MIMIC)–IV database, in which billed *International Classification of Diseases, Tenth Revision* (*ICD-10*) codes served as the ground truth. Two approaches were compared: a direct GPT-4.1 baseline prompted to predict *ICD-10* codes without constraints and GAVS, in which GPT-4.1 generated diagnostic entities that were each mapped onto the top 10 matching *ICD-10* codes through vector search.

Results: Across the 63 outputs, 62 (98.4%) satisfied all evaluation criteria, with the only exception being a minor ordering inconsistency in one repetition of case 14. For GAVS, the 958 admissions contained 8576 assigned *ICD-10* subcategory codes (1610 unique). The vanilla LLM produced 131,329 candidate codes, whereas GAVS produced 136,920. At the subcategory level, the vanilla LLM achieved 17.95% average recall (15.86% weighted), while GAVS achieved 20.63% (18.62% weighted), a statistically significant improvement (P<.001). At the category level, performance converged (32.60% vs 32.58% average weighted recall; P=.99).

Conclusions: GARAG demonstrated a workflow that grounds management plans in diagnosis-specific, peer-reviewed guideline evidence, preserving fine-grained clinical detail during retrieval. GAVS significantly improved fine-grained diagnostic coding recall compared with a direct LLM baseline. Together, these frameworks illustrate how LLM-based methods can enhance clinical decision support and medical coding. Both were subsequently integrated into Rhazes Clinician, a clinician-facing web application that orchestrates LLM agents to call specialized tools, providing a single interface for physician use. Further independent validation and large-scale studies are required to confirm generalizability and assess their impact on patient outcomes.

(JMIR Biomed Eng 2025;10:e66691) doi:10.2196/66691

KEYWORDS

AI assistant; large language model; LLM; GPT-4; retrieval-augmented generation; RAG; generation-assisted retrieval-augmented generation; GARAG; generation-assisted vector search; GAVS; medical informatics; digital health; generative AI in medicine; medical web application; automated medical coding; AI diagnosis; artificial intelligence

Introduction

Health Care's Diverse Challenges

Health care is facing profound challenges that urgently require innovative solutions. Medical errors [1], overwhelming administrative burdens [2], understaffing [3,4], spiraling costs [5], and clinician burnout [6,7] threaten the foundations of medical care. Despite this, the health care sector has lagged behind other industries in adopting new technologies.

To address the above challenges, a subset of the authors has developed an AI-powered web app called Rhazes that helps doctors with paperwork and analytical tasks along their clinical workflow. Rhazes, like many other digital health startups, aims to offer integrated tools to health care professionals to match many of the growing needs in health care worsened by a staffing crisis [8-10]. AI-powered tools have been shown to have the potential to automate rote tasks [11], reduce errors [12,13], cut costs for everyone [14], improve clinician well-being and their patient-centeredness [15], and ultimately deliver better patient outcomes [12,16]. However, significant barriers and challenges persist in ensuring the safe and effective integration of AI within health care systems globally [17-19].

Diagnostic Errors

Diagnostic errors in health care are more common than generally recognized and often receive less attention in both clinical practice and research. There are 2 ways to estimate their prevalence: autopsies and expert opinions. Various studies suggest a range between 5% and 20% [20-22], with the actual figure likely falling somewhere in between. Considering the 1.3 billion health care visits annually in the United States as an example, this percentage translates to a staggering 65 to 260 million diagnostic errors each year in the United States alone [23]. The rate of false negatives varies widely, from as low as 2.2% for myocardial infarction to an alarming 62.1% for spinal abscesses [20]. The National Academy of Medicine underscores the severity of this issue, attributing diagnostic errors to approximately 10% of patient deaths and 6% to 17% of hospital complications, with vascular events, infections, and cancers constituting around 75% of the serious harms from these errors [24].

Diagnostic errors arise from several causes. No-fault errors, such as those due to atypical disease presentations, are difficult to prevent. System-related issues, including delays in testing and communication failures, also play a role, though comprehensive data are limited. Cognitive errors, however, are the most frequent, with Graber et al finding that cognitive factors—such as faulty perception, failed heuristics, and cognitive biases—contribute to 74% of diagnostic errors [25]. This suggests that a substantial proportion of these errors could be mitigated through targeted interventions aimed at clinical decision-making and judgment. In malpractice claims, nearly 90% involve failures in these domains, regardless of the underlying condition [26].

All Roads Lead to Burnout

Clinicians are responsible not only for making diagnoses but also for managing patient care thereafter. This process involves 3 major challenges. The first is staying up to date with a rapidly expanding medical knowledge base, which is estimated to double every 73 days [27] alongside frequently updated clinical guidelines. To address this, many physicians rely on peer-reviewed online clinical guideline databases, such as subscription-based services including UpToDate [28], DynaMed [29], as well as freely available resources like StatPearls [30]. The second challenge is identifying the most relevant guideline for a specific clinical scenario. The third, and often most complex, is tailoring these guidelines to the unique needs of each patient, taking into account individual characteristics, medical history, preferences, and socioeconomic context.

Medical errors often occur within the broader context of systemic pressures. One major factor is the administrative workload placed on clinicians. Studies indicate that for every hour spent in direct patient care, physicians spend approximately 2 additional hours on documentation and other administrative tasks [31]. This environment can contribute to situations where clinical duties become secondary to administrative responsibilities.

In the United States, more than half of physicians report at least 1 symptom of burnout [32], representing an estimated annual economic cost of US $4.6 billion or US $7600 per physician [33]. In the United Kingdom, physician burnout rates reached a record high in 2021 according to the annual national training survey [34].

Medical Coding

Medical coding has clinical, statistical, and billing-related usages. Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT; maintained by the International Health Terminology Standards Development Organization) is a terminology that provides clinicians with precise patient-specific information, including symptoms, diagnoses, procedures, and social contexts [35]. In the National Health Service (NHS), UK, SNOMED-CT is used for clinical coding, specifically to safely and accurately exchange information between health care providers. It is recorded at the point-of-care level and integrated into electronic health records (EHRs) as required by Fast Health Care Interoperability Resources [35], a health care data sharing standard. Most general practitioner clinics employ medical coders to translate patient findings into a mix of SNOMED-CT and in-house diagnostic codes for the most common cases. SNOMED-CT allows for more precise coding, as it not only comprises over 340,000 clinical [36] and 1.4 million drug-related codes but also describes the relationship between these codes, essentially functioning as an ontology [35]. In the United Kingdom, diagnostic codes using the *International Classification of Diseases, Eleventh Revision* (*ICD-11*) standard (published by the World Health Organization) and procedural codes using the Operating Procedure Codes Supplement (OPCS; published by NHS England) standard are recorded after the clinical event for statistical purposes, whereas in the United States, *ICD-11* is used for mainly billing purposes [37]. For coding procedures in the United States, the Health Care Common Procedure Coding System (HCPCS; published by the Centers for Medicare and Medicaid Services) is used. It has 2 levels: level 1 comprises Current Procedural Terminology (CPT)

codes (published by the American Medical Association), which is used to bill for procedures done by health care professionals, and level 2 can be used to bill for products, supplies, and services used outside the physician's office such as ambulatory services or orthotics [38]. Automated medical coding is needed for 2 main reasons: one is accuracy and the other is efficiency: the average coding accuracy is around 80% [37], with 83% in the United Kingdom and 89% in Scotland [39], and just the coding of backlog cases can take anywhere from several months to over a year [40].

Large Language Models Could Help

Given the recent progress in artificial intelligence (AI), it has been proposed to help with various aspects of clinical work, including scribing and diagnosis [13,41,42]. GPT-4, a large language model (LLM) developed by OpenAI, has shown promise in medical applications with its passing of the medical board exam in multiple countries and languages [43-45]. A peer-reviewed study assessing the diagnostic ability of GPT-4 and Pathways Language Model 2 on 1000 electronic patient records reported that GPT-4 achieved a 93.9% diagnostic hit rate (lower bound), validated by 3 medical doctors [46]. Furthermore, the authors found that a quick and accurate automated diagnostic evaluation may be possible by presenting the ground truth data to GPT-4 and asking it to assess the diagnostic predictions made by LLMs [46,47]. This can then be used to rapidly benchmark different models and prompting strategies. A report published by OpenAI and Penda Health [48] claimed that AI Consult, a tool powered by LLMs, reduced diagnostic errors by 16% and treatment errors by 13% for 39,849 patient visits in Kenya.

When it comes to management planning, LLMs can revolutionize medical search and find recommendations for a specific clinical scenario by automatically citing the relevant guidelines. Examples include recently developed online medical search tools such as Elsevier Clinical Key [49] and MedWise [50]. The next level of automation is AI analyzing and extracting the relevant details from the EHR to adapt the clinical guidelines to the unique needs of the patient, essentially crafting a personalized treatment plan. An application capable of doing this is called a clinical decision support (CDS) [51] system, and industry examples of such tools include Glass Health [52] and Rhazes [53].

Another important application of LLMs in clinics is notetaking. Automated documentation leveraging ambient listening has shown promise in reducing clinician burden and improving the experience of doctor–patient interactions for both parties [54,55]. In addition, clinical evaluation of existing scribing tools such as Tortus [56], DeepScribe [57], Nuance Dragon Ambient Experience (DAX) [58,59], and Rhazes [60] has indicated enhanced documentation quality [56], increase in billed diagnostic codes, and potential time and cost savings [57,58,60]. However, such tools can cost US $1850 per month per clinician [58] and cause a worsening of after-hours electronic health records (EHR) usage [59]. In fact, Haberle et al found that Dragon Ambient Experience did not benefit documentation, productivity, or even patient experience but helped with provider engagement [59]. Ma et al argue that ambient AI scribes can even reduce time spent on the EHR, but further studies are needed to identify the users benefiting most from such technology [61].

Even though computer-assisted medical coding has been shown to improve coding accuracy [62], automating the clinical coding system appeared out of reach prior to the generative AI revolution due to technological and implementation-level challenges [63]. Non-LLM–based encoder–decoder-type models were shown to really struggle with identifying less frequent codes [64]; however, retrieval-augmented generation (RAG)–enhanced LLMs were recently found to be preferable to provider coders in terms of coding accuracy [65]. Generative AI seems to have made a big contribution toward the full automation of medical coding, and while we found no peer-reviewed evaluation paper to date, the authors of the previously cited paper, affiliated with Corti AI [66], who were researching non-LLM-based methods [64], are now leveraging generative AI to automate medical coding [66].

Methods

Ethical Considerations

The Medical Information Mart for Intensive Care (MIMIC)–IV [67] is a publicly available database and was previously ethically approved by the institutional review boards at Beth Israel Deaconess Medical Center (2001P001699) and the Massachusetts Institute of Technology (0403000206) in accordance with the tenets of the Declaration of Helsinki. The waiver of the requirement for informed consent was included in the institutional review board approval, as all protected health information was deidentified [67]. One of the authors (PS) was granted access to the database after completing training in human research (CITI Human Research certification number: 54889098) and signing a data use agreement in PhysioNet (agreement number 64081). The experiments described in this paper were conducted on Microsoft Azure (Azure OpenAI service) according to the "Responsible use of MIMIC data with online services like GPT" guidance by PhysioNet [68]. The code associated with this publication has been shared in an open repository, and information is provided in the "Data Availability" section of this manuscript.

Generation-Assisted Retrieval-Augmented Generation for Clinical Decision Support

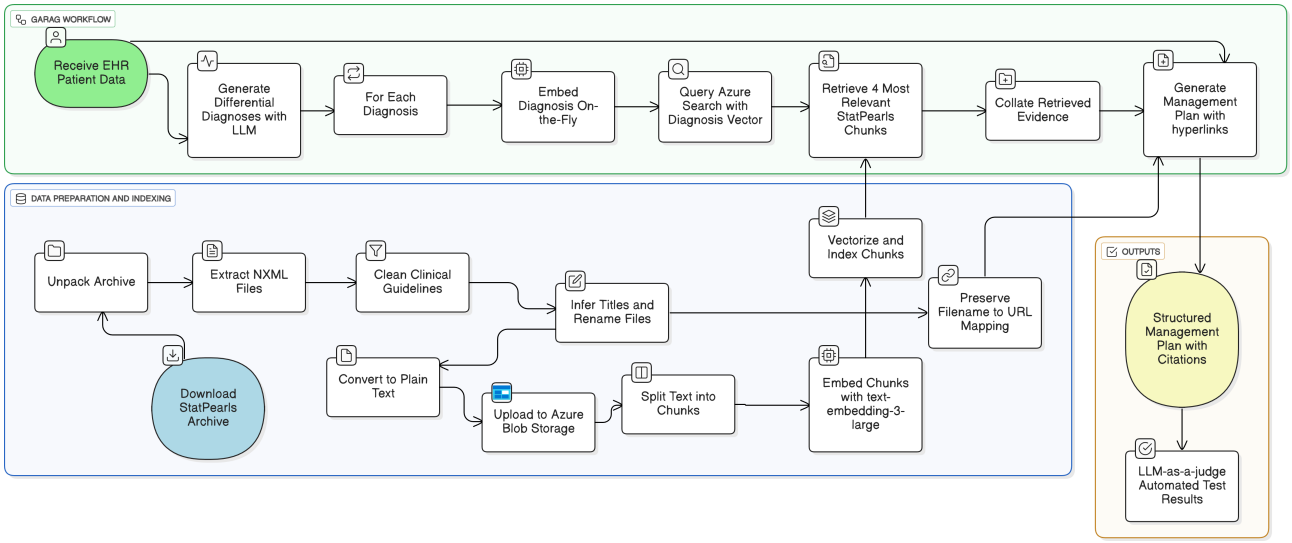
Here we demonstrate how to build a prototype for an AI-driven CDS, in particular, for crafting patient-specific management plans with verifiable citations from StatPearls [30], a point-of-care medical database with peer-reviewed clinical guidelines. On March 2, 2025, a comprehensive archive of clinical guidelines from StatPearls (approximately 1.5 gigabytes in size) was downloaded for use as the reference corpus in the RAG process. After unpacking the archive, a total of 9559 nxml files were obtained, each corresponding to the management of a distinct medical condition. These files were subsequently cleaned to remove nonclinical and extraneous information, including licenses, credits, warranties, publishing details, user prompts (eg, "Comment on this article"), and reference sections, so that only clinically relevant content remained. The title of

each file was automatically inferred from the text and used as a filename, thereby linking each document to the medical event it described. The cleaned files were then converted into plain text format and uploaded to Azure Blob Storage. For citation purposes, a mapping was preserved between each inferred filename and the original download URL from StatPearls.

To enable semantic search and retrieval, the corpus was indexed within Azure Search Service. The indexing pipeline comprised a data source connection to Azure Blob Storage, a search index with fields for filename, chunk identifier, chunk text, and embedding vector; a text-splitting skill with a maximum chunk size of 4000 tokens and an overlap of 100 tokens; and an embedding skill using text-embedding-3-large model, OpenAI's latest and best embedding model to date [69]. An indexer was then executed to vectorize and index the entire collection of document chunks.

Building upon this foundation, we developed a proof-of-concept workflow, which we termed generation-assisted retrieval-augmented generation (GARAG). GARAG proceeds in 3 stages. First, given EHR data, an LLM (specifically GPT-4.1) is prompted to generate a structured list of differential diagnoses. Second, for each diagnosis, the system queries the indexed StatPearls corpus through Azure Search Service, employing the Hierarchical Navigable Small World approximate nearest neighbor algorithm (with parameters M=4, efConstruction=400, and efSearch=500) and cosine similarity as the distance metric. The 4 most relevant text chunks are retrieved for each candidate diagnosis. Third, the LLM is prompted again with the patient data and the retrieved evidence sources. From this input, the model generates a structured management plan covering investigations, treatment suggestions, supportive management, other considerations, risks, and references. Citations are automatically hyperlinked to the original StatPearls sources via the preserved filename-to-URL mapping (Figure 1). A key advantage of GARAG is its ability to ground recommendations in guideline-specific, peer-reviewed sources tailored to each predicted diagnosis. This targeted retrieval avoids the information dilution that can occur with standard RAG approaches, where embedding the entire case may obscure fine-grained clinical details. By structuring the workflow around diagnosis-specific guideline retrieval, GARAG ensures that management plans are directly aligned with authoritative clinical references.

Figure 1. Generating a personalized evidence-based management plan using StatPearls and the GARAG framework. EHR: electronic health record; GARAG: generation-assisted retrieval-augmented generation; LLM: large language model.



For clarity, we have provided a Python Jupyter Notebook that demonstrates our implementation of such a GARAG system and the automated tests we have run to confirm that the instructions are being followed by the LLM. To contrast the GARAG workflow to a traditional RAG workflow, we also provide the reader with a Python prototype that generates treatment plans with StatPearls references using RAG.

Generation-Assisted Vector Search for Automated Medical Coding

Next, we demonstrate how to build an AI tool for automated medical coding using a method we named generation-assisted vector search (GAVS). Unlike RAG, in which retrieval precedes generation, GAVS inverts the process: generation is performed first and retrieval follows. In this pipeline, an Azure OpenAI LLM is few-shot prompted to read EHR text and to enumerate, with maximal granularity, the clinically relevant entities for coding (eg, diagnoses/comorbidities/abnormalities, procedures/treatments/services, and drugs). Each generated item is subsequently embedded with text-embedding-3-large and matched against a vector database representing the target coding ontology.

For procedural coding, the coding ontology was derived from the official 2025 CPT release downloaded as an Excel workbook from the Centers for Disease Control and Prevention government website. The worksheet containing all CPT entries ("ALL 2025 CPT Codes") was parsed to retain the canonical code identifier and its short description; rows with missing codes were removed. Each row was converted into a LangChain [70] Document whose embedding encodes the description, while the metadata preserves both the CPT Code and Title. Because CPT descriptions are short, no chunking was required.

Embeddings were stored in a Pinecone serverless index [71] and queried via the LangChain [70] PineconeVectorStore at runtime. During inference, each model-generated item (eg,procedure) in a structured list output is embedded on the fly and used to perform semantic search, retrieving the top 10 nearest CPT entries by cosine similarity. The returned results include both semantic scores and the canonical CPT codes via the stored metadata, allowing the system to report human-readable candidates (code+title) alongside each extracted clinical item.

The same pattern generalizes to diagnostic and pharmacological coding by substituting the target ontology (eg, *ICD-11* or SNOMED CT for diagnoses; SNOMED CT for drugs) and constructing an analogous vector store with description-level embeddings and code identifiers preserved in metadata. For transparency and reproducibility, we provide a Jupyter Notebook demonstrating the full CPT workflow end-to-end, including data acquisition, runtime index creation, LLM-based generation, and similarity search-based mapping. Crucially, the index is created and populated programmatically at runtime if absent (index name cpt-cdc-2025-text-embedding-3-large, dimension 3072, metric cosine, region eu-west-1); if present, the pipeline connects to the existing index without reingestion. We also provide links to download coding ontologies from the official publisher websites in the format of single or multiple Excel files: CPT from the Centers for Medicare and Medicaid Services [72], *ICD-11* from the World Health Organization [73], and SNOMED from the NHS Digital website [74].

To quantitatively assess the benefits of GAVS over direct LLM prompting, we conducted a proof-of-concept evaluation for *International Classification of Diseases, Tenth Revision* (*ICD-10*) coding following the methods of Sarvari and Al-Fagih [47]. We initially sampled 1000 admissions at random from MIMIC-IV, of which 42 did not have officially assigned *ICD-10* codes and were excluded, resulting in a final cohort of 958 admissions. For each admission, the set of billed *ICD-10* codes served as the ground truth. Across this cohort, there were 8576 total *ICD-10* codes at the subcategory (full code) level, comprising 1610 unique subcategories. When mapped to parent categories, the total number decreased to 7311 codes across 540 unique categories. The discrepancy (8576 vs 7311) reflects cases in which the Python library used for mapping [75] did not recognize certain overly specific subcategory codes, in which case no parent category was assigned. The full *ICD-10* ontology includes 95,109 valid codes, defining the candidate space for prediction. Two approaches were compared. In the vanilla LLM method, GPT-4.1 was prompted directly to predict *ICD-10* codes for each admission, without external constraints. In the GAVS method, GPT-4.1 was first prompted to generate granular diagnostic entities, which were then embedded and matched to the top 10 most similar *ICD-10* codes through vector search (Azure Search Service) of the official ontology. Both methods used identical LLM configuration and dataset preprocessing, ensuring comparability across experiments. The primary outcome was recall, defined as the fraction of ground-truth codes correctly predicted. Two variants were calculated, consistent with [46] (1) average (per-admission mean) recall, averaged across the 958 admissions, and (2) aggregate (weighted average) recall, defined as the total number of correctly predicted codes divided by the total number of codes (8576) or 1−(∑missed codes across admissions / ∑true codes across admissions).

Precision was not reported, as discussed previously [47], because billing records are not a reliable gold standard for false positive determination: clinically valid diagnoses often go unbilled, and multiple codes may be acceptable matches (especially when working with incomplete data). In this context, precision metrics would therefore be misleading. For statistical comparison, following [47], we applied a 2-proportion *z* test to evaluate differences between methods in recovered versus missed ground-truth codes.

For clarity, we provide the reader with a Python Jupyter notebook, demonstrating the entire automated coding prediction (including the vanilla GPT-4.1 and GAVS methods for predicting *ICD-10* codes) and evaluation workflow both at the subcategory and category levels.

Results

GARAG: Citation Integrity and Relevance

Using the LLM-as-a-judge method [47,76], we evaluated the GARAG workflow on 21 clinical test cases created by a subset of the authors who are medical professionals. Each case was executed 3 times independently to assess reproducibility, yielding a total of 63 runs. Performance was assessed using four criteria: (1) correctness of references, (2) absence of duplicate citations, (3) adherence to citation formatting standards, and (4) contextual appropriateness of the generated management plan, including whether it explicitly addressed the presented diagnoses. Across all 63 runs, 62 satisfied all 4 criteria, corresponding to a success rate of 98.4%. The single exception occurred in case 14 during its first repetition, in which references were accurate but displayed a minor ordering inconsistency, with "[3]" appearing before "[2]." Importantly, no spurious references were observed, all citations could be traced directly to their StatPearls sources, and all management plans were judged clinically relevant. These findings demonstrate that GARAG provides a highly reliable and reproducible workflow for generating clinical management plans with proper citation handling, with only minor formatting issues detected across repeated executions.

GAVS: Diagnostic Coding Performance

We next evaluated the GAVS method for automated diagnostic coding on 958 randomly selected MIMIC-IV hospital admissions. Across these cases, there were 8576 total assigned *ICD-10* codes at the subcategory level, spanning 1610 unique subcategories. When collapsed to categories using a Python mapping library [75], this corresponded to 7311 total codes across 540 unique categories. The full *ICD-10* ontology contains 95,109 valid codes, underscoring the scale of the prediction task. Two approaches were compared: (1) a direct LLM baseline, in which GPT-4.1 was prompted to predict *ICD-10* codes without constraints, and (2) GAVS, in which GPT-4.1 first generated granular diagnostic entities that were each mapped onto the top 10 matching *ICD-10* codes through vector search over the official ontology. Across all admissions, the vanilla

LLM produced 131,329 candidate codes, while GAVS produced 136,920. At the subcategory level, the vanilla LLM achieved a mean recall of 17.95% (15.86% weighted), whereas GAVS achieved 20.63% (18.62% weighted), representing a statistically significant improvement ($P<.001$, 2-proportion z test). Notably, GAVS generated 11,254 unique predicted subcategories, compared with 15,572 unique subcategories from the vanilla LLM, suggesting that the vanilla approach was more diffuse in its predictions, whereas GAVS concentrated predictions on a narrower and more relevant set of codes. At the category level, the vanilla LLM achieved a mean recall of 34.05% (32.60% weighted), while GAVS achieved 33.57% (32.58% weighted). The difference was not statistically significant ($P=.99$). GAVS produced 1192 unique predicted categories, compared with 913 unique categories for the vanilla LLM.

Integration Into Rhazes Clinician

Building on the GARAG and GAVS prototypes, we developed a clinician-facing agentic web application that integrates documentation assistance, management planning (GARAG), automated coding (GAVS), and differential diagnosis tools and is accessible via the Rhazes website [53]. To maximize accessibility, the system was deployed as a Progressive Web Application, enabling installation and seamless use across desktop and mobile platforms without requiring a native app. The application was implemented using a modern web architecture: a Next.js full-stack framework with React (TypeScript) for the front end, a Node.js backend, and a PostgreSQL database accessed through the Prisma object-relational mapper. Hosting was provided on Aptible [77], a platform-as-a-service offering secure, Health Insurance Portability and Accountability Act–ready infrastructure. Within Rhazes, user queries are handled by LLM agents that route requests to the most appropriate tools. These include the management planning (GARAG) and medical coding (GAVS) pipelines described above, a documentation assistant for completing predefined templates, and a differential diagnosis tool that was previously evaluated in Sarvari and Al-Fagih [47]. The orchestration layer was built on LangGraph [70], allowing for parallel tool execution and a persistent shared conversation history across agents. The system, certified under Cyber Essentials [78,79], is used by thousands of doctors and supports integration with major EHR systems, including Epic (32.8% market share in 2021) and Cerner (23.2% market share in 2021) [80].

Discussion

Principal Findings

In this study, we introduced Rhazes, an AI assistant for doctors designed to handle paperwork and analytical tasks in clinical medicine. Rhazes aims to free physicians from the burden of documentation and to help them provide better care for more patients. Within this broader system, the GARAG and GAVS frameworks demonstrate the feasibility of embedding structured guardrails into LLM-based clinical workflows. GARAG ensures that management plans are grounded in peer-reviewed guideline sources with properly formatted references, achieved through a diagnosis-first retrieval workflow that increases the likelihood of relevant sources being cited. GAVS applies a similar principle to coding, improving fidelity through a 2-step process in which diagnostic entities are generated first and then deterministically mapped to valid ontology terms via vector search. Both methods represent proof-of-concept prototypes that were subsequently deployed within the Rhazes Clinician platform. Although GARAG was evaluated on a smaller case set compared with GAVS, its strength lies in preserving fine-grained diagnostic information during retrieval. By generating diagnoses first and then retrieving guideline evidence for each one, GARAG avoids the information dilution that occurs when the entire patient record is embedded at once, ensuring that management plans remain tightly linked to diagnosis-specific guidance.

Taken together, the evaluation results indicate that GAVS improves resolution at the subcategory level without sacrificing performance at the broader category level. Beyond this quantitative advantage, GAVS has 3 qualitative benefits that strengthen its reliability and scalability. First, GAVS guarantees that every predicted code is part of the official coding ontology. Because predictions are drawn directly from a vector search over the ontology, the system cannot hallucinate nonexistent codes—a risk that remains with unconstrained LLM outputs. Second, GAVS is flexible across coding systems. Adapting it to a different *ICD* version, or to CPT/SNOMED, or to institution-specific ontologies requires no retraining or prompt engineering. One simply replaces the vector database with embeddings of the target ontology's code descriptions, and the method functions seamlessly. Third, GAVS enhances explainability. The LLM provides a structured list of diagnostic predictions together with textual reasoning, and each prediction is then mapped deterministically to a small, fixed set of candidate codes via cosine similarity. This 2-step design ensures systematic and interpretable outputs. By contrast, a vanilla LLM generates codes as a single sequence based on statistical likelihood, with no guarantee of coverage, ordering, or manageable length, making its reasoning harder to audit and its predictions less scalable.

Future Work

There are many feature improvements we envisage adding to Rhazes Clinician soon. First, we plan to experiment with new embedding models such as Guided In-Sample Selection of Training Negatives-large-embedding-v0 [81], which has been identified as a good fit for clinical tasks in a previous study [82]. The change in the embedding model means that we will have to reindex the latest versions of the clinical and coding guidelines we have been using for GARAG and GAVS. From a platform perspective, Rhazes already supports *ICD*, CPT, and SNOMED codes. We plan to extend this support to the full HCPCS [38] (including level 2) as well as OPCS. These additions will broaden coverage across clinical and administrative workflows. From an evaluation perspective, future work will focus on systematically assessing whether GAVS' advantage over an unconstrained LLM generalizes across coding ontologies beyond *ICD-10*. We will design blinded, head-to-head comparisons—similar in spirit to Klang et al [65]—spanning CPT, SNOMED, HCPCS level 2, OPCS, and clinic-specific ontologies, with physicians and LLMs independently adjudicating results. Because these ontologies

(particularly the less common or locally maintained ones) are less likely to be represented in pretraining data, our a priori hypothesis is that the relative benefit of GAVS will be larger than what we observed for *ICD-10*. As part of this program, we aim to construct and share a deidentified, gold-labeled coding dataset suitable for benchmarking across methods. Additional methodological work will examine the effect of the vector-search candidate set size (eg, top-k), alternative embedding models [81,82], and improved parent-mapping resources to reduce unresolvable cases during category aggregation. Finally, we plan to extend our EHR integration offerings: we aim to support Egton Medical Information Systems, the leading EHR for UK primary care clinics, and SystmOne, the second most popular EHR for UK general practitioners [83]. These integrations will facilitate prospective, multisite evaluations and subgroup analyses while maintaining interoperability with existing clinical systems.

Limitations

The GARAG workflow was tested on a relatively small set of 21 author-designed cases. While reproducibility was high, independent validation on larger and more varied case sets is needed. The GAVS evaluation, while based on a sizable cohort of 958 admissions, relied on billing records as the gold standard. Because billing data do not fully capture the clinical picture of each admission, it is not possible to definitively establish precision, as some well-reasoned diagnostic predictions may go unbilled [46,47]. Moreover, the underlying MIMIC-IV dataset has well-recognized constraints: it lacks clinical notes, physical examination findings, and certain test results such as electrocardiograms, and it is drawn from a single hospital in Boston, MA. This means the data are subject to demographic and institutional biases and may not generalize to other patient populations. Finally, some specific *ICD-10* codes could not be mapped to parent categories due to library limitations. These factors highlight the need for further testing against richer clinical datasets, across multiple institutions, and with more comprehensive ontology mappings. Taken together, these component-level limitations reflect broader challenges in deploying LLM-driven systems like Rhazes into clinical practice. The effectiveness of an AI co-pilot hinges on its accuracy across diverse clinical scenarios. In general, evaluation of clinical performance of LLMs is challenging due to the lack of transparency when it comes to versions, prompts, human evaluations, LLM-as-a-judge evaluations [47,76], patient data, and due to the nonexistence of gold-labeled data sets for many clinical applications [84]. Care must be taken to accurately assess AI for improved patient outcomes and to avoid statistically flawed evaluations [85]. In practice, AI tools can exhibit degraded performance when used outside the conditions of their training (out-of-distribution use). Even models that performed well in development or obtained regulatory clearance have underperformed in new settings due to poor generalization. This raises the risk of missed diagnoses or incorrect management plans if the AI encounters patient data that differs from its training distribution. Continuous validation of the system on local patient populations is therefore critical to ensure reliability in AI-generated recommendations [86]. AI models learn from historical data, so any biases or gaps in those data can lead to skewed or inequitable outcomes [87]. If the training dataset underrepresents certain demographics or conditions, the model suggestions may be less accurate for those groups, potentially perpetuating health disparities. For example, studies have found some clinical AI algorithms perform significantly worse for female patients or racial minorities, underdiagnosing these groups compared to others [88,89]. Such bias not only affects accuracy but also violates principles of fairness in care. Ensuring the data used by AI co-pilots are diverse and representative is essential to minimize this risk. We must also be mindful of other harmful biases LLMs may learn during training [84], as well as the risks that over-reliance on AI systems may bring to medicine (eg, automation bias) [90].

AI in Health Care Ethics

Under the General Data Protection Regulation, health care organizations can often process patient data for care without explicit consent, provided they have a valid lawful basis (Article 6) and meet a special category condition (Article 9) [91], such as provision of health care services. This lawful basis should naturally extend to the data processors used, such as the AI scribes. However, alongside lawfulness, transparency is important. Patients should be informed when an algorithm is involved in their diagnosis or treatment planning. Research indicates that disclosing the use of an AI tool is essential to patients; a recent study found that patients strongly prefer to be informed when AI assists in their care and recommended that explicit consent for AI involvement be obtained during the clinical workflow [92]. In the context of Rhazes, this means clinicians should be transparent about the AI's role—explaining to patients that an AI system will analyze their data and contribute to suggestions. Such transparency not only respects patient autonomy but also helps build trust, as patients are more likely to accept AI-derived recommendations if they understand and agree to their use.

An AI system must consistently uphold the core principles of medical ethics—beneficence, nonmaleficence, autonomy, and justice. One concern is that software like Rhazes might, in some situations, propose an option that, while data-driven, conflicts with a patient's values or broader ethical norms. For example, an AI might prioritize treatments based on statistical outcomes or cost-effectiveness, which could unintentionally de-emphasize a patient's personal preference for quality of life. If Rhazes recommends an aggressive treatment purely because it maximizes survival odds, but the patient prioritizes comfort, blindly following the AI would undermine patient autonomy. Human clinicians must interpret Rhazes' outputs through the lens of their professional ethics and clinical judgment. They should override or adjust recommendations that do not fit the patient's individual context or the ethical standards of care. In essence, Rhazes should support clinical decisions that are not only effective but also ethically sound, with the physician ensuring final decisions align with the principle of autonomy and patient-centered care. Several approaches can address these ethical concerns and ensure that AI tools are used responsibly in health care. One key strategy is incorporating explainability into the AI model. Rather than acting as a "black box," Rhazes provides interpretable reasoning or an explanation for its suggestions (for instance, highlighting which patient factors or

medical evidence led to a given diagnostic recommendation). Explainable AI methods help clinicians and patients understand why a recommendation was made, which is vital for trust and for verifying that the recommendation makes ethical and clinical sense. Another strategy is clinician oversight and accountability. Rhazes is intended to assist, not replace, the clinician; therefore, protocols should emphasize that the human provider retains ultimate responsibility for diagnosis and treatment decisions. By maintaining clear accountability—where the clinician must review and approve AI-generated plans—the risk of blind adoption of incorrect suggestions is reduced. Studies on automation bias mitigation have noted that training users and stressing their accountability can counter overreliance [93]. Regular training sessions for clinicians on the proper use of Rhazes, including case studies of when the AI errs, can sharpen their judgment on when to trust the AI and when to apply caution. Finally, patient education about AI in health care can help. Patients should be informed in understandable terms what Rhazes is and what role it plays in their care. When patients understand that the AI is a tool used by their doctor (and not a substitute for the doctor), it can alleviate fears of a purely machine-driven care plan. Surveys have shown that both doctors and patients feel anxious if they do not understand AI's involvement [94], so educational efforts (leaflets, consent discussions, etc) can demystify the technology. In summary, through explainable AI design, strong human oversight, and educational transparency, Rhazes can be deployed in a way that upholds ethical standards and supports clinicians and patients alike.

Compliance Requirements for AI Tools in Hospitals

Any digital health technology company operating within the United Kingdom collecting or processing any form of personal data must comply with UK General Data Protection Regulation and the Data Protection Act [95]. Companies processing personal data in the United Kingdom must be registered with the Information Commissioner's Office (ICO) [96]. In addition to the usual requirements around processing personal data, it is likely that a health-tech company will be processing sensitive personal health data, which would classify as special category data. This can bring some additional requirements, such as the need or recommendation to complete a Data Protection Impact Assessment [97]. There must also be appropriate contracts and Data Processing Agreements [98] in place between an NHS organization and the digital health supplier, between which personal data may flow.

If a digital health supplier is looking to work with NHS organizations and will be interacting with NHS patient data, they will need to complete the Data Security and Protection Toolkit (DSPT) [99]. This is not a requirement in the private sector or in direct-to-consumer models. If an organization is an IT supplier with 50+ staff members and has a turnover of at least £10 million (US $13.16 million) and supplies digital goods and services to the NHS, the company must also undertake an independent audit/assessment [99]. Organizations handling patient data may require the following personnel: a data protection officer (DPO), a senior information risk owner, and a Caldicott guardian [99]. A DPO [100] is required if a company's core activities consist of large-scale processing of special category data. DPOs help to monitor internal compliance, inform on data protection obligations, provide advice regarding Data Protection Impact Assessments and act as a point of contact for data subjects and the ICO. The Senior Information Risk Owner is a senior member of the organization whose roles are to promote a culture that values and protects ICO information, own information risk management policies and processes and ensure they are implemented, advise on information risk management processes and provide assurance, and own the incident management framework. The Caldicott Guardian [101] is a senior person responsible for overseeing the use and sharing of patient information by protecting the confidentiality of people's health and care information.

Companies aiming to deploy in the NHS also must go through information security and technical assurance. UK Cyber Essentials [78,102,103] is a self-assessment that any company looking to work with the public sector must comply with. The general recommendation is for companies to comply with Cyber Essentials Plus, which involves both a self-assessment and an external audit. Often, companies will use ISO 27001 to demonstrate a higher level of security than required by just meeting DSPT requirements, but it is not generally mandated in health care organizations. ISO 27001 is an internationally recognized standard for information security [102] that is not health care specific. It requires companies to implement Information Security Management Systems and focuses on risk assessment. It requires independent certification by an accredited body. Digital health technologies deployed in NHS organizations in England will also need to comply with the digital clinical safety standards DCB0129/0160. This is required by law, under section 250 of the Health and Social Care Act 2012 [104]. Both the manufacturer (DCB0129) and deploying health or social care organization (DCB01060) are required to complete a clinical risk assessment, including key documentation. This process is overseen by an appropriately qualified Clinical Safety Officer. There is also a requirement to monitor and record any incidents post-deployment. Penetration testing is required to assess the security of digital health technologies deployed in NHS organizations as part of the NHS Digital Technology Assessment Criteria (DTAC) [105], which mandates that any identified vulnerabilities should be appropriately remediated. The NHS DTAC [105] is a framework that brings together legislation and best practice in 5 core areas: clinical safety, data protection, technical security, interoperability, and accessibility and usability, which incorporates the aforementioned NHS DSPT and DCB0129 standards. The DTAC is a national baseline criteria for digital health technologies being deployed within NHS health care organizations and can be used by health care organizations to assess suppliers as part of their due diligence process. Due to the complexity of navigating NHS compliance frameworks, specialized firms have emerged to help digital health companies accelerate clinical assurance processes; these include Assuric [106], Vanta [107], and Naq [108].

If the AI product meets the definition of software as a medical device (SaMD), companies would also need to comply with medical device regulation and achieve appropriate certifications before being available for use in the open market. This would be the case if the intended purpose and functionality of the

product extends into diagnosis, prevention, monitoring, prediction, prognosis, treatment, or alleviation of disease, as defined by European Union (EU) Medical Device Regulation [109]. SaMD is defined as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device" [110]. In the United Kingdom, medical devices are classified by risk to class 1, class 2a, class 2b, and class 3, with class 1 being low risk to patients and class 3 being high risk to patients. Manufacturers face a greater scope of work and evidentiary burden when dealing with higher-risk products. In the United Kingdom, low-risk class 1 devices require manufacturers to make a self-declaration of conformity to the Medicines and Healthcare Products Regulatory Agency (MHRA). Other classes require involvement and approval from an approved body (an organization designated by the MHRA to assess the conformity of products before they are placed on the market), granting a UK Conformity Assessed mark, the equivalent to a Conformité Européenne (CE) mark in the EU.

There have been several AI-as-a-medical device products on the market for some time, primarily in the category of diagnostic radiology or dermatology tools, one example being Skin Analytics, which recently achieved regulatory approval for autonomous AI skin cancer detection system Deep Ensemble for Recognition of Malignancy in Europe, receiving class III CE marking. This is the first legally authorized AI to independently make clinical decisions on skin cancer without oversight. Deep Ensemble for Recognition of Malignancy achieves 99.8% accuracy rate in ruling out cancer, surpassing the performance of dermatologists who typically achieve 98.9% [111]. However, there is yet to be a generative AI product that has been certified as a medical device in the United Kingdom or EU. In the United States, Modella AI's generative AI co-pilot, PathChat, has received device designation by the Food and Drug Administration. This is the first regulatory approval of a clinical-grade generative AI co-pilot [112] and is the first of likely many more SaMD generative AI applications.

The NHS 10-Year Reform Plan sets a clear direction for modernizing care delivery, with a strong emphasis on digital transformation, integrated community services, and reducing strain on clinical staff [113]. One area gaining significant momentum is the deployment of AI-enabled ambient scribing tools, which offer practical relief from administrative overhead by automatically transcribing and summarizing clinical encounters [113]. The plan explicitly highlights the need to streamline documentation and use responsible automation to release clinician time for patient care [113]. In parallel, NHS England's technical guidance on ambient voice technology adds further clarity and outlines key regulatory considerations for these tools [114]. Pure transcription tools are generally not considered medical devices. However, where generative AI features extend into summarization, providing prompts, generating structured clinical notes, letters, or codes, they are likely to qualify as SaMD. Such tools would then require UK Conformity Assessed or CE marking, MHRA registration, and a full clinical safety case under DCB0129/0160. Beyond regulatory certification, NHS organizations are expected to ensure integration with existing EHRs through standards such as Fast Health Care Interoperability Resources, HL7, and SNOMED CT, maintain strong human oversight to mitigate diagnostic drift or foreseeable misuse, and implement a clear post-deployment monitoring framework. This includes mechanisms for clinicians to flag transcription errors, routine audits of scribe outputs, and attention to bias risks, particularly for patients with regional accents, dialects, or speech impairments.

Conclusion

While AI has been rapidly evolving over the last 2 years, progress has not been reciprocated in the health care industry, a heavily regulated space with many financial, staffing, and quality-of-service-type problems. Due to the lack of gold-labeled datasets and human evaluation protocols for LLM-generated text, recent AI in health care innovations was driven by well-funded industry players who were able to start generating evidence by securing hospital pilots early. So far, most companies seek to innovate in administrative workflows that avoid direct patient care as this comes with lesser regulatory burden. There seems to be a regulatory gray area surrounding workflows which could ultimately affect patient care should doctors over-rely on AI. Examples include AI scribing and clinical document generation, with only a few AI notetaker tools evaluated in academic journals with often conflicting and lack of reproducible results.

In this article, we reviewed the need for AI tools in health care and the current state of the industry, including dominant players and their progress. During this review, we demonstrated, firsthand, how such tools may be created and how they may be used by physicians. We discussed key implementation considerations for Rhazes Clinician, an AI assistant for doctors. We described in detail the methods used to create a CDS and an admin assistant for doctors, including the LLMs deployed, the clinical guidelines used, the RAG hyperparameters, and the cloud services used. We also introduced a new method for medical coding that we dubbed GAVS for Generation-Augmented Vector Search and an improved RAG workflow for CDS that we named GARAG. GARAG highlights the value of diagnosis-specific retrieval, allowing management plans to stay closely linked to diagnostic evidence while avoiding the information dilution that occurs when entire patient records are embedded in typical RAG workflows. We showed that GAVS statistically significantly improves *ICD-10* coding predictions. For both treatment planning and medical coding, we provided Jupyter notebooks that demonstrate the (albeit simplified) implementation of these Rhazes tools. Our goal with this is to contribute to the academic discussion about AI tools for health care and encourage academics as well as industry players to share their datasets and novel methods in order to accelerate the deployment of transparent AI tools in hospitals.

Acknowledgments

The authors would like to thank the thousands of doctors who to this date signed up to try out the Rhazes app. Their support, feedback, and interaction with the tool have been an invaluable source of information and inspiration for improving the product and its capabilities. The authors have received no funding to conduct this study.

Data Availability

The Medical Information Mart for Intensive Care (MIMIC)–IV data is available to approved researchers on PhysioNet, and the SQL code used to transform this dataset is available at GitHub [115].

The Jupyter notebook showcasing the generation-assisted retrieval-augmented generation system for treatment plan generation is publicly available at GitHub [116].

The Jupyter notebook showcasing the retrieval-augmented generation system for treatment plan generation is publicly available at GitHub [117].

The Jupyter notebook showcasing the generation-assisted vector search (GAVS) system for automated Current Procedural Terminology (CPT) coding is publicly available at GitHub [118].

The Jupyter notebook containing the implementation of the GAVS system for automated *International Classification of Diseases, Tenth Revision* (*ICD-10*) coding, together with its comparison to the vanilla GPT-4.1 coding approach and the full evaluation on a subset of the MIMIC-IV dataset, is publicly available at GitHub [119].

CPT data are available to download from the Centers for Medicare & Medicaid Services website [72].

International Classification of Diseases, Eleventh Revision (*ICD-11*) data are available in the following file [73]. In addition, mappings between *ICD-10* and *ICD-11* can be found in the following file [120].

Systematized Nomenclature of Medicine Clinical Terms (SNOMED) clinical codes can be found at NHS TRUD [74].

SNOMED drug codes can be found at NHS TRUD as well [121].

Conflicts of Interest

PS, ZA, and AA-C are currently affiliated with Rhazes, which is a health tech startup discussed throughout the paper. PJ and RT are affiliated with Assuric, a company providing digital health compliance services, mentioned once in the paper. AI has previously been affiliated with Rhazes but is currently not affiliated with Rhazes. The authors invite the readers to independently verify the results of this article using the scripts shared in the Data Availability Statement.

References

1. Rodziewicz TL, Houseman B, Vaqar S, Hipskind JE. Medical Error Reduction and Prevention: StatPearls Publishing; 2025. [Medline: 29763131]
2. Herd P, Moynihan D. Health care administrative burdens: centering patient experiences. Health Serv Res 2021 Oct;56(5):751-754. [doi: 10.1111/1475-6773.13858] [Medline: 34515996]
3. Rimmer A. Staff shortages are affecting doctors' mental health, survey finds. BMJ 2023 May 17:1121. [doi: 10.1136/bmj.p1121]
4. Haddad LM, Annamaraju P, Toney-Butler TJ. Nursing Shortage: StatPearls Publishing; 2025. [Medline: 29630227]
5. Berwick DM. Salve Lucrum: the existential threat of greed in US health care. JAMA 2023 Feb 28;329(8):629-630. [doi: 10.1001/jama.2023.0846] [Medline: 36716043]
6. Singh R, Volner K, Marlowe D. Provider Burnout: StatPearls Publishing; 2025. [Medline: 30855914]
7. Khullar D. Burnout, professionalism, and the quality of US health care. JAMA Health Forum 2023 Mar 3;4(3):e230024. [doi: 10.1001/jamahealthforum.2023.0024] [Medline: 36961455]
8. Shemtob L, Asanati K, Pahl N, Majeed A. What needs to be done to address staffing shortages in health and social care? Br J Gen Pract 2023 Mar;73(728):102-103. [doi: 10.3399/bjgp23X732045] [Medline: 36823045]
9. McCarey M. Brexit adds further complexity to the health and care staffing crisis. BMJ 2023 Jan 31;380:208. [doi: 10.1136/bmj.p208] [Medline: 36720481]
10. Riley RF, Alasnag M, Batchelor WB, et al. The ongoing national medical staffing crisis: impacts on care delivery for interventional cardiologists. J Soc Cardiovasc Angiogr Interv 2022;1(3):100307. [doi: 10.1016/j.jscai.2022.100307] [Medline: 35814944]
11. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J 2021 Jul;8(2):e188-e194. [doi: 10.7861/fhj.2021-0095] [Medline: 34286183]
12. Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform 2020 Jul 24;8(7):e18599. [doi: 10.2196/18599] [Medline: 32706688]

13. Topol EJ. Toward the eradication of medical diagnostic errors. Science 2024 Jan 26;383(6681):eadn9602. [doi: 10.1126/science.adn9602] [Medline: 38271508]
14. Khanna NN, Maindarkar MA, Viswanathan V, et al. Economics of artificial intelligence in healthcare: diagnosis vs. treatment. Healthcare (Basel) 2022 Dec 9;10(12):2493. [doi: 10.3390/healthcare10122493] [Medline: 36554017]
15. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak 2023 Apr 20;23(1):73. [doi: 10.1186/s12911-023-02162-y] [Medline: 37081503]
16. Olawade DB, David-Olawade AC, Wada OZ, Asaolu AJ, Adereni T, Ling J. Artificial intelligence in healthcare delivery: prospects and pitfalls. J Med Surg Public Health 2024 Aug;3:100108. [doi: 10.1016/j.glmedi.2024.100108]
17. Esmaeilzadeh P. Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: a perspective for healthcare organizations. Artif Intell Med 2024 May;151:102861. [doi: 10.1016/j.artmed.2024.102861] [Medline: 38555850]
18. Lambert SI, Madi M, Sopka S, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111. [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
19. Razai MS, Al-Bedaery R, Bowen L, Yahia R, Chandrasekaran L, Oakeshott P. Implementation challenges of artificial intelligence (AI) in primary care: perspectives of general practitioners in London UK. PLoS ONE 2024;19(11):e0314196. [doi: 10.1371/journal.pone.0314196] [Medline: 39570873]
20. Newman-Toker DE, Wang Z, Zhu Y, et al. Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: toward a national incidence estimate using the “Big Three”. Diagnosis (Berl) 2021 Feb 23;8(1):67-84. [doi: 10.1515/dx-2019-0104] [Medline: 32412440]
21. Neale G, Hogan H, Sevdalis N. Misdiagnosis: analysis based on case record review with proposals aimed to improve diagnostic processes. Clin Med (Lond) 2011 Aug;11(4):317-321. [doi: 10.7861/clinmedicine.11-4-317] [Medline: 21853823]
22. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016 May 3;353:i2139. [doi: 10.1136/bmj.i2139] [Medline: 27143499]
23. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014 Sep;23(9):727-731. [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
24. Balogh EP, Miller BT. Improving Diagnosis in Health Care: National Academies Press; 2015. [doi: 10.17226/21794]
25. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med 2005 Jul 11;165(13):1493-1499. [doi: 10.1001/archinte.165.13.1493] [Medline: 16009864]
26. David E. Newman-Toker MD, PhD S, Shervin Badihian MD, et al. Diagnostic Errors in the Emergency Department: A Systematic Review: Agency for Healthcare Research and Quality (US); 2022.
27. Densen P. Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc 2011;122(48–58):48-58. [Medline: 21686208]
28. UpToDate. Wolters Kluwer. URL: https://www.wolterskluwer.com/en/solutions/uptodate [accessed 2025-10-31]
29. Charbonneau DH, James LN. DynaMed Plus®: an evidence-based clinical reference resource. Med Ref Serv Q 2018 Mar;37(2):168-176. [doi: 10.1080/02763869.2018.1439221] [Medline: 29558334]
30. StatPearls. URL: https://www.statpearls.com [accessed 2025-10-31]
31. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
32. Shanafelt TD, Hasan O, Dyrbye LN, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. Mayo Clin Proc 2015 Dec;90(12):1600-1613. [doi: 10.1016/j.mayocp.2015.08.023] [Medline: 26653297]
33. Han S, Shanafelt TD, Sinsky CA, et al. Estimating the attributable cost of physician burnout in the United States. Ann Intern Med 2019 Jun 4;170(11):784-790. [doi: 10.7326/M18-1422] [Medline: 31132791]
34. Wise J. Burnout among trainees is at all time high, GMC survey shows. BMJ 2022 Jul 19:o1796. [doi: 10.1136/bmj.o1796]
35. Chang E, Sung S. Use of SNOMED CT in large language models: scoping review. JMIR Med Inform 2024 Oct 7;12:e62924. [doi: 10.2196/62924] [Medline: 39374057]
36. Kersloot MG, van Putten FJP, Abu-Hanna A, Cornet R, Arts DL. Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies. J Biomed Semantics 2020 Nov 16;11(1):14. [doi: 10.1186/s13326-020-00231-z] [Medline: 33198814]
37. Horsky J, Drucker EA, Ramelson HZ. Accuracy and completeness of clinical coding using ICD-10 for ambulatory visits. AMIA Annu Symp Proc 2017;2017:912-920. [Medline: 29854158]
38. HCPCS - general information. Centers for Medicare and Medicaid Services. 2025. URL: https://www.cms.gov/medicare/coding-billing/healthcare-common-procedure-system [accessed 2025-10-31]
39. Dong H, Falis M, Whiteley W, et al. Automated clinical coding: what, why, and where we are? NPJ Digit Med 2022 Oct 22;5(1):159. [doi: 10.1038/s41746-022-00705-7] [Medline: 36273236]
40. Alonso V, Santos JV, Pinto M, et al. Problems and barriers during the process of clinical coding: a focus group study of coders’ perceptions. J Med Syst 2020 Feb 8;44(3):62. [doi: 10.1007/s10916-020-1532-x] [Medline: 32036459]

41. Mesko B. The ChatGPT (generative artificial intelligence) revolution has made artificial intelligence approachable for medical professionals. J Med Internet Res 2023 Jun 22;25:e48392. [doi: 10.2196/48392] [Medline: 37347508]
42. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023 Jun 28;25:e48568. [doi: 10.2196/48568] [Medline: 37379067]
43. Madrid-García A, Rosales-Rosado Z, Freites-Nuñez D, et al. Harnessing ChatGPT and GPT-4 for evaluating the rheumatology questions of the Spanish access exam to specialized medical training. Sci Rep 2023 Dec 13;13(1):22129. [doi: 10.1038/s41598-023-49483-6] [Medline: 38092821]
44. Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish medical final examination. Sci Rep 2023 Nov 22;13(1):20512. [doi: 10.1038/s41598-023-46995-z] [Medline: 37993519]
45. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023 Oct 1;13(1):16492. [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
46. Sarvari P, Al-Fagih Z, Ghuwel A, Al-Fagih O. A systematic evaluation of the performance of GPT-4 and PaLM2 to diagnose comorbidities in MIMIC-IV patients. Health Care Sci 2024 Feb;3(1):3-18. [doi: 10.1002/hcs2.79] [Medline: 38939167]
47. Sarvari P, Al-Fagih Z. Rapidly benchmarking large language models for diagnosing comorbid patients: comparative study leveraging the LLM-as-a-judge method. JMIRx Med 2025 Aug 29;6:e67661. [doi: 10.2196/67661] [Medline: 40880236]
48. Korom R, Kiptinness S, Adan N, et al. AI-based clinical decision support for primary care: a real-world study. arXiv. Preprint posted online on Jul 22, 2025. [doi: 10.48550/arXiv.2507.16947]
49. Elsevier ClinicalKey. Elsevier. 2025. URL: https://www.elsevier.com/en-gb/products/clinicalkey [accessed 2025-10-31]
50. Medwise. URL: https://www.medwise.ai [accessed 2025-11-04]
51. Elhaddad M, Hamam S. AI-driven clinical decision support systems: an ongoing pursuit of potential. Cureus 2024 Apr;16(4):e57728. [doi: 10.7759/cureus.57728] [Medline: 38711724]
52. Glass Health. URL: https://www.glass.health [accessed 2025-10-31]
53. Rhazes AI. URL: https://www.rhazes.ai [accessed 2025-10-31]
54. Tierney AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical socumentation. NEJM Catalyst 2024 Feb 21;5(3). [doi: 10.1056/CAT.23.0404]
55. Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. J Am Med Inform Assoc 2025 Feb 1;32(2):375-380. [doi: 10.1093/jamia/ocae295] [Medline: 39657021]
56. Balloch J, Sridharan S, Oldham G, et al. Use of an ambient artificial intelligence tool to improve quality of clinical documentation. Future Healthc J 2024 Sep;11(3):100157. [doi: 10.1016/j.fhj.2024.100157] [Medline: 39371531]
57. Doshi GK, Jensen TL, Graziano A, Enenmoh C, Lindsey J. Use of ambient AI scribing: impact on physician administrative burden and patient care. JCO Oncol Pract 2024 Oct;20(10_suppl):418-418. [doi: 10.1200/OP.2024.20.10_suppl.418]
58. Cao DY, Silkey JR, Decker MC, Wanat KA. Artificial intelligence-driven digital scribes in clinical documentation: pilot study assessing the impact on dermatologist workflow and patient encounters. JAAD Int 2024 Jun;15:149-151. [doi: 10.1016/j.jdin.2024.02.009] [Medline: 38571698]
59. Haberle T, Cleveland C, Snow GL, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. J Am Med Inform Assoc 2024 Apr 3;31(4):975-979. [doi: 10.1093/jamia/ocae022] [Medline: 38345343]
60. Nisar A, Nandakumar A, Elhamri M, et al. 3232 AI-assisted transcription in healthcare: enhancing efficiency and quality in clinical documentation. Emerg Med J 2025 Jun;42(6):e1. [doi: 10.1136/emermed-2025-RED.1]
61. Ma SP, Liang AS, Shah SJ, et al. Ambient artificial intelligence scribes: utilization and impact on documentation time. J Am Med Inform Assoc 2025 Feb 1;32(2):381-385. [doi: 10.1093/jamia/ocae304] [Medline: 39688515]
62. Campbell S, Giadresco K. Computer-assisted clinical coding: a narrative review of the literature on its benefits, limitations, implementation and impact on clinical coding professionals. HIM J 2020 Jan;49(1):5-18. [doi: 10.1177/1833358319851305]
63. Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. NPJ Digit Med 2023 Feb;6(1):16. [doi: 10.1038/s41746-023-00768-0]
64. Edin J, Junge A, Havtorn JD, et al. Automated medical coding on MIMIC-III and MIMIC-IV: a critical review and replicability study. Presented at: SIGIR '23; Jul 23-27, 2023. [doi: 10.1145/3539618.3591918]
65. Klang E, Tessler I, Apakama DU, et al. Assessing retrieval-augmented large language model performance in emergency department ICD-10-CM coding compared to human coders. medRxiv. 2024 Oct 17 p. 2024.10.15.24315526. [doi: 10.1101/2024.10.15.24315526] [Medline: 39484238]
66. Corti AI. URL: https://www.corti.ai [accessed 2025-10-31]
67. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 3;10(1):1. [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
68. Responsible use of MIMIC data with online services like GPT. PhysioNet. 2023. URL: https://physionet.org/news/post/gpt-responsible-use [accessed 2025-10-31]
69. New embedding models and API updates. OpenAI. 2024. URL: https://openai.com/index/new-embedding-models-and-api-updates [accessed 2025-10-31]
70. LangChain. URL: https://www.langchain.com [accessed 2025-10-31]

71. Pinecone. URL: https://www.pinecone.io [accessed 2025-10-31]
72. License for use of Current Procedural Terminology, Fourth Edition ("CPT®"). Centers for Medicare & Medicaid Services. 2024. URL: https://www.cms.gov/license/ama?file=/files/zip/list-codes-effective-january-1-2025-published-november-26-2024.zip [accessed 2025-10-31]
73. SimpleTabulation-ICD-11-MMS-en. International Classification of Diseases. 2024. URL: https://icdcdn.who.int/static/releasefiles/2024-01/SimpleTabulation-ICD-11-MMS-en.zip [accessed 2025-10-31]
74. SNOMED CT. NHS TRUD. 2025. URL: https://isd.digital.nhs.uk/trud/users/authenticated/filters/0/categories/40/items/279/releases [accessed 2025-10-31]
75. simple-icd-10 2.1.1. Python Package Index. 2025. URL: https://pypi.org/project/simple-icd-10/ [accessed 2025-10-31]
76. Kahng M, Tenney I, Pushkarna M, et al. LLM comparator: interactive analysis of side-by-side evaluation of large language models. IEEE Trans Vis Comput Graph 2025 Jan;31(1):503-513. [doi: 10.1109/TVCG.2024.3456354] [Medline: 39255096]
77. Aptible. URL: https://www.aptible.com [accessed 2025-10-31]
78. Cyber essentials. National Cyber Security Centre. 2025. URL: https://www.ncsc.gov.uk/cyberessentials/overview [accessed 2025-10-31]
79. Essentials certification – CyberSmart Rhazes AI Cyber. Blockmark Registry. 2024. URL: https://registry.blockmarktech.com/certificates/d7d89e73-f326-4586-ba4a-397a0281f741 [accessed 2025-10-31]
80. Holmgren AJ, Apathy NC. Trends in US hospital electronic health record vendor market concentration, 2012-2021. J Gen Intern Med 2023 May;38(7):1765-1767. [doi: 10.1007/s11606-022-07917-3] [Medline: 36348217]
81. Solatorio AV. GISTEmbed: guided in-sample selection of training negatives for text embedding fine-tuning. arXiv. Preprint posted online on Feb 26, 2024. [doi: 10.48550/arXiv.2402.16829]
82. Soffer S, Glicksberg BS, Kovatch P, et al. A scalable framework for benchmarking embedding models for semantic medical tasks. medRvix. Preprint posted online on Aug 20, 2024. [doi: 10.1101/2024.08.14.24312010]
83. Sheikh A, Anderson M, Albala S, et al. Health information technology and digital innovation for national learning health and care systems. Lancet Digit Health 2021 Jun;3(6):e383-e396. [doi: 10.1016/S2589-7500(21)00005-4] [Medline: 33967002]
84. Moreno AC, Bitterman DS. Toward clinical-grade evaluation of large language models. Int J Radiat Oncol Biol Phys 2024 Mar 15;118(4):916-920. [doi: 10.1016/j.ijrobp.2023.11.012] [Medline: 38401979]
85. Byrne DW, Domenico HJ, Moore RP. Artificial intelligence for improved patient outcomes-the pragmatic randomized controlled trial is the secret sauce. Korean J Radiol 2024 Feb;25(2):123-125. [doi: 10.3348/kjr.2023.1016] [Medline: 38238014]
86. Mahmood U, Shukla-Dave A, Chan HP, et al. Artificial intelligence in medicine: mitigating risks and maximizing benefits via quality assurance, quality control, and acceptance testing. BJR Artificial Intelligence 2024 Mar 4;1(1). [doi: 10.1093/bjrai/ubae003]
87. Zack T, Lehman E, Suzgun M, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024 Jan;6(1):e12-e22. [doi: 10.1016/S2589-7500(23)00225-X] [Medline: 38123252]
88. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
89. Cao J, Zhang X, Shahinian V, et al. Generalizability of an acute kidney injury prediction model across health systems. Nat Mach Intell 2022 Dec;4(12):1121-1129. [doi: 10.1038/s42256-022-00563-8] [Medline: 38148789]
90. Jabbour S, Fouhey D, Shepard S, et al. Measuring the impact of AI in the diagnosis of hospitalized patients. JAMA 2023 Dec 19;330(23):2275. [doi: 10.1001/jama.2023.22295]
91. Regulation (EU) 2016/679 of the European Parliament and of the Council. Legislation.gov.uk. 2016. URL: https://www.legislation.gov.uk/eur/2016/679/contents [accessed 2025-10-31]
92. Park HJ. Patient perspectives on informed consent for medical AI: a web-based experiment. Digit Health 2024;10:20552076241247938. [doi: 10.1177/20552076241247938] [Medline: 38698829]
93. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127. [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
94. Li W, Liu X. Anxiety about artificial intelligence from patient and doctor-physician. Patient Educ Couns 2025 Apr;133:108619. [doi: 10.1016/j.pec.2024.108619] [Medline: 39721348]
95. Data protection. GOV.UK. 2025. URL: https://www.gov.uk/data-protection [accessed 2025-10-31]
96. Data protection principles, definitions, and key terms. Information Commissioner's Office. 2025. URL: https://ico.org.uk/for-organisations/advice-for-small-organisations/key-data-protection-terms-you-need-to-know [accessed 2025-10-31]
97. Data protection impact assessments. Information Commissioner's Office. 2025. URL: https://ico.org.uk/for-organisations/law-enforcement/guide-to-le-processing/accountability-and-governance/data-protection-impact-assessments [accessed 2025-10-31]
98. Data protection agreements and contracts. AI and Digital Regulations Service for health and social care. 2023. URL: https://www.digitalregulations.innovation.nhs.uk/regulations-and-guidance-for-adopters/all-adopters-guidance/data-protection-agreements-and-contracts [accessed 2025-10-31]
99. NHS England Data Security and Protection Toolkit. URL: https://www.dsptoolkit.nhs.uk [accessed 2025-10-31]

100. Data protection officers. Information Commissioner's Office. 2025. URL: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/data-protection-officers [accessed 2025-10-31]
101. The UK Caldicott Guardian Council. URL: https://www.ukcgc.uk [accessed 2025-10-31]
102. Frameworks that can help. NHS England. 2025. URL: https://digital.nhs.uk/cyber-and-data-security/guidance-and-assurance/data-security-and-protection-toolkit-assessment-guides/guide-9---it-protection/frameworks-that-can-help [accessed 2025-10-31]
103. Cyber security and resilience for health or care services. AI and Digital Regulations Service for health and social care. 2025. URL: https://www.digitalregulations.innovation.nhs.uk/regulations-and-guidance-for-adopters/all-adopters-guidance/cyber-security-and-resilience-for-health-or-care-services [accessed 2025-10-31]
104. DCB0129: clinical risk management. NHS England Digital. 2023. URL: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections [accessed 2025-10-31]
105. Fort J. What are NHS DTAC penetration testing requirements? Blaze. 2024. URL: https://www.blazeinfosec.com/post/nhs-dtac-penetration-testing-requirement [accessed 2025-10-31]
106. Assuric. URL: https://www.assuric.com [accessed 2025-10-31]
107. Vanta. URL: https://www.vanta.com [accessed 2025-10-31]
108. Naq. URL: https://www.naqcyber.com [accessed 2025-10-31]
109. The Medical Devices Regulations. Legislation.gov.uk. 2002. URL: https://www.legislation.gov.uk/uksi/2002/618/contents [accessed 2025-10-31]
110. Crafting an intended purpose in the context of software as a medical device (SaMD). GOV.UK. 2023. URL: https://www.gov.uk/government/publications/crafting-an-intended-purpose-in-the-context-of-software-as-a-medical-device-samd/crafting-an-intended-purpose-in-the-context-of-software-as-a-medical-device-samd [accessed 2025-10-31]
111. DERM makes medical history as world's first autonomous skin cancer detection system is approved for clinical decisions in Europe. Skin Analytics. 2025. URL: https://skin-analytics.com/news/regulatory-certification/derm-class-iii-ce-mark [accessed 2025-10-31]
112. Press release: Modella AI's generative AI co-pilot pathchat receives FDA breakthrough device designation. Modella AI. 2025. URL: https://modella.ai/pathchat-fda-breakthrough-designation.html [accessed 2025-10-31]
113. Fit for the future: 10 year health plan for England. GOV.UK. 2025. URL: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future/fit-for-the-future-10-year-health-plan-for-england-executive-summary [accessed 2025-10-31]
114. Guidance on the use of AI-enabled ambient scribing products in health and care settings. NHS England. 2025. URL: https://www.england.nhs.uk/long-read/guidance-on-the-use-of-ai-enabled-ambient-scribing-products-in-health-and-care-settings [accessed 2025-10-31]
115. MIMIC-SQL. GitHub. URL: https://github.com/sarvarip/MIMIC-SQL [accessed 2025-11-04]
116. IndexingPipelines/clean_gpt41_garagpipelineandeval.ipynb. GitHub. URL: https://github.com/rhazes-dev/IndexingPipelines/blob/main/Clean_GPT41_GARAGPipelineAndEval.ipynb [accessed 2025-11-04]
117. IndexingPipelines/treatmentplanragpipeline_clean.ipynb. GitHub. URL: https://github.com/rhazes-dev/IndexingPipelines/blob/main/TreatmentPlanRAGPipeline_clean.ipynb [accessed 2025-11-04]
118. IndexingPipelines/cpt_gavs_pipeline_clean.ipynb. GitHub. URL: https://github.com/rhazes-dev/IndexingPipelines/blob/main/CPT_GAVS_pipeline_clean.ipynb [accessed 2025-11-04]
119. MIMIC-GAVS-EVAL/cleaned_icd10_gavs_mimic_eval.ipynb. GitHub. URL: https://github.com/sarvarip/MIMIC-GAVS-EVAL/blob/main/CLEANED_ICD10_GAVS_MIMIC_EVAL.ipynb [accessed 2025-11-04]
120. Mapping. International Classification of Diseases. URL: https://icdcdn.who.int/static/releasefiles/2024-01/mapping.zip [accessed 2025-11-04]
121. SNOMED drug codes. NHS TRUD. URL: https://isd.digital.nhs.uk/trud/users/authenticated/filters/0/categories/40/items/280/releases [accessed 2025-11-04]

Abbreviations

AI: artificial intelligence
CDS: clinical decision support
CPT: Current Procedural Terminology
DPO: data protection officer
DSPT: Data Security and Protection Toolkit
DTAC: Digital Technology Assessment Criteria
EHR: electronic health record
GARAG: generation-assisted retrieval-augmented generation
GAVS: generation-assisted vector search

HCPCS: Health Care Common Procedure Coding System
***ICD-10*:** *International Classification of Diseases, Tenth Revision*
***ICD-11*:** *International Classification of Diseases, Eleventh Revision*
ICO: Information Commissioner's Office
LLM: large language model
MHRA: Medicines and Healthcare Products Regulatory Agency
MIMIC: Medical Information Mart for Intensive Care
NHS: National Health Service
OPCS: Operating Procedure Codes Supplement
RAG: retrieval-augmented generation
SaMD: software as a medical device
SNOMED-CT: Systematized Nomenclature of Medicine–Clinical Terms

Edited by SAA Shah; submitted 20.09.24; peer-reviewed by A Gandra, F Shah-Mohammadi, H Wang; revised version received 28.09.25; accepted 05.10.25; published 10.11.25.

Please cite as:
Sarvari P, Al-fagih Z, Abou-Chedid A, Jewell P, Taylor R, Imtiaz A
Challenges and Solutions in Applying Large Language Models to Guideline-Based Management Planning and Automated Medical Coding in Health Care: Algorithm Development and Validation
JMIR Biomed Eng 2025;10:e66691
URL: https://biomedeng.jmir.org/2025/1/e66691
doi:10.2196/66691

Influence of Pre-Existing Pain on the Body's Response to External Pain Stimuli: Experimental Study

Burcu Ozek, PhD; Zhenyuan Lu, PhD; Srinivasan Radhakrishnan, PhD; Sagar Kamarthi, PhD

Mechanical and Industrial Engineering Department, Northeastern University, 360 Huntington Avenue, Boston, MA, United States

Corresponding Author:
Sagar Kamarthi, PhD
Mechanical and Industrial Engineering Department, Northeastern University, 360 Huntington Avenue, Boston, MA, United States

Abstract

Background: Accurately assessing pain severity is essential for effective pain treatment and desirable patient outcomes. In clinical settings, pain intensity assessment relies on self-reporting methods, which are subjective to individuals and impractical for noncommunicative or critically ill patients. Previous studies have attempted to measure pain objectively using physiological responses to an external pain stimulus, assuming that the participant is free of internal body pain. However, this approach does not reflect the situation in a clinical setting, where a patient subjected to an external pain stimulus may already be experiencing internal body pain.

Objective: This study investigates the hypothesis that an individual's physiological response to external pain varies in the presence of preexisting pain.

Methods: We recruited 39 healthy participants aged 22 - 37 years, including 23 female and 16 male participants. Physiological signals, electrodermal activity, and electromyography were recorded while participants were subject to a combination of preexisting heat pain and cold pain stimuli. Feature engineering methods were applied to extract time-series features, and statistical analysis using ANOVA was conducted to assess significance.

Results: We found that the preexisting pain influences the body's physiological responses to an external pain stimulus. Several features—particularly those related to temporal statistics, successive differences, and distributions—showed statistically significant variation across varying preexisting pain conditions, with *P* values <.05 depending on the feature and stimulus.

Conclusions: Our findings suggest that preexisting pain alters the body's physiological response to new pain stimuli, highlighting the importance of considering pain history in objective pain assessment models.

(JMIR Biomed Eng 2025;10:e70938) doi:10.2196/70938

KEYWORDS

pain measurement; sensors; physiological signals; hypothesis testing; pain assessment; ANOVA

Introduction

Accurate pain assessment is vital for ensuring proper treatment and helping patients receive the necessary care to reduce discomfort and prevent complications. Yet, current pain assessment tools and methods, which rely on patients' description of their pain using scales or descriptive measures, often fall short of clinical expectations [1]. These methods are ineffective for noncommunicative patients, such as infants or critically ill patients under sedation or mechanical ventilation. They are also inherently subjective, as pain perception varies widely between individuals [2-5]. These limitations increase the risk of misdiagnosis and mistreatment, highlighting the need for more objective and reliable pain assessment methods [6,7].

To address the limitations of self-reported pain assessments, physiological signals offer a promising alternative. Signals such as skin conductance, heart rate, and muscle activity provide objective data that can reflect the body's response to pain. Unlike self-reporting, physiological signals do not depend on a patient's ability to communicate, making them particularly suitable for critically ill or noncommunicative patients. By monitoring these signals in real-time, health care providers can gain an accurate and continuous understanding of a patient's pain levels, paving the way for timely and appropriate interventions. This shift toward objective, data-driven pain assessment can help reduce the variability and inaccuracies associated with traditional methods, enhancing health care providers' assessments [8,9].

Several studies have explored data-driven approaches for assessing pain through physiological signals [10-12]. These studies primarily collected data such as skin conductance, electromyography (EMG), electrocardiography, and electroencephalography during controlled pain stimuli experiments [9,13-15]. The BioVid Heat Pain Database is one of the most well-known, aiming to differentiate between various pain levels by analyzing physiological responses to heat pain

[9]. Other studies, like Rojas et al [16] and Lin et al [14], also gathered data from participants exposed to heat or cold stimuli, applying machine learning techniques to classify pain levels. These studies have demonstrated the potential of physiological signals for objective pain assessment and established valuable datasets for pain assessment research [9,14,17,18].

While the aforementioned studies provide promising results, they mainly focus on healthy participants responding to a single type of externally induced pain stimulus. One crucial factor that remains underexplored is the impact of preexisting conditions, such as chronic pain, postsurgical pain, or injury pain, that a patient is experiencing when the patient is administered an external pain stimulus. A few studies have investigated different patient populations, such as patients with chronic pain (back pain and shoulder pain) [11,19-22], patients in postsurgery [23], patients who are injured [24], patients with orthopedic trauma [25], patients with musculoskeletal trauma [26], and patients with cancer (eg, breast cancer [27]). These studies have provided insights into pain assessment in these populations, but they have not fully explored how preexisting pain interacts with new pain stimuli in terms of physiological responses.

Although the literature has begun exploring objective pain assessment for a single source of external pain stimuli, insights from medical research reveal that preexisting pain influences responses to new pain stimuli, underscoring the importance of considering preexisting pain. Sacco et al [20] found that individuals without chronic pain (without preexisting pain) exhibit an adaptive response to acute pain (new pain) by activating internal pain regulation mechanisms, including the release of natural painkillers and an increase in blood pressure, which temporarily reduces sensitivity. However, in patients with chronic pain, this adaptive mechanism can become disrupted, leading to heightened sensitivity to both acute and chronic pain. Similarly, Moscato et al [22] found that the autonomic signals of patients with chronic low back pain show differences compared to those of healthy individuals, both at rest and when subjected to a noxious stimulus, as evaluated through a set of physiological indicators. Lee et al [26] showed that preexisting pain can impact specific biomarkers, such as IL−1β, affecting how the body processes musculoskeletal trauma as a new pain. Raza et al [27] also found that women with chronic breast pain experienced more severe postoperative pain, highlighting preexisting pain as a predictor of adverse pain outcomes. In patients with trauma, Fetzh et al [24] observed that preexisting pain serves as a significant predictor for long-term pain following severe injury, emphasizing the complex interaction between pain history and physiological responses.

Although chronic pain is often referenced in the literature, the goal of this study is neither to simulate nor to assess chronic pain specifically. Instead, we use “preexisting pain” as a broader effect that can include various types of ongoing pain, such as postsurgical pain, injury-related pain, or other chronic and nonchronic conditions. Our aim is to investigate how any form of preexisting pain—regardless of origin—might influence the physiological response to a new external pain stimulus.

Our hypothesis is that preexisting pain significantly alters physiological responses to new pain stimuli. For instance, patients with chronic pain or postsurgical pain may show distinct physiological signals—such as changes in skin conductance or EMG—compared to healthy individuals when encountering new pain. To test this hypothesis, we conducted an experimental study examining how different levels of preexisting pain influence physiological responses to new pain stimuli. Understanding these responses could lead to accurate and personalized pain assessments.

In our experiments, we designated “heat pain” as a form of preexisting pain and “cold pain” as a new external stimulus. Heat pain and cold pain were studied at 3 levels: zero, low, and high. We conducted experiments with 9 combinations of no-heat, low-heat, high-heat, no-cold, low-cold, and high-cold pain. We recorded electrodermal activity (EDA) and EMG as time series data during these experimental conditions. Following data collection, we used feature engineering methods to extract features from these time series. We identified distributions, simple temporal statistics, linear and nonlinear autocorrelation, successive differences, and fluctuation analysis as pain-sensitive features. Next, we applied an ANOVA test to investigate whether physiological responses to cold pain stimuli exhibit statistical differences across three levels of preexisting heat pain. By analyzing variations in EDA and EMG features across different pain exposure levels, we aim to gain insights into how preexisting pain modulates the body’s response to new pain.

The aim of this study is to investigate how varying levels of preexisting heat pain affect the physiological response to new cold pain stimuli, using EDA and EMG signals as objective markers.

To our knowledge, this work represents the first experimental study that explores the EDA and EMG features that exhibit statistically significant differences across varying preexisting heat pain levels in response to an external stimulus.

Methods

Ethical Considerations

The research protocol was approved by the Northeastern University Institutional Review Board (IRB #22-11-06). The methods for this study adhered to the guidelines outlined in the Belmont Report. Northeastern University holds a Federal Wide Assurance with the US Department of Health & Human Services, ensuring our compliance with the principles of the Common Rule, 45 CFR 46. Before the experiment, the researcher orally explained the experimental procedure to each participant, the participant’s role, and other relevant information. In addition, the researcher presented each participant with a written consent form to read. The researcher obtained written informed consent from each participant before commencing the experiment. The research team kept participants’ data confidential and anonymized, securely storing all data with access limited to the research team only. No identifying information was included in the manuscript or any related materials. Participants were compensated with a gift card.

Participants

In total, 39 participants were recruited, with 31 completing the experiments. The remaining 8 participants chose not to continue the experiment due to discomfort from the heat pain. The study included 23 female and 16 male participants, with ages ranging from 22 to 37 years, with an average age of 26.1 (SD 3.57) years. All participants were healthy, and none reported experiencing pain before the experiment.

Inclusion and Exclusion Criteria

Participants were recruited from the Northeastern University community, including students, faculty, and staff. Inclusion criteria required participants to be between 18 and 50 years of age, in good general health, and not currently experiencing chronic pain or other medical conditions that could interfere with physiological responses. Only English-speaking individuals were included to ensure clear communication and understanding of study procedures. Pregnant individuals were excluded from participation to ensure their comfort and to avoid the introduction of additional physiological variability. There were no exclusion criteria related to gender, race or ethnicity, socioeconomic status, or literacy level.

Measured Physiological Signals

This study examined two physiological signals, EDA and EMG, to capture responses to pain stimuli.

Electrodermal Activity (EDA)

EDA serves as an indicator of neurocognitive stress through changes in the skin's electrical conductance [28]. Closely linked to the sympathetic branch of the autonomic nervous system, EDA can sense and transmit information about environmental changes, including temperature, pressure, and pain [29-31]. Consequently, EDA reflects emotional and cognitive states, making it a valuable physiological marker across various applications [32].

During emotional arousal or cognitive stress, sweat gland stimulation induces fluctuations in skin conductance, measured by EDA. These changes, largely beyond conscious control, capture subconscious physiological responses to emotions and stress, providing an objective means of assessing an individual's state [33].

In pain assessment, EDA plays a crucial role by offering a quantitative and objective measure of physiological responses to pain. It provides valuable insights into pain intensity, complementing self-reporting to enhance pain assessment accuracy in research and clinical settings [28,34]. EDA encompasses data related to both slow shifts (tonic component) and the signal's rapid alterations (phasic changes). Our analysis focused on gathering information from the tonic component, specifically skin conductance level.

Electromyography (EMG)

EMG is the electrical signal produced by skeletal muscle activity. These signals originate from motor neurons, which are integral components of the central nervous system. Since EMG signals are a reflection of neuromuscular activity, they find application in the diagnosis of conditions such as muscle injuries, nerve damage, and muscle dysfunction arising from neurological and muscular disorders [35-37]. EMG is an excellent choice for developing an objective pain assessment tool because of its unique ability to measure muscle activity directly. It allows real-time monitoring of muscle responses to understand pain intensity, location, and characteristics [14,38,39].

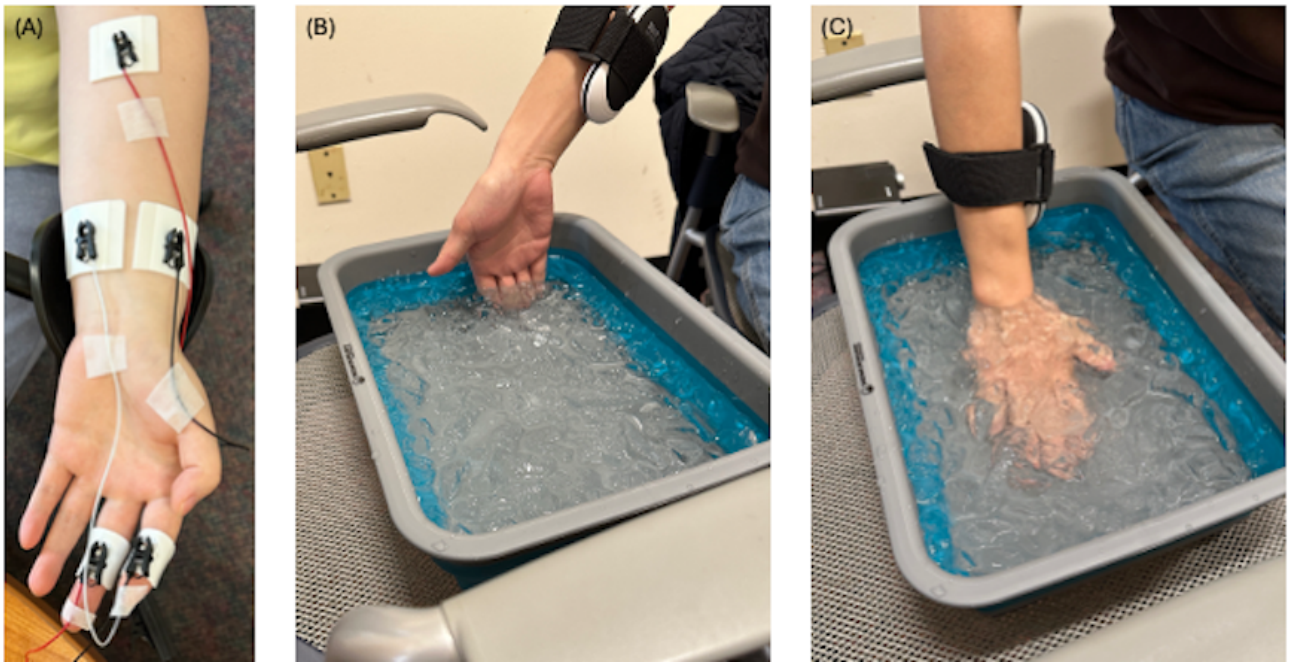
Design of the Experiment

The physiological data were collected using the BIOPAC MP160 data acquisition and analysis systems with AcqKnowledge software (BIOPAC Systems, Inc). Smart amplifiers recorded EMG and EDA. Heat stimulation was delivered using OCOOPA Hand Warmers, which offered two temperature settings: 37 and 45 °C. These temperatures were measured and monitored using a BIOPAC SKT (Skin Temperature) Smart Amplifier. Cold stimulation was provided through iced water, with the temperature continuously monitored using a thermometer. In these experiments, heat pain acts as preexisting pain, while cold pain acts as a new pain stimulus.

Using temperature-based modalities for both preexisting (heat) and new (cold) pain stimuli allowed us to design a consistent, safe, and replicable experimental setup. Temperature stimuli are well-established in pain research and offer practical advantages regarding ecological validity and participant safety. Moreover, the thermal approach enabled controlled comparisons of physiological responses across different pain levels while minimizing variability introduced by mechanical or electrical alternatives.

EDA data were collected using the BIOPAC EDA Smart Amplifier attached to the ring and index fingers of the participant's nondominant hand. Before attaching the sensors to the fingers, the skin was cleaned with wet wipes, and GEL101A was applied to the electrodes to improve conductivity, enhance signal quality, and reduce impedance. EMG data were acquired using the BIOPAC EMG Smart Amplifier, with three electrodes attached to the participant's nondominant forearm. The skin in the sensor placement area was prepared by cleaning it with wet wipes, followed by abrasion and application of ELPREP. GEL100 was applied to the electrodes to improve contact. To minimize motion artifacts, all cables were secured with medical tape. Hand warmers were fastened to the participant's dominant forearm using a strap. Figure 1A shows the picture of the placement of the electrodes.

Figure 1. Data acquisition setup and experimental setup for pain stimuli. (A) EDA data were gathered from the ring and index fingers of the participant's nondominant hand, while EMG data were recorded using three electrodes positioned on the participant's nondominant forearm. (B,C) Hand warmers, serving as heat pain stimuli, were fastened to the participant's dominant forearm using a strap. Cold pain stimuli were induced by iced water when participants placed their fingers or hands in the iced water, depending on the stimulus level the participant is expected to receive in the design of experiments: for low-level cold pain stimulus, participants placed fingers in the iced water, and for high-level cold pain stimulus, participants placed the hand in the iced water.



The experiment consisted of two types of pain stimuli: (1) heat pain caused by attaching hand warmers to the forearm and (2) cold pain induced by placing fingers or hands in ice water. Each type of pain had low and high levels. The heat and cold pain stimuli were applied to the dominant hand, while physiological signals were collected from the nondominant hand. At the end of each step, participants were asked to report their pain levels on a scale of 0 to 10. The participants are given a 4-minute relaxation break at the beginning of each data collection session.

We collected baseline data from each participant without inducing any type of pain stimulus. The rest of the experimental procedure consisted of two phases. In the first phase, we collected data from four steps; in Step 1, only the low-level cold pain was applied; in Step 2, only the high-level cold was applied; in Step 3, only low-level heat pain was applied; and in Step 4, only high-level heat pain was applied. In the second phase of the experiments, we applied a different combination of heat and cold pain levels to examine their combined effect in Steps 5 through 8.

The experimental procedures for the first phase involved four steps. First, the participant placed their fingers in iced water and held them there for 8 seconds, representing low-level cold pain. In the second step, they placed their dominant hand in iced water for 8 seconds, representing high-level cold pain. In the third step, using a hand warmer attached to the participant's dominant forearm, they were subjected to 37 °C heat for 1.5 minutes, which caused low-level preexisting heat pain. In the final step of the first phase, using a hand warmer attached to the participant's dominant forearm, they were subjected to 45 °C heat for 1.5 minutes, which caused high-level preexisting heat pain.

The second phase of the experiment involved four additional steps. In the fifth step of the experiment, the participant wore a hand warmer on their nondominant forearm, experiencing a temperature of 37 °C for 1.5 minutes. After 80 seconds into the heat pain stimulus, the participant placed their fingers in iced water for 8 seconds. This scenario represents the simultaneous application of low preexisting heat pain and new low cold pain. In the sixth step, the participant repeated Step 5 with the hand warmer on their nondominant forearm, but at a temperature of 45 °C. Again, after 80 seconds, they placed their fingers in iced water for 8 seconds. This scenario represents the simultaneous application of high preexisting heat pain and new low cold pain. In the seventh step, the participant wore the hand warmer on their nondominant forearm at 37 °C for 1.5 minutes. After 80 seconds had elapsed, they immersed their dominant hand in iced water for 8 seconds. This scenario represents the simultaneous application of low preexisting heat pain and new high cold pain. In the eighth and final step, the participant repeated Step 7 with the hand warmer on their nondominant forearm at 45 °C for 1.5 minutes. After 80 seconds, they immersed their dominant hand in iced water for 8 seconds. This scenario represents the simultaneous application of high preexisting heat pain and new high cold pain.

Figure 1B,C illustrates how the hand warmer is positioned on the forearm and how the fingers or hand are placed in the ice water.

Signal Processing

Both EDA and EMG signals were recorded at a data acquisition rate of 2000 samples per second (2 kHz). For EDA, a low-pass filter with a 1.0 Hz frequency cutoff was used to eliminate high-frequency noise [30,40].

We processed EMG signals through a comb bandstop transformation to eliminate interference from the power line frequency (50 Hz) [41]. The comb bandstop transformation aims to effectively suppress or eliminate interference originating from the power line frequency (50 Hz), ensuring a relatively noise-free EMG signal for analysis and interpretation. Subsequently, a finite impulse response bandpass filter was applied, specifying a low-frequency cutoff at 28 Hz and a high-frequency cutoff at 500 Hz [42]. This step was implemented to filter out both high and low artifacts, such as motion artifacts, and to focus on the EMG signal within the frequency range of 28 to 500 Hz.

Recognizing that the EMG signal centers around 0, a rectified version was generated by averaging samples in sets of 100. This approach makes analysis easy by eliminating negative values and retaining the magnitude of the signal.

To analyze EMG further, the root mean square (RMS) was calculated using a window size of 100 samples. This measurement meaningfully represents the signal's characteristics because EMG is centered around 0.

Feature Extraction

In this study, we derived features from EDA and EMG using the "Canonical Time-series Characteristics" outlined by Lubba et al [43]. These features encompass fundamental statistical metrics of time-series data, stationarity measures, entropy, linear correlations, nonlinear time-series analysis techniques, linear and nonlinear model parameters, predictive capabilities, and fits. Specifically, we identified the subset of 22 features highlighted as the most informative by Lubba et al [43]. These features are listed in Table 1. Following all the data processing and extraction steps, we obtained 22 features from EDA, EMG, rectified EMG, and RMS of EMG signals; this resulted in a total of 22×4=88 features. Then, we applied z-transformation to normalize all features for each participant, using the participant-specific mean and SD.

Table . Time-series feature categories and descriptions using the "Canonical Time-series Characteristics" defined by Lubba et al [43].

Feature category	Features
Distribution	• Mode of *z*-scored distribution (5-bin histogram) • Mode of *z*-scored distribution (10-bin histogram)
Simple temporal statistics	• The longest period of consecutive values above the mean • Time intervals between successive extreme events above the mean • Time intervals between successive extreme events below the mean
Linear autocorrelation	• The first 1/e crossing of the autocorrelation function • The first minimum of the autocorrelation function • Total power in the lowest fifth of frequencies in the Fourier power spectrum • Centroid of the Fourier power spectrum • Mean error from a rolling 3-sample mean forecasting
Nonlinear autocorrelation	• Time-reversibility statistic, $(x_{t+1}-x_t)^3{}_t$ • Auto mutual information, $m=2$, $\tau=5$ • The first minimum of the auto-mutual information function
Successive differences	• Proportion of successive differences exceeding 0.04 σ (Mietus et al [44]) • The longest period of successive incremental decreases • Shannon entropy of two successive letters in equiprobable 3-letter symbolization • Change in correlation length after iterative differencing • Exponential fit to successive distances in 2D embedding space
Fluctuation analysis	• The proportion of slower timescale fluctuations that scale with DFA[a] (50% sampling) • The proportion of slower timescale fluctuations that scale with linearly rescaled range fits
Others	• Trace of covariance of the transition matrix between symbols in the 3-letter alphabet • Periodicity measure (Wang et al [45])

[a]DFA: detrended fluctuation analysis.

Statistical Testing

The initial analysis aims to identify statistically significant features for class differentiation. This includes using the ANOVA test, which assesses variations among the means of various groups. It is applied in various situations to ascertain whether there are any significant differences between the means of the groups [46,47]. The null hypothesis asserts that the means

of the groups are the same, while the alternative hypothesis posits that the means are not equal.

(1)H0:μ1=μ2H1:μ1≠μ2

We reject the null hypothesis if the calculated *P* value is less than the chosen significance level, say, .05.

We used ANOVA to assess the statistical differences in the means of extracted time series features derived from physiological signals. The sample comprises 31 observations. The normality of data, which is a requisite for ANOVA, is confirmed through the Kolmogorov-Smirnov Test for normality of data and examination of quantile-quantile plots (Q-Q plots) for each individual feature. A significance level of .05 is set for the ANOVA test, which is conducted as a 2-tailed analysis.

Results

The following sections present the results of statistical comparisons of EMG and EDA signal features across different combinations of heat and cold pain levels.

Significant Features in the Presence and Absence of Pre-existing Pain

Table 2 summarizes the statistically significant differences (*P*<.05) in EMG and EDA features across experimental groups. Each row corresponds to a specific hypothesis involving two groups. For example, the first row compares Group 1 (participants who experienced low-level cold pain without preexisting heat pain) with Group 2 (participants who experienced the same low-level cold pain while also experiencing mild preexisting heat pain). This comparison examines feature-level differences across EMG and EDA signals under these two conditions.

Table . Statistically significant feature categories and the average *P* values of features within each category for different hypotheses, aiming to study the influence of the presence or absence of pre-existing pain on external pain stimuli between symbols in the 3-letter set.

Groups	EMG,[a] (*P* value)	RMS[b] of EMG, (*P* value)	Rectified EMG, (*P* value)	EDA,[c] (*P* value)
• Group 1: low-level cold pain without any pre-existing pain • Group 2: low-level cold pain with mild pre-existing heat pain	• Linear autocorrelation (<.001) • Successive differences (.002) • Distribution (.02) • Others (.006) • Statistics (.02)	• Linear autocorrelation (.004) • Successive differences (.01)	• Linear autocorrelation (.003) • Successive differences (.001)	• Statistics (.02)
• Group 1: low-level cold pain without any pre-existing pain • Group 2: low-level cold pain with severe pre-existing heat pain	• Linear autocorrelation (.004) • Successive differences (.005) • Others (.02) • Statistics (.02)	• Linear autocorrelation (.02)	• Linear autocorrelation (.02) • Successive differences (.03)	• Statistics (.02) • Others (.04)
• Group 1: high-level cold pain without any pre-existing pain • Group 2: high-level cold pain with mild pre-existing heat pain	• No significant features	• No significant features	• No significant features	• Successive differences (.03)
• Group 1: high-level cold pain without any pre-existing pain • Group 2: high-level cold pain with severe pre-existing heat pain	• No significant features	• No significant features	• Successive differences (.03)	• Others (.03)

[a]EMG: electromyography.

[b]RMS: root mean square.

[c]EDA: electrodermal activity.

For low-level cold pain without any pre-existing pain (Group 1) versus low-level cold pain with mild pre-existing heat pain (Group 2), significant differences were observed in EMG features related to linear autocorrelation, including the "first minimum and the first 1/e crossing of the autocorrelation function." In the EDA signal, temporal statistics, specifically "time intervals between successive extreme events," showed statistically significant differences.

For low-level cold pain without any pre-existing pain (Group 1) versus low-level cold pain with severe pre-existing heat pain (Group 2), EMG features related to linear autocorrelation, such as the first minimum and 1/e crossing of the autocorrelation function, were significantly different. The EDA features that

showed the differences included “time intervals between successive extreme events” and the “longest period of consecutive values above the mean.”

For high-level cold pain without any pre-existing pain (Group 1) versus high-level cold pain with mild pre-existing heat pain (Group 2), the distinguishing features were found in the EDA signal’s successive differences, particularly the “longest period of successive incremental decreases.”

For high-level cold pain without any pre-existing pain (Group 1) versus high-level cold pain with severe pre-existing heat pain (Group 2), statistically significant differences were observed in the rectified EMG signal for features related to successive differences, including the “change in correlation length after iterative differencing” and the “longest period of successive incremental decreases.” In the EDA signal, differences were observed in the “trace of covariance of the transition matrix between symbols in the 3-letter set.”

Significant Features in the Mild and Severe Cases of Pre-existing Pain

Table 3 presents the signals and their respective features that exhibit statistically significant differences ($P<.05$) among the groups. In this section, two hypotheses are investigated. The first hypothesis aims to compare physiological signals to assess the influence of mild and severe pre-existing pain in Groups 1 and 2; Group 1 includes signals from participants subjected to low-level cold pain while already experiencing mild pre-existing heat pain; Group 2 includes signals from participants subjected to low-level cold pain while already experiencing severe pre-existing heat pain. The second hypothesis involves comparing the groups to assess the impact of mild and severe pre-existing heat pain on participants when they are subjected to high-level cold pain. Figure 2 visually illustrates the distribution of the most statistically significant features for each of the two hypotheses.

Table . Statistically significant feature categories and individual features for distinguishing the influence of varying levels of pre-existing pain on the response to low and high levels of cold pain.

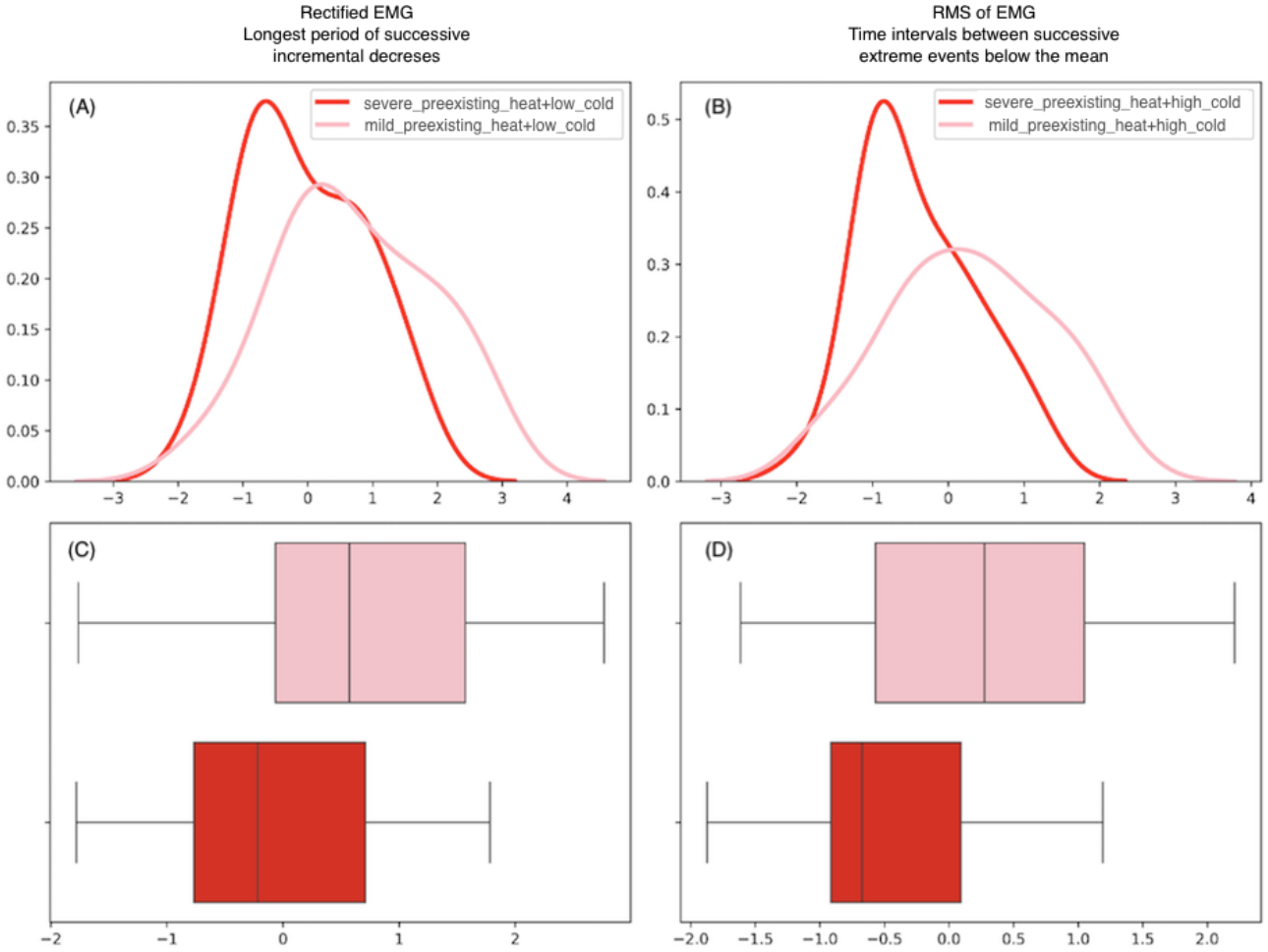
Hypotheses and signal	Feature category	Feature (*P* value)
Group 1: low-level cold pain with mild pre-existing heat pain; Group 2: low-level cold pain with severe pre-existing heat pain		
EMG[a]	Distribution	Mode of *z*-scored distribution (10-bin histogram; .03)
RMS[b] of EMG	Successive differences	Longest period of successive incremental decreases (.01)
RMS of EMG	Statistics	Longest period of consecutive values above the mean (.03)
Group 1: high-level cold pain with mild pre-existing heat pain; Group 2: high-level cold pain with severe pre-existing heat pain		
Rectified EMG	Successive differences	Longest period of successive incremental decreases (.007)
Rectified EMG	Statistics	Longest period of consecutive values above the mean (.01)
EMG	Distribution	Mode of *z*-scored distribution (10-bin histogram; .03)
EMG	Statistics	Time intervals between successive extreme events below the mean (.04)
RMS of EMG	Statistics	Time intervals between successive extreme events below the mean (.005)
RMS of EMG	Statistics	Time intervals between successive extreme events above the mean (.009)
Rectified EMG	Statistics	Time intervals between successive extreme events below the mean (.01)
Rectified EMG	Statistics	Time intervals between successive extreme events above the mean (.01)
Rectified EMG	Successive differences	Change in correlation length after iterative differencing (.03)
EDA[c]	Statistics	Time intervals between successive extreme events below the mean (.01)
EDA	Others	Trace of covariance of transition matrix (.02)

[a]EMG: electromyography.

[b]RMS: root mean square.

[c]EDA: electrodermal activity.

Figure 2. Distribution of features with the influence of pre-existing heat pain: (A,B) Illustrate the probability density of two significant EMG features under low- and high-level cold pain conditions. (C,D) Present the corresponding boxplots for each feature, comparing the mild and severe pre-existing pain conditions. EMG: electromyography; RMS: root mean square.



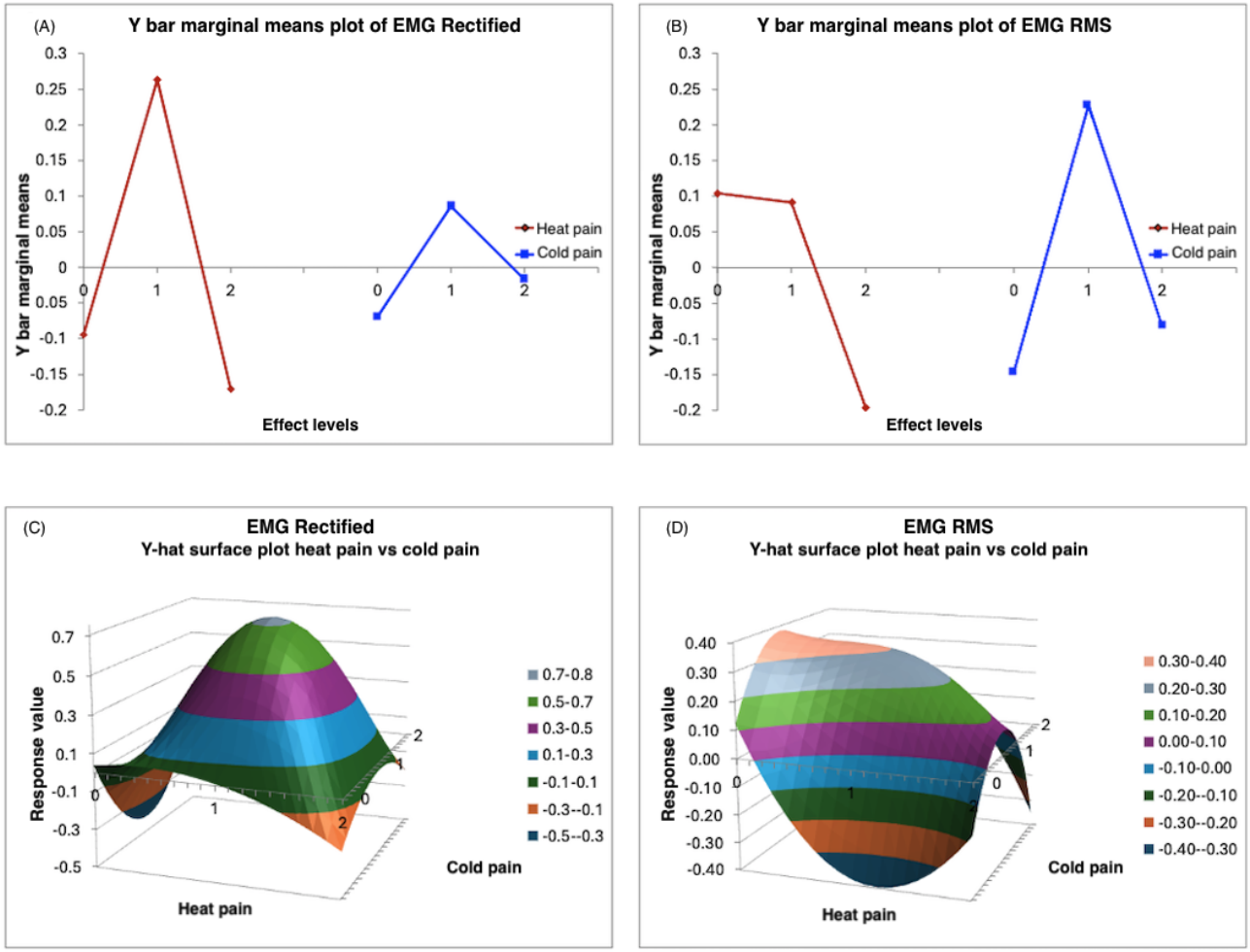
Hypothesis 1 examines the influence of mild and severe pre-existing heat pain on the body's response to low-level cold pain. Significant differences were observed in the EMG signal's "mode of *z*-scored distribution." RMS of EMG showed variations in successive differences and statistics, specifically related to "the longest period of incremental decreases" and "the longest period of consecutive values above the mean." Similar patterns were found in the rectified EMG signal.

Hypothesis 2 investigates the influence of mild and severe pre-existing heat pain on the body's response to high-level cold pain. The "mode of *z*-scored distribution" of EMG exhibited significant differences across the groups. RMS of EMG also showed variations in statistics related to "time intervals between successive extreme events below and above the mean." Rectified EMG signals differed in features pertaining to successive differences and statistics. Additionally, EDA signals showed significant differences in the "trace of covariance of the transition matrix."

Heat and Cold Pain Interactions

This section presents a response surface analysis using marginal mean plots and surface plots. It examines how varying levels of heat and cold pain affect two statistically significant features: the rectified EMG's "longest period of successive incremental decreases" and the RMS of EMG's "time intervals between successive extreme events below the mean." Figure 3A and C show the rectified EMG response values, while Figure 3B and D display the RMS of EMG response values. The analysis includes pain levels coded as 0 (no pain), 1 (mild pain), and 2 (severe pain).

Figure 3. Interaction effects of pre-existing heat pain and new cold pain on EMG features. (A,B) Marginal means plots illustrating how rectified EMG and RMS values vary across different levels of heat pain (0, 1, 2) and cold pain (0, 1, 2). (C,D) Surface plots depicting the variation in responses under various combinations of heat and cold pain levels. EMG: electromyography; RMS: root mean square.



Discussion

Principal Findings

This study was guided by the hypotheses that (1) pre-existing body pain alters the physiological response to a new pain stimulus relative to the physiological response in the absence of pre-existing pain, and (2) pre-existing pain of different intensities produces distinguishably different physiological patterns in response to a new pain stimulus. The use of multimodal physiological signals, EDA, and EMG provides insight into the underlying mechanisms and supports the potential for objective, signal-based pain assessment in complex pain scenarios.

This study found that pre-existing heat pain significantly influences physiological responses to new cold pain stimuli, as indicated by features from EDA and EMG, particularly successive differences, temporal statistics, and distribution features, demonstrating noticeable sensitivity to varying pain combinations.

Comparative Analysis: Features Sensitive to the Presence and Absence of Pre-existing Pain

This section evaluates how the presence or absence of pre-existing heat pain influences physiological responses when the body encounters a cold pain stimulus. EMG signals exhibit significant variation across groups, particularly in features such as “linear autocorrelation” and “successive differences,” while EDA signals indicate differences through statistical features.

When an external low-level cold pain is applied, the pre-existing heat pain, mild or severe, consistently leads to marked changes in both EMG and EDA signals. In the EMG signal, features such as the “first minimum and the first 1/e crossing of the autocorrelation function” capture distinct temporal characteristics of muscle activity. The first minimum identifies a key point of dissimilarity, while the 1/e crossing reflects the timescale at which the signal’s autocorrelation declines to approximately 36.8% of its peak. In the EDA signal, the “time intervals between successive extreme events” and the “longest period of consecutive values above the mean” emerge as distinguishing features. These results indicate that low-level cold pain elicits prominently different physiological features in the presence and absence of pre-existing heat pain.

When a high-level cold pain stimulus is applied, the EDA signal’s sensitivity to successive differences, particularly the “longest period of successive incremental decreases,” emerges as a distinguishing feature. This feature identifies continuous patterns where EDA consistently decreases from one point to the next and the trace of covariance of the “transition matrix

between symbols in the 3-letter set." This method involves encoding and simplifying the EDA signal into sequences, allowing for the analysis of how these sequences change and relate to each other over time, highlighting its utility in capturing autonomic dynamics influenced by layered pain conditions. Similarly, rectified EMG features tied to successive differences are important: the "change in correlation length after iterative differencing" and the "longest period of successive incremental decreases" further underscore the complementary roles of multimodal physiological measurements.

High-intensity cold pain appears to overshadow the physiological responses associated with pre-existing heat pain. Under these conditions, significant differences are limited and primarily observed in EDA and rectified EMG signals. The overwhelming nature of high-level cold pain reduces the detectability of pre-existing pain effects, making it difficult to distinguish their individual contributions to the physiological response. Despite this, certain features remain sensitive. In the EDA signal, successive differences, particularly the "longest period of successive incremental decreases," identify continuous patterns where EDA consistently decreases from one point to the next. Additionally, the "trace of covariance of the transition matrix between symbols in the 3-letter" set captures how patterns evolve over time, offering insights into autonomic dynamics under layered pain conditions. Similarly, rectified EMG features related to successive differences, including the "change in correlation length after iterative differencing" and the "longest period of successive decreases," emphasize the value of combining multimodal physiological measurements to capture subtle effects that may persist despite dominant pain stimuli.

Together, these findings suggest that the influence of pre-existing heat pain on the body's physiological response is more discernible when cold pain is mild, particularly through EMG and EDA signals. In contrast, high-intensity cold pain may mask these effects, making it difficult to detect the physiological changes due to pre-existing pain. Understanding these interactions between physiological responses due to external and pre-existing pains is essential for interpreting pain states in complex and overlapping pain scenarios. The presence of statistically significant and diverse features supports the notion that pre-existing heat pain has a measurable impact on physiological responses.

Comparative Analysis: Significant Features in the Mild and Severe Cases of Pre-existing Pain

This section examines how the severity of pre-existing heat pain, ranging from mild to severe, influences the body's physiological response when exposed to a new cold pain stimulus. The findings reveal distinct alterations in EMG and EDA signals that differentiate these pain intensities.

When participants experience low-level cold pain while the body is already encountering varying degrees of pre-existing heat pain, the physiological responses captured through EMG are particularly sensitive to the severity of pre-existing heat pain. Features like "mode of *z*-score distribution," which refers to the value or range of values that occur most frequently, indicate shifts in the most dominant muscle activity patterns. Additionally, the RMS of EMG shows differences in features related to successive differences and statistics, specifically the "longest period of incremental decreases" and the "longest period of consecutive values above the mean." The first feature refers to the duration in the time series where the EMG signal's RMS consistently decreases incrementally. In simpler terms, it identifies the most extended continuous period during which the RMS values decrease step by step. The second feature pertains to the time series duration in which the EMG signal's RMS values remain consistently above the mean. This duration captures the longest continuous segment where the RMS values are consistently higher than the average. These features are further supported by similar patterns observed in the rectified EMG signal, reinforcing the robustness of these distinctions.

In the high-level cold pain condition, EMG signals continue to reveal statistically significant differences across pre-existing pain intensities. The "mode of the *z*-scored distribution" emerges as an important marker, indicating distinctive patterns in muscle activity under mild and severe pre-existing heat pain conditions. Analysis of the RMS of EMG signals unveils notable variations in statistics involving "time intervals between successive extreme events above or below the mean." This observation highlights the complex temporal dynamics associated with the interaction of high-level cold pain and the severity of pre-existing pain. The distribution of rectified EMG signals further reinforces these findings, highlighting distinct patterns in successive differences and statistics, which contribute to the differentiation of the influence of different pre-existing pain conditions. Beyond EMG, EDA signals also contribute to this differentiation. The "trace of the covariance of the transition matrix" emerges as a key feature. This reveals variation in how these patterns evolve over time under different pre-existing pain conditions. The inclusion of EDA signals in our analysis deepens our understanding of physiological responses to the influence of varying pre-existing pain intensities.

The results of this study highlight that the body exhibits distinct responses to cold pain stimuli when experiencing mild versus severe pre-existing heat pain. These findings highlight the intricate relationship between pain conditions and physiological responses. The identified features within EMG and EDA signals offer valuable insights into the body's mechanisms, highlighting the influence of pre-existing pain on physiological signals.

Analysis of Heat and Cold Pain Interactions

Response surface analysis provides comprehensive insight into how varying levels of pre-existing heat pain and externally introduced cold pain interact to influence physiological responses.

Figure 3A displays the marginal means plot of the rectified EMG response. Marginal means plots illustrate the responses by considering only the level of one type of pain, independent of the levels of any other type of pain. For example, the response is the strongest when the heat pain is mild. Similarly, when considering only cold pain, the response peaks again at the mildest level of pain. The surface plot represents the interactions between the two types of pain and their effects on the response. The surface plot in Figure 3Creveals a convex shape with a peak, indicating that the rectified EMG responses reach their

highest values when both heat and cold pain are at a mild level. The plot shows that the response is low when there is no heat pain and mild cold pain, and similarly low under severe heat pain with no accompanying cold pain.

Figure 3B presents a marginal mean plot of the RMS of EMG responses. Here, it is evident that mild cold levels yield the highest response values. Both “no heat” pain and “mild heat” pain conditions result in high response values, while severe heat pain significantly reduces the RMS of EMG responses. In Figure 3D, surface plots of the RMS of EMG responses are displayed.

Notably, as heat pain increases, the RMS value decreases, reaching its peak when heat pain is absent or mild. Conversely, instances of mild heat paired with no cold pain result in the lowest RMS of EMG response values.

These patterns underscore the importance of considering multidimensional pain contexts, as overlapping pain experiences can interact in nonintuitive ways that meaningfully alter physiological signatures.

Limitations

The relatively small and homogeneous sample size, consisting of 31 healthy young adults aged 22 to 37 years, is one of the shortcomings of this study. This may limit the generalizability of the findings to broader and clinically relevant populations. Additionally, the study was conducted in a controlled laboratory environment, which may not fully replicate real-world clinical settings, thus limiting its ecological validity. While the use of fixed-intensity heat and cold stimuli was effective for controlled experimental design, it may not capture the full complexity of pre-existing pain conditions and individual pain thresholds. Furthermore, the fixed-intensity nature of these stimuli does not account for interindividual variability in pain sensitivity, which could influence physiological responses. The devices used in this study did not support personalized stimulus calibration, which we recognize as a limitation.

Future Directions

Future work is open to expanding the sample population to include individuals from diverse age groups and clinical backgrounds, particularly those experiencing chronic or postsurgical pain, to improve the generalizability of findings. Validation in real-world clinical environments is also crucial for enhancing ecological validity. To better reflect the complexity of pain experiences, future studies should explore alternative or multimodal pain induction methods beyond heat and cold stimuli and incorporate personalized calibration to account for individual pain thresholds. Additionally, expanding the range of physiological signals—such as heart rate variability, electroencephalography, and functional neuroimaging—may offer a more comprehensive understanding of the neural and autonomic correlates of pain.

This study used statistical analysis to examine the significance of physiological differences across pain conditions. In future work, we will further explore machine learning models to analyze physiological responses to new external pain stimuli. This approach will enable us to assess the intensity of pre-existing pain caused by chronic conditions, injuries, or surgeries. By integrating machine learning, we aim to develop predictive models that can objectively assess pain intensity and support personalized, effective pain management, particularly in clinical settings where patients are unable to verbally communicate their pain levels.

Conclusions

Accurate pain assessment is crucial for the correct diagnosis and effective treatment of many diseases. While existing literature has developed tools for estimating pain levels based on physiological responses, these studies often focus on healthy individuals experiencing acute pain, overlooking the potential influence of pre-existing conditions, such as postsurgical pain, chronic pain, and physical discomfort, on the physiological signals triggered by acute pain. Acknowledging this factor is essential, as individuals may respond differently to new pain stimuli depending on the intensity of their pre-existing pain.

This study examined the impact of pre-existing heat pain through experimental research when participants were exposed to cold pain stimuli. We used heat pain as the pre-existing pain condition, cold pain as the new pain stimulus, and EMG and EDA as physiological signals. By using statistical tests, we observed significant differences in specific EDA and EMG signal features across varying levels of pre-existing heat pain and new cold pain combinations. Notably, simple temporal statistics (the most extended period of consecutive values, time intervals between successive extreme events), successive differences (change in correlation length after iterative differencing), distribution (mode of *z*-scored distribution), and autocorrelation (the first 1/e crossing of the autocorrelation function) emerged as primary feature categories that significantly varied across pre-existing heat pain and new cold pain intensity combinations.

Our investigation into the differences in EMG and EDA signals in the presence of different levels of pre-existing heat pain has revealed valuable insights. The distinction between the absence of pre-existing pain and the presence of mild or severe pre-existing heat pain, particularly when stimulated with new low-level cold pain, highlighted statistically significant differences in both EMG and EDA signals. Notably, when we switched to high-level cold pain, EDA emerged as a more reliable indicator of variation in pre-existing pain than EMG. During high-level cold pain, the time series features of “successive differences” proved to be effective indicators of the level of the pre-existing pain. Furthermore, our analysis of mild and severe pre-existing heat pain scenarios revealed that EMG exhibited statistically significant differences, particularly in response to the new low-level cold pain, whereas EDA remained relatively unchanged. However, when we switched to high-level cold pain, both EMG and EDA signal features exhibited statistically significant differences. Successive difference, temporal statistics, and distribution features of time series emerged as reliable indicators of the pre-existing heat pain in these cases. These findings shed light on the changes in EMG and EDA signals across different levels of pre-existing pain, advancing our understanding of physiological responses in pain assessment.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

BO, ZL, SR, and SK designed the experiment. BO conducted the experiments, analyzed the data, and prepared the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Huo J, Yu Y, Lin W, Hu A, Wu C. Application of AI in multilevel pain assessment using facial images: systematic review and meta-analysis. J Med Internet Res 2024 Apr 12;26:e51250. [doi: 10.2196/51250] [Medline: 38607660]
2. Liu R, Gutiérrez R, Mather RV, et al. Development and prospective validation of postoperative pain prediction from preoperative EHR data using attention-based set embeddings. NPJ Digital Med 2023 Nov 16;6(1):209. [doi: 10.1038/s41746-023-00947-z] [Medline: 37973817]
3. Ozek B, Lu Z, Radhakrishnan S, Kamarthi S. Uncertainty quantification in neural-network based pain intensity estimation. PLoS One 2024;19(8):e0307970. [doi: 10.1371/journal.pone.0307970] [Medline: 39088473]
4. Pouromran F, Radhakrishnan S, Kamarthi S. Exploration of physiological sensors, features, and machine learning models for pain intensity estimation. PLoS One 2021;16(7):e0254108. [doi: 10.1371/journal.pone.0254108] [Medline: 34242325]
5. Amidei J, Nieto R, Kaltenbrunner A, Ferreira De Sá JG, Serrat M, Albajes K. Exploring the capacity of large language models to assess the chronic pain experience: algorithm development and validation. J Med Internet Res 2025 Mar 31;27:e65903. [doi: 10.2196/65903] [Medline: 40163858]
6. Lu Z, Ozek B, Kamarthi S. Transformer encoder with multiscale deep learning for pain classification using physiological signals. Front Physiol 2023;14:1294577. [doi: 10.3389/fphys.2023.1294577] [Medline: 38124717]
7. Billot M, Ounajim A, Moens M, et al. The added value of digital body chart pain surface assessment as an objective biomarker: multicohort study. J Med Internet Res 2025 Apr 16;27:e62786. [doi: 10.2196/62786] [Medline: 40239206]
8. Ozek B, Lu Z, Pouromran F, Radhakrishnan S, Kamarthi S. Analysis of pain research literature through keyword co-occurrence networks. PLOS Digital Health 2023 Sep;2(9):e0000331. [doi: 10.1371/journal.pdig.0000331] [Medline: 37676880]
9. Walter S, Gruss S, Ehleiter H, et al. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. Presented at: 2013 IEEE International Conference on Cybernetics (CYBCO); Jun 13-15, 2013; Lausanne, Switzerland p. 128-131. [doi: 10.1109/CYBConf.2013.6617456]
10. Jiang M, Mieronkoski R, Syrjälä E, et al. Acute pain intensity monitoring with the classification of multiple physiological parameters. J Clin Monit Comput 2019 Jun;33(3):493-507. [doi: 10.1007/s10877-018-0174-8] [Medline: 29946994]
11. Luebke L, Gouverneur P, Szikszay TM, Adamczyk WM, Luedtke K, Grzegorzek M. Objective measurement of subjective pain perception with autonomic body reactions in healthy subjects and chronic back pain patients: an experimental heat pain study. Sensors (Basel) 2023 Oct 3;23(19):8231. [doi: 10.3390/s23198231] [Medline: 37837061]
12. Lancaster J, Mano H, Callan D, Kawato M, Seymour B. Decoding acute pain with combined eeg and physiological data. Presented at: 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER); May 25-28, 2017; Shanghai, China. [doi: 10.1109/NER.2017.8008404]
13. Guo Y, Wang L, Xiao Y, Lin Y. A personalized spatial-temporal cold pain intensity estimation model based on facial expression. IEEE J Transl Eng Health Med 2021;9:4901008. [doi: 10.1109/JTEHM.2021.3116867] [Medline: 34650836]
14. Lin Y, Xiao Y, Wang L, et al. Experimental exploration of objective human pain assessment using multimodal sensing signals. Front Neurosci 2022;16:831627. [doi: 10.3389/fnins.2022.831627] [Medline: 35221908]
15. Chu† Y, Zhao X, Yao† J, Zhao Y, Wu Z. Physiological signals based quantitative evaluation method of the pain. IFAC Proc Volumes 2014;47(3):2981-2986. [doi: 10.3182/20140824-6-ZA-1003.01420]
16. Rojas RF, Huang X, Ou KL. A machine learning approach for the identification of a biomarker of human pain using fNIRS. Sci Rep 2019 Apr 4;9(1):5645. [doi: 10.1038/s41598-019-42098-w] [Medline: 30948760]
17. Posada-Quintero HF, Kong Y, Chon KH. Objective pain stimulation intensity and pain sensation assessment using machine learning classification and regression based on electrodermal activity. Am J Physiol Regul Integr Comp Physiol 2021 Aug 1;321(2):R186-R196. [doi: 10.1152/ajpregu.00094.2021] [Medline: 34133246]
18. Ghita M, Ghita M, Copot D, Ionescu CM. Methodologically study for detection of thermal induced pain via skin impedance. : IEEE Presented at: 2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI); Jan 24-26, 2019; Herlany, Slovakia. [doi: 10.1109/SAMI.2019.8782776]

19. Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I. Painful data: the unbcmcmaster shoulder pain expression archive database. Presented at: 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG); Mar 21-25, 2011; Santa Barbara, CA, USA. [doi: 10.1109/FG.2011.5771462]
20. Saccò M, Meschi M, Regolisti G, et al. The relationship between blood pressure and pain. J Clinical Hypertension 2013 Aug;15(8):600-605. [doi: 10.1111/jch.12145]
21. Thiam P, Kessler V, Amirian M, et al. Multi-modal pain intensity recognition based on the SenseEmotion database. IEEE Trans Affective Comput 2019;12(3):743-760. [doi: 10.1109/TAFFC.2019.2892090]
22. Moscato S, Zhu W, Guo Y, et al. Comparison of autonomic signals between healthy subjects and chronic low back pain patients at rest and during noxious stimulation. Presented at: 8th National Congress of Bioengineering, GNB 2023 - Proceedings; Jun 21-23, 2023; Padova, Italy.
23. Lötsch J, Ultsch A, Kalso E. Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. Br J Anaesth 2017 Oct 1;119(4):821-829. [doi: 10.1093/bja/aex236] [Medline: 29121286]
24. Fetz K, Lefering R, Kaske S. Pre-trauma pain is the strongest predictor of persistent enhanced pain patterns after severe trauma: results of a single-centre retrospective study. Medicina (Kaunas) 2023 Jul 19;59(7):1327. [doi: 10.3390/medicina59071327] [Medline: 37512138]
25. Clay FJ, Watson WL, Newstead SV, McClure RJ. A systematic review of early prognostic factors for persisting pain following acute orthopedic trauma. Pain Res Manag 2012;17(1):35-44. [doi: 10.1155/2012/935194] [Medline: 22518366]
26. Lee JY, Fakhereddin M, MacDermid JC, Elliott JM, Schabrun SM, Walton DM. An exploration of blood marker×environment interaction effects on pain severity and interference scores in people with acute musculoskeletal trauma. Clin J Pain 2021 Oct 1;37(10):747-758. [doi: 10.1097/AJP.0000000000000961] [Medline: 34292185]
27. Raza MM, Zaslansky R, Gordon DB, et al. Chronic breast pain prior to breast cancer surgery is associated with worse acute postoperative pain outcomes. J Clin Med 2021 Apr 27;10(9):1887. [doi: 10.3390/jcm10091887] [Medline: 33925567]
28. Posada-Quintero HF, Chon KH. Innovations in electrodermal activity data collection and signal processing: a systematic review. Sensors (Basel) 2020 Jan 15;20(2). [doi: 10.3390/s20020479] [Medline: 31952141]
29. Subramanian S, Barbieri R, Brown EN. Point process temporal structure characterizes electrodermal activity. Proc Natl Acad Sci U S A 2020 Oct 20;117(42):26422-26428. [doi: 10.1073/pnas.2004403117] [Medline: 33008878]
30. Horvers A, Tombeng N, Bosse T, Lazonder AW, Molenaar I. Detecting emotions through electrodermal activity in learning contexts: a systematic review. Sensors (Basel) 2021 Nov 26;21(23):7869. [doi: 10.3390/s21237869] [Medline: 34883870]
31. Li S, Sung B, Lin Y, Mitas O. Electrodermal activity measure: a methodological review. Ann Tour Res 2022 Sep;96:103460. [doi: 10.1016/j.annals.2022.103460]
32. Braithwaite JJ, Watson DG, Jones R, Rowe M. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRS) for psychological experiments. Psychophysiology 2013;49(1):1017-1034.
33. Subramaniam SD, Dass B. Automated Nociceptive pain assessment using physiological signals and a hybrid deep learning network. IEEE Sensors J 2020;21(3):3335-3343. [doi: 10.1109/JSEN.2020.3023656]
34. Kusumaningrum A, Rustina Y, Abuzairi T, Ibrahim N. The skin conductance-based non-invasive pain assessment instrument for infants. Sri Lanka J Child Health 2022;51(3):448. [doi: 10.4038/sljch.v51i3.10249]
35. Gohel V, Mehendale N. Review on electromyography signal acquisition and processing. Biophys Rev 2020 Nov 10;12(6):1361-1367. [doi: 10.1007/s12551-020-00770-w] [Medline: 33169207]
36. Simao M, Mendes N, Gibaru O, Neto P. A review on electromyography decoding and pattern recognition for human-machine interaction. IEEE Access 2019;7:39564-39582. [doi: 10.1109/ACCESS.2019.2906584]
37. Geisser ME, Ranavaya M, Haig AJ, et al. A meta-analytic review of surface electromyography among persons with low back pain and normal, healthy controls. J Pain 2005 Nov;6(11):711-726. [doi: 10.1016/j.jpain.2005.06.008] [Medline: 16275595]
38. Kelati A, Nigussie E, Dhaou IB, Plosila J, Tenhunen H. Real-Time classification of pain level using zygomaticus and corrugator EMG features. Electronics (Basel) 2022;11(11):1671. [doi: 10.3390/electronics11111671]
39. Gruss S, Treister R, Werner P, et al. Pain intensity recognition rates via biopotential feature patterns with support vector machines. PLoS One 2015;10(10):e0140330. [doi: 10.1371/journal.pone.0140330] [Medline: 26474183]
40. van Bruinessen IR, van den Ende ITA, Visser LNC, van Dulmen S. The impact of watching educational video clips on analogue patients' physiological arousal and information recall. Patient Educ Couns 2016 Feb;99(2):243-249. [doi: 10.1016/j.pec.2015.08.022] [Medline: 26427309]
41. Ree A, Mayo LM, Leknes S, Sailer U. Touch targeting C-tactile afferent fibers has a unique physiological pattern: a combined electrodermal and facial electromyography study. Biol Psychol 2019 Jan;140:55-63. [doi: 10.1016/j.biopsycho.2018.11.006] [Medline: 30468895]
42. Heesink L, Geuze E. Pre-processing of electromyography startle data: a novel semi-automatic method. Presented at: Proceedings of Measuring Behavior; Aug 27-29, 2014; Wageningen, The Netherlands.
43. Lubba CH, Sethi SS, Knaute P, Schultz SR, Fulcher BD, Jones NS. catch22: canonical time-series characteristics: selected through highly comparative time-series analysis. Data Min Knowl Discov 2019;33(6):1821-1852. [doi: 10.1007/s10618-019-00647-x]

44. Mietus JE, Peng CK, Henry I, Goldsmith RL, Goldberger AL. The pNNx files: re-examining a widely used heart rate variability measure. Heart 2002 Oct;88(4):378-380. [doi: 10.1136/heart.88.4.378] [Medline: 12231596]
45. Wang X, Wirth A, Wang L. Structure-based statistical features and multivariate time series clustering. Presented at: Seventh IEEE international conference on data mining (ICDM 2007); Oct 28-31, 2007; Omaha, NE, USA.
46. Walpole RE, Myers RH, Myers SL, Ye K. Probability and Statistics for Engineers and Scientists: Macmillan; 1993, Vol. 5.
47. St»hle L, Wold S. Analysis of variance (ANOVA). Chemometr Intell Lab Syst 1989 Nov;6(4):259-272. [doi: 10.1016/0169-7439(89)80095-4]

Abbreviations

EDA: electrodermal activity
EMG: electromyography
RMS: root mean square

Edited by J Shaikh-Mohammed; submitted 09.01.25; peer-reviewed by J Henderson, Y Li; revised version received 19.06.25; accepted 20.06.25; published 20.08.25.

Please cite as:
Ozek B, Lu Z, Radhakrishnan S, Kamarthi S
Influence of Pre-Existing Pain on the Body's Response to External Pain Stimuli: Experimental Study
JMIR Biomed Eng 2025;10:e70938
URL: https://biomedeng.jmir.org/2025/1/e70938
doi:10.2196/70938

Thigh-Worn Sensor for Measuring Initial and Final Contact During Gait in a Mobility Impaired Population: Validation Study

Thomas Johnson, MEng; Janeesata Kuntapun, Msc; Craig Childs, PhD; Andrew Kerr, PhD

Department of Biomedical Engineering, University of Strathclyde, 106 Rottenrow, Glasgow, United Kingdom

Corresponding Author:
Thomas Johnson, MEng
Department of Biomedical Engineering, University of Strathclyde, 106 Rottenrow, Glasgow, United Kingdom

Abstract

Background: Adapting physical activity monitors to detect gait events (ie, at initial and final contact) has the potential to build a more personalized approach to gait rehabilitation after stroke. Meeting laboratory standards for detecting these events in impaired populations is challenging, without resorting to a multisensor solution. The Teager-Kaiser energy operator (TKEO) estimates the instantaneous energy of a signal; its enhanced sensitivity has successfully detected gait events from the acceleration signals of individuals with impaired mobility, but has not been applied to stroke.

Objective: This study aimed to test the criterion validity of TKEO gait event detection (and derived spatiotemporal metrics) using data from thigh mounted physical activity monitors compared with concurrent 3D motion capture in chronic survivors of stroke.

Methods: Participants with a history of stroke(n=13, mean age 59, SD 14 years), time since stroke (mean 1.5, SD 0.5 years), walking speed (mean 0.93_{ms}^{-1}, SD 0.38 m/s) performed two 10m walks at their comfortable speed, while wearing two ActivPAL 4+ (AP4) sensors (anterior of both thighs) and LED cluster markers on the pelvis and ankles which were tracked by a motion capture system. The TKEO signal processing technique was then used to extract gait events (initial and final contact) and calculate stance durations which were compared with motion capture data.

Results: There was very good agreement between the AP4 and motion capture data for stance duration (AP4 0.85s, motion capture system 0.88s, 95% CI of difference −0.07 to 0.13, intraclass correlation coefficient [3,1]=0.79).

Conclusions: The TKEO method for gait event detection using AP4 data provides stance time durations that are comparable with laboratory-based systems in a population with chronic stroke. Providing accurate stance time durations from wearable sensors could extend gait training out of clinical environments. Limitations include ecological and external validity. Future work should confirm findings with a larger sample of participants with a history of stroke.

Trial Registration: ClinicalTrials.gov NCT06787768; https://clinicaltrials.gov/study/NCT06787768

(JMIR Biomed Eng 2025;10:e80308) doi:10.2196/80308

KEYWORDS

accelerometers; Teager-Kaiser energy operator; stance phase estimation; stroke; thigh-worn sensor; spaciotemporal

Introduction

Rehabilitation improves recovery after stroke, with better outcomes when applied intensively and tailored to individual needs [1-4]. Wearable technology has the potential to support increased self-managed rehabilitation by providing performance feedback during everyday activities such as walking in the community but needs to be accurate [5].

As an important feature of independent living, the recovery of walking ability is a major rehabilitation goal [6]. Reduced walking ability is common after stroke, particularly in hemiplegic stroke, which causes specific gait impairments such as reduced stance time duration on the hemiplegic side [3,7]. Speed is the most widely reported and clinically accessible metric in stroke rehabilitation, used as a global measure of mobility recovery [8,9]. Speed alone, however, does not provide meaningful information on the underlying impairments needed to inform effective rehabilitation interventions [10]. One metric that could provide this insight and is clinically relevant and sensitive to change, is stance duration symmetry [11-13]. Measuring this metric during everyday rehabilitation is problematic but achievable with wearable sensors; however, such an approach should consider measurement burden and potential Hawthorne effects [14]. Integrating these metrics into existing systems, for example physical activity monitors, may be a more acceptable approach.

One well-established wearable device, designed for measuring free-living physical activity, is the ActivPAL4+ (AP4) activity monitor (PAL technologies), which is a uniaxial accelerometer attached to the anterior thigh, using proprietary algorithms to

measure physical activity and posture of a healthy and impaired populations within free-living environments [15]. This includes standing, sitting, walking (durations and transitions) and measurements of stepping cadence, step counts and energy expenditure [16]. The AP4 has good validity for measuring walking bouts at normal walking speeds, intraclass correlation coefficient [ICC (2,1)=0.78] when compared to direct observation [15,17-19].

Many algorithms and analysis techniques have been developed to measure gait parameters from wearable sensors, from frequency domain metrics to more complex approaches that calculate joint angular displacements [20,21]. To detect gait events specifically, most approaches have used peak detection algorithms and zero crossing techniques, but these have not been found to be robust [22-24]. Gait event detection using the Teager Kaiser Energy Operator (TKEO) with AP4 data has been attempted before with Huntington's disease, to determine initial (IC) and final contact (FC) events, with the resulting stance phase calculation consistently underestimated (0.08 s), compared to video analysis [25,26]. While this technique appears promising from tests in healthy and impaired participants populations data, there is a need to test with a hemiplegic population post stroke who stand to benefit from the enhanced gait rehabilitation, that could be enabled by the feedback from this approach.

The aim of this study, therefore, was to test the concurrent validity of a thigh-mounted physical activity sensor (AP4) for measuring the stance phase duration of hemiplegic gait, with a 3D motion capture system details acting as the gold standard measure. The hypothesis was that the AP4 would have acceptable levels of concurrent validity through an ICC (3,1) greater than 0.75, and a low (<0.1 s) absolute difference between the two systems experience.

Methods

Recruitment

The data were collected as part of a larger rehabilitation trial.

Study Design

Concurrent validation of an accelerometer-based system (AP4, PAL technologies, Glasgow, UK) was performed against 3D motion capture (Vicon) for measuring stance phase duration in participants of hemiplegic stroke.

Data Capture

Data from the AP4 and motion capture system were captured concurrently from participants with hemiplegic chronic stroke during two 10m walks in a gait laboratory. An AP4 (43 mmx26 mmx5 mm) was attached to the anterior surface of each thigh (ie, hemiplegic and nonhemiplegic side) with the acceleration data sampled at 50Hz. Marker clusters (Pulsars, Vicon) were attached to participants at three locations (posterior pelvis, lateral malleoli) using Velcro straps and tracked with 37 cameras (Viper, Vicon) sampling at 120 HZ. The 3D trajectory data for the clusters was captured using commercial software (Evoke, Vicon) and processed with a customized Python script (version 3.13.2, Python Software Foundation).

Stance Time Duration Calculation

Stance time durations were calculated from the motion capture data using a coordinate-based algorithm, modified to use ankle-placed cluster markers [27]. For the acceleration data, Teager and Kaiser [28] developed an algorithm that used the amplitude and frequency of a signal to discern its energy. This algorithm is defined as:

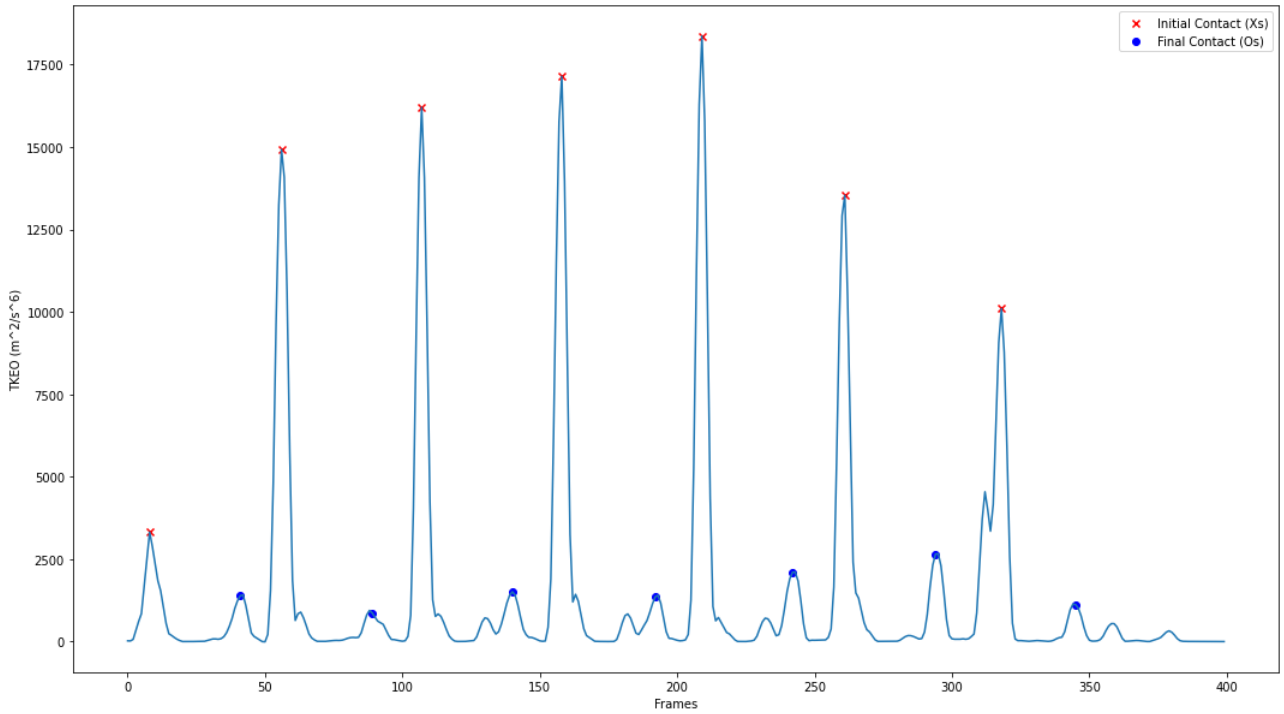
$$(1)\Psi[x(t)]=\dot{x}(t)-x(t)\ddot{x}(t)$$

with $\Psi x(t)$ equating the energy of the signal, x, at time, t. $\dot{x}(t)$ and $\ddot{x}t$ denoting the first and second derivative of the signal, x, respectively. The discrete time variant is required for determining specific gait events. To obtain this, a 3-sample symmetric difference is calculated, approximating the first and second derivatives. For a discrete one-dimensional accelerometer signal, xn, the TKEO signal, ϕn, is obtained through equation (1):

$$(2)\phi n=[2xn2+(xn+1-xn-1)2-xn(xn+2-xn-2)]4Ts2$$

Flood et al [26] used equation (2) to amplify accelerometer signal transient features for gait event detection. The TKEO, a nonlinear energy-tracking operator, is considered effective in amplifying sudden changes in signal energy. This allows for identifying gait-related events, such as IC and FC, as seen in Figure 1.

Figure 1. Example TKEO output featuring IC and FC locations. Each large peak in anterior-posterior (AP) acceleration corresponds to IC, red crosses. Processing removed these peaks and calibrated the surrounding regions. After further processing, the largest remaining peaks correspond to FC, blue circles. AP: ; IC: initial contact; FC: final contact; TKEO: Teager-Kaiser energy operator.



An initial high-pass filter, at 0.5Hz via a 4th-order Butterworth filter is used on the anterior-posterior (AP) acceleration signal of the thigh-mounted sensor. Equation (2) is then applied to obtain ϕn. Once ϕn was obtained, the processing of the signal was conducted in the same fashion as Lozano-García et al [25], but distinguished by using the constraints and calibration discussed by Flood et al [26].

Applying the TKEO technique meant that the AP4 could remain in its recommended location on the anterior thigh, enabling collection of physical activity parameters using the AP4's proprietary algorithms [15], as well as collection of stance duration data.

Data Analysis

A Kolmogorov-Smirnov test was conducted to determine the normality of the difference between the AP4 and Evoke cluster marker system (ECMS). A two-sample *t* test was then used to determine whether the AP4 was statistically different from that of the ECMS when measuring stance times. A Bland-Altman plot and 95% CI of the limits of agreement (LOA) were used to compare the AP4 to the ECMS. A two-way mixed-effect, absolute agreement, single-measures ICC (3,1) was calculated to determine absolute agreement between the AP4 and ECMS.

Ethical Considerations

This study received ethical approval from Strathclyde University Ethics Committee (UEC25/23: Kerr) and is a registered clinical trial (NCT06787768) [29,30]. All participants of the main trial were invited to take part in this substudy. All participants provided informed consent before their involvement in the study. Participants did not receive compensation, and all data were anonymized.

Results

The Montreal cognitive assessment (MoCA) is a screening instrument that evaluates general mental capabilities, such as visuospatial abilities, executive functions, and orientation to time and space. The MoCA is rated between 1 and 30, with increasing score dictating better cognitive ability [31]. The functional ambulation category (FAC) evaluates walking ability in 6 levels, with a score of 0 defining no ability to walk, or requiring the help of 2 physiotherapist, and a score of 5 defining full capability to walk independently, including stairs [32]. The Rivermead mobility index (RMI) is an outcome measure used to assess mobility after stroke, rated between 0 and 15, with increasing score indicating better mobility [33] (Table 1).

The AP4 (mean stance time=0.85) saw consistent underestimation in comparison to the ECMS (mean stance time=0.88). Despite this, both methods had a high agreement for stance time measurement (T=0.61, *P*=.54, 95% CI for difference=−0.07s, 0.13s, ICC [3,1]=0.79). The differences between the AP4 and ECMS were shown to be within a standard distribution (*D*=0.11, *P*=.10).

The Bland-Altman plot (Figure 2) displays a consistent spread between the techniques during the performed stroke gait which ranged between 0.49 and 1.90. This shows close agreement between the AP4 and ECMS, irrespective of the value. A bias of 0.03 seconds is reported, suggesting excellent accuracy of the AP4 in comparison to ECMS. The LOA were 0.28s and −0.22s, which contain the 95% of the recorded datapoints. This indicates good agreement between the AP4 and ECMS.

Table . Participant characteristics of validation study.

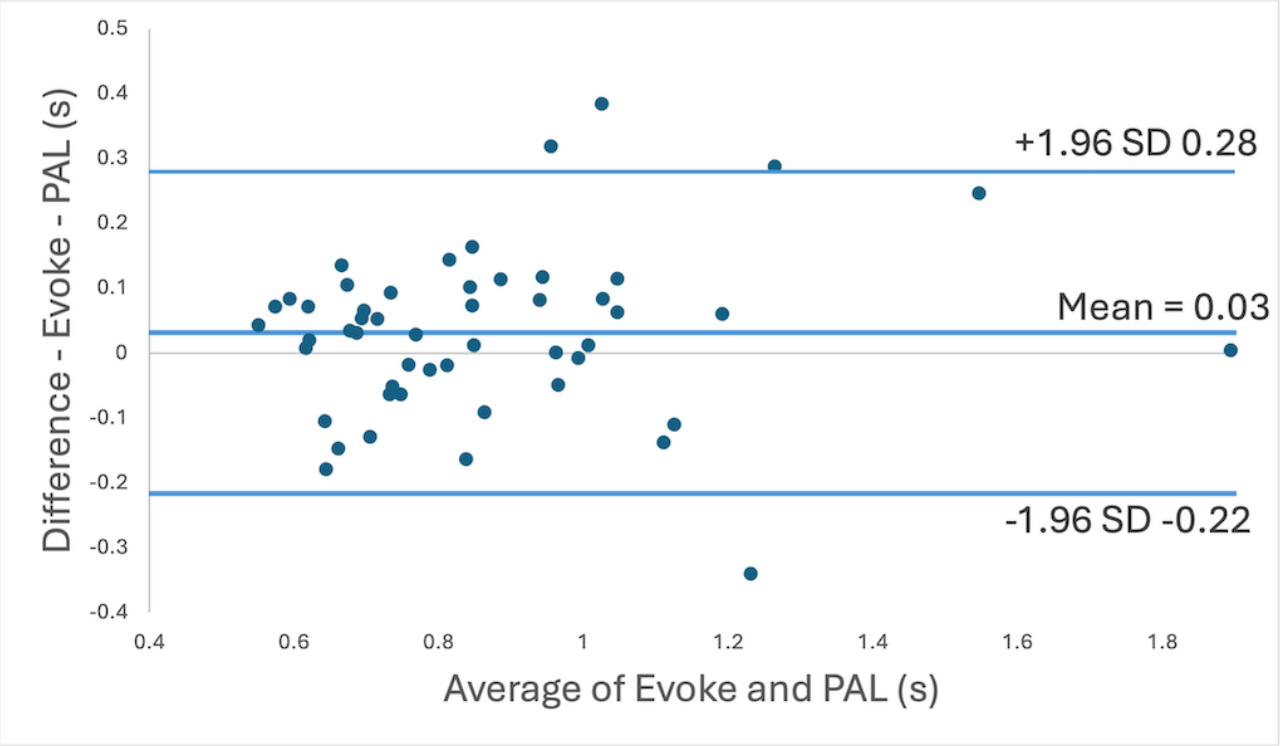
Characteristics	Participants (N=13)
Age (years)	30 - 78
Aphasia, n	5
Years since stroke, range	1 - 13
MoCA[a] (1-30), range	16 - 30
FAC[b] (0 - 5), range	1 - 5
Walking speed, range (m/s)	0.06 - 1.39
RMI[c]	9 - 15

[a]MoCA: Montreal cognitive assessment.

[b]FAC: functional ambulation category.

[c]RMI: Rivermead mobility index.

Figure 2. Bland-Altman plot comparing Evoke motion capture system stance time measurement using the Zeni technique with the AP4 using the Teager-Kaiser energy operator (TKEO) measurement technique. The X-axis shows the mean of the two measurements and the Y-axis shows the difference of the two measurements. The central line represents the mean difference with the outer lines representing 95% limits of agreement.



Discussion

This study aimed to test the concurrent validity of detecting gait events (ie, initial and final foot contact) and the derived metric—stance phase duration—from thigh acceleration data analyzed with the TKEO, compared with a gold standard 3D motion capture (Vicon) in chronic survivors of stroke. The excellent accuracy (bias=0.03 s) were similar to Flood et al (IC error=0.01s, FC error=0.02 s) [26]and Lozano-García et al (bias=0.08 s) [25] . In the context of a typical stance duration (nonparetic =1.01 s), the reported difference of 0.03 s (2 sampling frames) represents a 2.97% difference, with performance comparable to other commonly used methods of wearable sensor gait parameter measurement, collected by Pacini Panebianco [34] et al. The reported bias of 0.03 seconds could be considered clinically unimportant. A study by Wang [35] et al denoted stance time averages and variability from survivors of stroke at different gait speeds. They noted 0.90 (SD 0.34) seconds and 1.01 (SD 0.41) seconds for the paretic and nonparetic sides, respectively. The magnitude of the bias, therefore, will have little effect on the readings that the AP4 will make by using the TKEO technique. These findings should be confirmed with a larger sample that preserves the diverse nature of the current sample (speed ranged between $0.06 - 1.39_{ms^{-1}}$).

This positive outcome opens opportunity to expand the output of wearable sensors currently used as physical activity monitors, to provide feedback on gait parameters (stance duration,

symmetry and weight transference) during the rehabilitation of patients with hemiplegic stroke, without increasing measurement burden.

The results should be interpreted in light of the study's limitations. The testing took place in a laboratory setting, reducing the generalizability of results. This controlled environment may not reflect everyday gait [36,37]. To obtain ecologically valid gait data, there is a need to capture in free-living environments. The sample of participants with a history of chronic stroke means that only a specific subset (chronic) of the population with stroke has been validated for this technique. Future studies are encouraged to include subacute populations who may have more variable gait parameters. The study had a small sample size, limiting the statistical power and potentially not representing the whole population, although the range of walking ability (walking speed 0.06 - 1.39$_{ms}^{-1}$, FAC 1 - 5; Table 1) is reassuringly wide.

In conclusion, the AP4 sensor, in conjunction with the TKEO technique, has been validated against a gold standard 3D Motion Capture system, for stance duration measurement in participants with chronic stroke. However, this positive finding is limited by the study's setting and small sample. Future work should consider a bigger sample and collect gait data in free-living environments. The outcomes of this study could be exploited to enhance the function of wearable sensors, in order to provide the gait parameters valuable for self-rehabilitation after stroke, such as symmetry.

Acknowledgments

The research was supported through a PhD studentship from the University of Strathclyde. We are thankful for the support of the Sir Jules Thorn Center for the Co-creation of Rehabilitation Technology.

The ActivPAL 4+ sensors were provided gratis by PAL technologies.

Data Availability

The datasets generated or analyzed during this study are available in the University of Strathclyde, Glasgow repository [38].

Conflicts of Interest

PAL technologies part-funded the PhD studentship, whilst providing ActivPAL 4+ sensors during the study. The PhD studentship has also been part-funded by the University of Strathclyde.

References

1. Sivertsen M, De Jaegher H, Arntzen EC, Alstadhaug KB, Normann B. Embodiment, tailoring, and trust are important for co - construction of meaning in physiotherapy after stroke: a qualitative study. Physiotherapy Res Intl 2022 Jul;27(3):e1948. [doi: 10.1002/pri.1948]
2. Rincé G, Couturier C, Berrut G, Dylis A, Montero-Odasso M, Deschamps T. Impact of an individual personalised rehabilitation program on mobility performance in older-old people. Aging Clin Exp Res 2021 Oct;33(10):2821-2830. [doi: 10.1007/s40520-021-01812-3] [Medline: 33625687]
3. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. The Lancet 2011 May;377(9778):1693-1702. [doi: 10.1016/S0140-6736(11)60325-5]
4. Mawson S, Nasr N, Parker J, Zheng H, Davies R, Mountain G. Developing a personalised self-management system for post stroke rehabilitation; utilising a user-centred design methodology. Disability and Rehabilitation: Assistive Technology 2014 Nov;9(6):521-528. [doi: 10.3109/17483107.2013.840863]
5. Harris NR, Sthapit D. Towards a personalised rehabilitation system for post stroke treatment. Presented at: 2016 IEEE Sensors Applications Symposium (SAS); Apr 20-22, 2016; Catania, Italy. [doi: 10.1109/SAS.2016.7479848]
6. Taylor-Piliae RE, Latt LD, Hepworth JT, Coull BM. Predictors of gait velocity among community-dwelling stroke survivors. Gait Posture 2012 Mar;35(3):395-399. [doi: 10.1016/j.gaitpost.2011.10.358] [Medline: 22119886]
7. Chang MC, Lee BJ, Joo NY, Park D. The parameters of gait analysis related to ambulatory and balance functions in hemiplegic stroke patients: a gait analysis study. BMC Neurol 2021 Jan 27;21(1):38. [doi: 10.1186/s12883-021-02072-4] [Medline: 33504334]
8. Hsu AL, Tang PF, Jan MH. Analysis of impairments influencing gait velocity and asymmetry of hemiplegic patients after mild to moderate stroke. Arch Phys Med Rehabil 2003 Aug;84(8):1185-1193. [doi: 10.1016/s0003-9993(03)00030-3] [Medline: 12917858]
9. Grau-Pellicer M, Chamarro-Lusar A, Medina-Casanovas J, Serdà Ferrer BC. Walking speed as a predictor of community mobility and quality of life after stroke. Top Stroke Rehabil 2019 Jul;26(5):349-358. [doi: 10.1080/10749357.2019.1605751] [Medline: 31063439]
10. Tasseel-Ponche S, Delafontaine A, Godefroy O, et al. Walking speed at the acute and subacute stroke stage: a descriptive meta-analysis. Front Neurol 2022;13:989622. [doi: 10.3389/fneur.2022.989622] [Medline: 36226075]

11. Dewar ME, Judge G. Temporal asymmetry as a gait quality indicator. Med Biol Eng Comput 1980 Sep;18(5):689-693. [doi: 10.1007/BF02443147] [Medline: 7464295]
12. Patterson KK, Gage WH, Brooks D, Black SE, McIlroy WE. Changes in gait symmetry and velocity after stroke: a cross-sectional study from weeks to years after stroke. Neurorehabil Neural Repair 2010;24(9):783-790. [doi: 10.1177/1545968310372091] [Medline: 20841442]
13. Wonsetler EC, Bowden MG. A systematic review of mechanisms of gait speed change post-stroke. Part 1: spatiotemporal parameters and asymmetry ratios. Top Stroke Rehabil 2017 Sep;24(6):435-446. [doi: 10.1080/10749357.2017.1285746] [Medline: 28220715]
14. Zhou L, Rackoll T, Ekrod L, et al. Monitoring and visualizing stroke rehabilitation progress using wearable sensors. Presented at: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 15-19, 2024; Orlando, FL, USA. [doi: 10.1109/EMBC53108.2024.10782489]
15. Wu Y, Petterson JL, Bray NW, Kimmerly DS, O'Brien MW. Validity of the activPAL monitor to measure stepping activity and activity intensity: a systematic review. Gait & Posture 2022 Sep;97:165-173. [doi: 10.1016/j.gaitpost.2022.08.002]
16. Edwardson CL, Winkler EAH, Bodicoat DH, et al. Considerations when using the activPAL monitor in field-based research with adult populations. J Sport Health Sci 2017 Jun;6(2):162-178. [doi: 10.1016/j.jshs.2016.02.002] [Medline: 30356601]
17. Yaqoob I, Gusso S, Simpson M, Meiring RM. Agreement between the activPAL accelerometer and direct observation during a series of gait and sit-to-stand tasks in people living with cervical dystonia. Front Neurol 2024;15:1286447. [doi: 10.3389/fneur.2024.1286447] [Medline: 38725651]
18. Kirk AG, Kimmel LA, Behm KJ, Peiris CL, Ekegren CL. Validity of the activPAL and ActiGraph for measuring sitting time and steps in hospitalised orthopaedic patients with altered weight bearing. Disabil Rehabil 2024 Jan;46(2):378-386. [doi: 10.1080/09638288.2022.2157896] [Medline: 36541196]
19. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. Phys Ther 2017 May 1;97(5):581-588. [doi: 10.1093/ptj/pzx010] [Medline: 28339904]
20. Gurchiek RD, Garabed CP, McGinnis RS. Gait event detection using a thigh-worn accelerometer. Gait & Posture 2020 Jul;80:214-216. [doi: 10.1016/j.gaitpost.2020.06.004]
21. Strick JA, Farris RJ, Sawicki JT. A novel gait event detection algorithm using a thigh-worn inertial measurement unit and joint angle information. J Biomech Eng 2024 Apr 1;146(4):044502. [doi: 10.1115/1.4064435] [Medline: 38183222]
22. Selles RW, Formanoy MAG, Bussmann JBJ, Janssens PJ, Stam HJ. Automated estimation of initial and terminal contact timing using accelerometers; development and validation in transtibial amputees and controls. IEEE Trans Neural Syst Rehabil Eng 2005 Mar;13(1):81-88. [doi: 10.1109/TNSRE.2004.843176] [Medline: 15813409]
23. Köse A, Cereatti A, Della Croce U. Bilateral step length estimation using a single inertial measurement unit attached to the pelvis. J Neuroeng Rehabil 2012 Feb 8;9(1):9. [doi: 10.1186/1743-0003-9-9] [Medline: 22316235]
24. Lee HK, Hwang SJ, Cho SP, Lee DR, You SH, Lee KJ. Novel algorithm for the hemiplegic gait evaluation using a single 3-axis accelerometer. Presented at: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Sep 3-6, 2009; Minneapolis, MN. [doi: 10.1109/IEMBS.2009.5333650]
25. Lozano-Garcia M, Doheny EP, Mann E, et al. Estimation of gait parameters in Huntington's disease using wearable sensors in the clinic and free-living conditions. IEEE Trans Neural Syst Rehabil Eng 2024;32:2239-2249. [doi: 10.1109/TNSRE.2024.3407887] [Medline: 38819972]
26. Flood MW, O'Callaghan BPF, Lowery MM. Gait event detection from accelerometry using the Teager–Kaiser Energy Operator. IEEE Trans Biomed Eng 2020;67(3):658-666. [doi: 10.1109/TBME.2019.2919394]
27. Zeni JA, Richards JG, Higginson JS. Two simple methods for determining gait events during treadmill and overground walking using kinematic data. Gait & Posture 2008 May;27(4):710-714. [doi: 10.1016/j.gaitpost.2007.07.007]
28. Kaiser DA, Kaiser JF. Estimation of power systems amplitudes, frequencies, and phase characteristics using energy operators. Presented at: 2012 IEEE Energy Conversion Congress and Exposition (ECCE); Sep 15-20, 2012; Raleigh, NC, USA. [doi: 10.1109/ECCE.2012.6342719]
29. Kerr A, Grealy M, Slachetka M, et al. A participatory model for cocreating accessible rehabilitation technology for stroke survivors: user-centered design approach. JMIR Rehabil Assist Technol 2024;11:e57227-e57227. [doi: 10.2196/57227]
30. Kerr A, Keogh M, Slachetka M, Grealy M, Rowe P. An intensive exercise program using a technology-enriched rehabilitation gym for the recovery of function in people with chronic stroke: usability study. JMIR Rehabil Assist Technol 2023;10:e46619. [doi: 10.2196/46619]
31. Chiti G, Pantoni L. Use of Montreal Cognitive Assessment in patients with stroke. Stroke 2014 Oct;45(10):3135-3140. [doi: 10.1161/STROKEAHA.114.004590] [Medline: 25116881]
32. Mehrholz J, Wagner K, Rutte K, Meiβner D, Pohl M. Predictive validity and responsiveness of the functional ambulation category in hemiparetic patients after stroke. Arch Phys Med Rehabil 2007 Oct;88(10):1314-1319. [doi: 10.1016/j.apmr.2007.06.764]
33. Roorda LD, Green JR, Houwink A, et al. The Rivermead mobility index allows valid comparisons between subgroups of patients undergoing rehabilitation after stroke who differ with respect to age, sex, or side of lesion. Arch Phys Med Rehabil 2012 Jun;93(6):1086-1090. [doi: 10.1016/j.apmr.2011.12.015]

34. Pacini Panebianco G, Bisi MC, Stagni R, Fantozzi S. Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements. Gait & Posture 2018 Oct;66:76-82. [doi: 10.1016/j.gaitpost.2018.08.025]
35. Wang Y, Mukaino M, Ohtsuka K, et al. Gait characteristics of post-stroke hemiparetic patients with different walking speeds. Int J Rehabil Res 2020 Mar;43(1):69-75. [doi: 10.1097/MRR.0000000000000391] [Medline: 31855899]
36. Bock O, Beurskens R. Changes of locomotion in old age depend on task setting. Gait & Posture 2010 Oct;32(4):645-649. [doi: 10.1016/j.gaitpost.2010.09.009]
37. Atrsaei A, Corrà MF, Dadashi F, et al. Gait speed in clinical and daily living assessments in Parkinson's disease patients: performance versus capacity. NPJ Parkinsons Dis 2021 Mar 5;7(1):24. [doi: 10.1038/s41531-021-00171-0] [Medline: 33674597]
38. Johnson T, Kerr A. Data for "thigh-worn sensor for measuring initial and final contact during gait in a mobility impaired population: a validation study". University of Strathclyde, Glasgow. URL: https://pureportal.strath.ac.uk/en/datasets/data-for-thigh-worn-sensor-for-measuring-initial-and-final-contac/

Abbreviations

AP4: ActivPAL4+
ECMS: Evoke cluster marker system
FAC: functional ambulation category
FC: final contact
IC: initial contact
ICC: intraclass correlation coefficient
LOA: limits of agreement
MoCA: Montreal cognitive assessment
RMI: Rivermead mobility index
TKEO: Teager-Kaiser energy operator

Edited by J Sarvestan; submitted 08.07.25; peer-reviewed by A Torad, J guo; revised version received 01.10.25; accepted 02.10.25; published 30.10.25.

Please cite as:
Johnson T, Kuntapun J, Childs C, Kerr A
Thigh-Worn Sensor for Measuring Initial and Final Contact During Gait in a Mobility Impaired Population: Validation Study
JMIR Biomed Eng 2025;10:e80308
URL: https://biomedeng.jmir.org/2025/1/e80308
doi:10.2196/80308

Research Letter

Can Artificial Intelligence Diagnose Knee Osteoarthritis?

Mihir Tandon[1], BA; Nitin Chetla[2], BS; Adarsh Mallepally[3]; Botan Zebari[4], BS; Sai Samayamanthula[2], BA; Jonathan Silva[1], BS; Swapna Vaja[5], BS; John Chen[1], BS; Matthew Cullen[1], BS; Kunal Sukhija[6], MD

[1]Albany Medical College, Albany, NY, United States

[2]University of Virginia School of Medicine, Charlottesville, VA, United States

[3]School of Medicine, Virginia Commonwealth University, Richmond, VA, United States

[4]St. James School of Medicine, Binghamton, NY, United States

[5]Rush Medical College, Chicago, IL, United States

[6]Kaweah Health, Visalia, CA, United States

Corresponding Author:
Mihir Tandon, BA
Albany Medical College
43 New Scotland Ave
Albany, NY, 12208
United States
Phone: 1 3322488708
Email: tandonm@amc.edu

Related Article:

This is a corrected version. See correction statement: https://biomedeng.jmir.org/2025/1/e82980

Abstract

This study analyzed the capability of GPT-4o to properly identify knee osteoarthritis and found that the model had good sensitivity but poor specificity in identifying knee osteoarthritis; patients and clinicians should practice caution when using GPT-4o for image analysis in knee osteoarthritis.

(JMIR Biomed Eng 2025;10:e67481) doi:10.2196/67481

KEYWORDS

large language model; ChatGPT; GPT-4o; radiology; osteoarthritis; machine learning; X-rays; osteoarthritis detection

Introduction

Osteoarthritis often affects the knee, causing pain and disability, and is typically diagnosed by X-ray [1]. Advancements in artificial intelligence (AI) offer potential to automate image analysis, reducing diagnostic burden [2]. Given its widespread availability, tools like ChatGPT have potential as point-of-care diagnostic aids. AI has already been incorporated on the physician side through clinical decision support systems and robotic surgery. On the patient side, AI is used in applications such as virtual health assistants [3].

Orthopedic surgeons, radiologists, and primary care physicians can use AI tools to streamline their workflows and reduce errors while analyzing imaging for pathologies like osteoarthritis. Moreover, patients use ChatGPT to analyze their imaging to further understand their condition [4]. The ability of AI to read other radiological images (eg, computed tomography angiograms) has been shown to be subpar [5]. However, studies have shown that AI can perform well with X-rays [6]. As such, it is increasingly important for physicians to understand AI's strengths and limitations to assess its use in imaging and guide patients using AI for self-diagnosis.

Methods

We queried ChatGPT (using the GPT-4o version) and assessed its performance in classifying 500 X-ray images of normal knees and 500 images of knees with osteoarthritis from a publicly available Kaggle database [7]. Images were verified based on consensus among radiologists. A single standardized prompt was used: "This is an x-ray image found on examination, the multiple-choice question is as follows. Based on the x-ray image, does the patient have A) no osteoarthritis, B) osteoarthritis." Key metrics included accuracy, sensitivity, and specificity. No

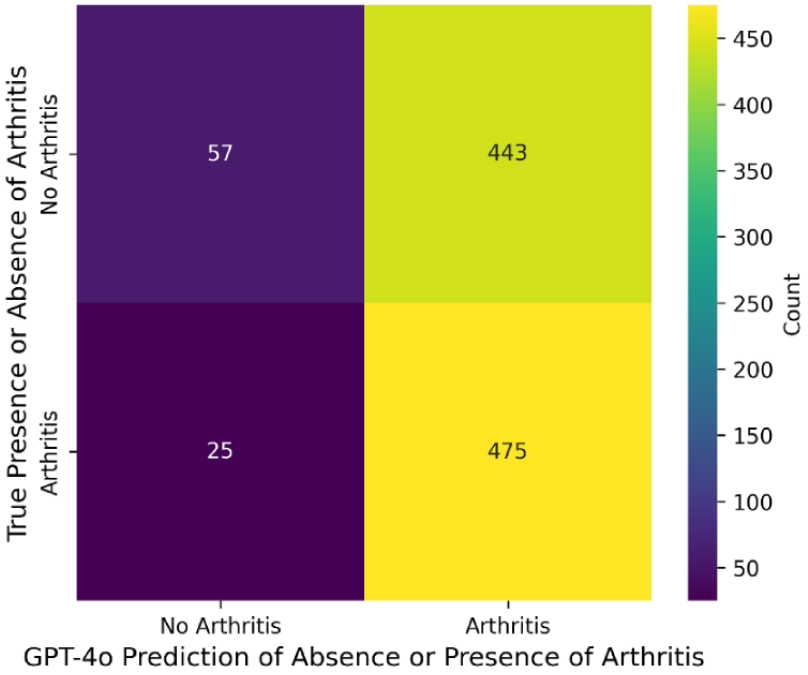
images were rejected by ChatGPT. The code used for statistical analysis is included in Multimedia Appendix 1.

Results

The model's performance in distinguishing osteoarthritis from nonosteoarthritis knee X-rays was mixed. The high recall (0.950, 95% CI 0.964-0.943) suggests that the model was sensitive in identifying arthritis cases, while the low specificity (0.114, 95% CI 0.134-0.104) indicated a poor ability to correctly identify nonosteoarthritis cases. The F_1-score (0.670, 95% CI 0.699-0.655) balanced precision and recall, showing moderate effectiveness, but the precision (0.517, 95% CI 0.548-0.501) reflected that about half the predicted osteoarthritis cases were correct. Accuracy was 0.532 (95% CI 0.563-0.516). Figure 1 shows sensitivity and specificity.

The binomial test, where the null hypothesis assumed the model's accuracy was 50% or less, indicated that the model was statistically better than random chance (P=.02). Additionally, the χ^2 test (P<.001) indicated a strong dependence between the model's predictions and the actual labels, demonstrating that its classifications were not purely random. However, the significance of this test should be interpreted with caution, as it does not necessarily reflect high accuracy or clinical reliability.

Figure 1. Sensitivity and specificity of Chat-GPT4o in analyzing knee osteoarthritis X-rays.



Discussion

The model had difficulty distinguishing between "not arthritis" and "arthritis." While the recall for arthritis was high (0.950), indicating strong performance in identifying true arthritis cases, the low specificity (0.114) reflects a significant number of false positives, with many nonarthritis cases misclassified as arthritis. This bias toward predicting arthritis lowered precision (0.517) and accuracy (0.532); similar misclassification issues have been reported in other ChatGPT studies [8].

Limitations include, first, that the prompt was binary. A binary prompt was used because it would have been difficult to analyze data obtained with an open-ended prompt. Second, the dataset was small; a larger dataset would have yielded more robust conclusions.

Even with its limitations, this study presents important data on GPT4o's use in imaging for diagnosing osteoarthritis. This is vital, as our understanding of tools like this in health care contexts is limited. These results suggest a need for better class balance and improved feature differentiation. Similar misclassification patterns have been noted in previous studies, where overlapping features led to false positives [9]. A higher-resolution, more comprehensively annotated osteoarthritis dataset could improve model training, enhancing overall accuracy, sensitivity, and specificity. Thus, future work should focus on analyzing larger datasets and refining the model to handle more nuanced cases more effectively, improving performance statistics. Using image preprocessing techniques, such as contrast enhancement and noise reduction, and including metadata like medical history and clinical presentation could also help distinguish osteoarthritis from anatomical variations.

Our results suggest that clinicians should use ChatGPT cautiously and as a screening tool prior to their own validation to help mitigate misclassification. Clinicians should also educate patients about the risks of using AI for self-diagnosis of osteoarthritis based on X-rays. Despite its shortcomings, AI has potential for developing more reliable diagnostic models for osteoarthritis.

Authors' Contributions

Conceptualization: NC (lead), MT (equal), KS (equal)
Data curation: AM (lead), MT (equal), SS (supporting), SV (supporting), JC (supporting)
Formal analysis: JC (lead), JS (supporting), MC (supporting), SV (supporting), AM (supporting)
Funding acquisition: KS (lead)
Investigation: SS (lead), KS (equal), BZ (supporting), SV (supporting)
Methodology: MT (lead), NC (equal), KS (equal), AM (supporting)
Resources: SV (lead), JC (supporting)
Software: JC (lead), AM (supporting)
Supervision: KS (lead), MT (equal), NC (equal)
Validation: JS (lead), JC (equal), MC (equal)
Visualization: MT (lead), MC (equal), SS (supporting)
Writing – original draft: MT (lead), NC (equal), BZ (supporting), SS (supporting), AM (supporting)
Writing – review & editing: JS (lead), SV (equal), JC (equal), MC (supporting), KS (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Code for analysis and prompting.
[DOCX File , 17 KB - biomedeng_v10i1e67481_app1.docx]

References

1. Choi MS, Lee DK. The effect of knee joint traction therapy on pain, physical function, and depression in patients with degenerative arthritis. J Kor Phys Ther 2019 Oct 31;31(5):317-321 [FREE Full text] [doi: 10.18857/jkpt.2019.31.5.317]
2. Bejarano A. The benefits of artificial intelligence in radiology: transforming healthcare through enhanced diagnostics and workflow efficiency. Rev Contemp Sci Acad Stud 2023 Aug 30;3(8):1-4. [doi: 10.55454/rcsas.3.08.2023.005]
3. Chatterjee I, Ghosh R, Sarkar S, Das K, Kundu M. Revolutionizing innovations and impact of artificial intelligence in healthcare. Int J Multidiscip Res 2024 May 14;6(3):19333. [doi: 10.36948/ijfmr.2024.v06i03.19333]
4. Zhang Z, Citardi D, Wang D, Genc Y, Shan J, Fan X. Patients' perceptions of using artificial intelligence (AI)-based technology to comprehend radiology imaging data. Health Informatics J 2021;27(2):14604582211011215 [FREE Full text] [doi: 10.1177/14604582211011215] [Medline: 33913359]
5. Young A, Tan K, Tariq F, Jin MX, Bluestone AY. Rogue AI: cautionary cases in neuroradiology and what we can learn from them. Cureus 2024 Mar;16(3):e56317 [FREE Full text] [doi: 10.7759/cureus.56317] [Medline: 38628986]
6. Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, et al. Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents. JAMA Netw Open 2020 Oct 01;3(10):e2022779 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22779] [Medline: 33034642]
7. Kabir F. Osteoarthritis prediction. Kaggle. URL: https://www.kaggle.com/datasets/farjanakabirsamanta/osteoarthritis-prediction [accessed 2024-09-01]
8. Dalalah D, Dalalah OM. The false positives and false negatives of generative AI detection tools in education and academic research: the case of ChatGPT. Int J Manag Educ 2023 Jul;21(2):100822. [doi: 10.1016/j.ijme.2023.100822]
9. Truhn D, Weber CD, Braun BJ, Bressem K, Kather JN, Kuhl C, et al. A pilot study on the efficacy of GPT-4 in providing orthopedic treatment recommendations from MRI reports. Sci Rep 2023 Dec 17;13(1):20159 [FREE Full text] [doi: 10.1038/s41598-023-47500-2] [Medline: 37978240]

Abbreviations

AI: artificial intelligence

Edited by S Rizvi, T Leung; submitted 12.10.24; peer-reviewed by Y Chaibi, A Jahnen, M Nayak; comments to author 25.02.25; revised version received 13.03.25; accepted 25.03.25; published 23.04.25.

Please cite as:
Tandon M, Chetla N, Mallepally A, Zebari B, Samayamanthula S, Silva J, Vaja S, Chen J, Cullen M, Sukhija K
Can Artificial Intelligence Diagnose Knee Osteoarthritis?
JMIR Biomed Eng 2025;10:e67481
URL: https://biomedeng.jmir.org/2025/1/e67481
doi:10.2196/67481
PMID:40266670

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org