JMIR Neurotechnology

For research exploring how technologies (e.g. information technology, neural engineering, neural interfacing, clinical data science, robotics, eHealth/mHealth) can be applied in clinical neuroscience (e.g., neurology, neurosurgery, neuroradiology) to prevent, diagnose, and treat neurological disorders.

Volume 4 (2025) ISSN 2817-092X Editors-in-Chief: Pieter Kubben MD, PhD

Contents

Viewpoint

Original Papers

Transforming Perceptions: Exploring the Multifaceted Potential of Generative AI for People With Cognitive Disabilities

Dorit Hadar Souval[1], MA, PhD; Yuval Haber[2], MA, PhD; Amir Tal[3], MA, PhD; Tomer Simon[4], BSc, MPH, PhD; Tal Elyoseph[5], MA; Zohar Elyoseph[6,7], MA, PhD

[1]Max Stern Academic College of Emek Yezreel, Tel Adashim, Israel

[2]Interdisciplinary Studies Unit, PhD Program of Hermeneutics and Cultural Studies, Bar-Ilan University, Ramat Gan, Israel

[3]Tel Aviv University, Tel Aviv, Israel

[4]Microsoft Israel R&D Center, Tel Aviv, Israel

[5]Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel

[6]Imperial College London, Fulham Palace Rd, London, United Kingdom

[7]Faculty of Education, University of Haifa, Haifa, Israel

Corresponding Author:
Zohar Elyoseph, MA, PhD
Imperial College London, Fulham Palace Rd, London, United Kingdom

Abstract

Background: The emergence of generative artificial intelligence (GenAI) presents unprecedented opportunities to redefine conceptions of personhood and cognitive disability, potentially enhancing the inclusion and participation of individuals with cognitive disabilities in society.

Objective: We aim to explore the transformative potential of GenAI in reshaping perceptions of cognitive disability, dismantling societal barriers, and promoting social participation for individuals with cognitive disabilities.

Methods: This study is a critical review of current literature in disability studies, artificial intelligence (AI) ethics, and computer science, integrating insights from disability theories and the philosophy of technology. The analysis focused on 2 key aspects: GenAI as a social mirror reflecting societal values and biases, and GenAI as a cognitive partner for individuals with cognitive disabilities.

Results: This paper proposes a theoretical framework for understanding the impact of GenAI on perceptions of cognitive disability. It introduces the concepts of GenAI as a "social mirror" that reflects and potentially amplifies societal biases and as a "cognitive copilot" providing personalized assistance in daily tasks, social interactions, and environmental navigation. This paper also presents a novel protocol for developing AI systems tailored to the needs of individuals with cognitive disabilities, emphasizing user involvement, ethical considerations, and the need to address both the opportunities and challenges posed by GenAI.

Conclusions: Although GenAI has great potential for promoting the inclusion and empowerment of individuals with cognitive disabilities, realizing this potential requires a change in societal attitudes and development practices. This paper calls for interdisciplinary collaboration and close partnership with the disability community in the development and implementation of GenAI technologies. Realizing the potential of GenAI for promoting the inclusion and empowerment of individuals with cognitive disabilities requires a multifaceted approach. This involves a shift in societal attitudes, inclusive AI development practices that prioritize the needs and perspectives of the disability community, and ongoing interdisciplinary collaboration. This paper emphasizes the importance of proceeding with caution, recognizing the ethical complexities and potential risks alongside the transformative possibilities of GenAI technology.

(JMIR Neurotech 2025;4:e64182) doi:10.2196/64182

KEYWORDS

generative artificial intelligence; cognitive disability; social participation; AI ethics; assistive technology; cognitive disorder; societal barriers; social inclusion; disability study; social mirror; cognitive partner; empowerment; user involvement; GenAI; artificial intelligence; neurotechnology; neuroinformatics; digital health; health informatics; neuroscience; mental health; computer science; machine learning

Introduction

In the era of generative artificial intelligence (GenAI), traditional notions of personhood and normality are being challenged [1-4]. Technological advances are blurring the boundaries between human and machine capabilities, offering an opportunity to expand the limits of social inclusion and promote change in attitudes toward people with disabilities [1]. As artificial intelligence (AI) systems demonstrate increasingly sophisticated cognitive abilities, they prompt us to reconsider what qualities define personhood and human intelligence. This paper examines the potential of GenAI to disrupt limiting conceptions of morality and humanity, focusing on the implications of GenAI for the social status of people with cognitive disabilities. This paper also proposes a practical toolkit for GenAI development and engineering professionals—product managers, data scientists, and developers—to help incorporate these insights into their work.

Cognitive disability refers to a wide range of impairments affecting cognitive functions such as learning, problem-solving, judgment, communication, and social interaction [5]. Examples of cognitive disabilities include intellectual disability, attention-deficit/hyperactivity disorder, autism spectrum disorders, specific learning disabilities (such as dyslexia), and brain injuries (such as traumatic brain injury or stroke) [5-7]. It is important to emphasize the variety of individuals with cognitive disabilities, each one possessing a unique combination of strengths, impairments, and potential, which means that cognitive disabilities require personalized approaches to intervention. While recognizing the diverse nature of cognitive disabilities and the need for tailored solutions, this paper focuses on the general potential of GenAI to improve the lives of people across the spectrum of cognitive disabilities.

Engaging with the integration of GenAI and individuals with cognitive disabilities is a new direction in the use of technology in the field of disability. The potential for AI to support and empower this population lies in its ability to perform cognitive tasks such as reasoning, planning, decision-making, and communication—areas that are challenging for people with cognitive disabilities [8-10]. The ability of AI to remove barriers and open new paths for inclusive and equitable participation makes it especially relevant for this population [11]. An in-depth analysis of this ability requires examining the philosophical and ethical implications of AI for conceptions of humanity and morality, questions that directly determine how society views and accommodates individuals with cognitive disabilities. These are fundamental inquiries into the nature of intelligence, personhood, consciousness, and human agency, which largely determine the degree of participation and inclusion for this group.

Personhood and AI: An Opportunity for Paradigm Shift

The concept of personhood, which emerged as a central topic in bioethical debates surrounding topics such as abortion, stem cell research, and euthanasia, has evolved into a complex and multifaceted construct that now spans multiple disciplines [12]. Inherently normative in nature, personhood involves value judgments and ethical considerations regarding how we ought to treat and perceive others rather than merely describing observable facts. Personhood is not rooted exclusively in our biology and experiences but in our essence and identity. This identity, however, is not formed in isolation; it is dynamically shaped in an intricate interaction between self-perception and the perception of others and interaction with them. Rosfort [13] argued that this conceptualization of personhood reveals its profoundly relational and social nature, demonstrating how identity and perception of self-worth are inextricably woven into interactions and the broader human context.

The concept of “personhood” has long served as a central criterion in bioethical discussions, determining which entities deserve moral consideration and rights [3]. As a result, this notion has also functioned as a mechanism of exclusion, denying basic rights and opportunities to those deemed cognitively “abnormal” [14].

For example, historically, people with cognitive disabilities were excluded from the public sphere and denied the right to make decisions for themselves [15,16]. Even today, despite significant progress in discourse and work based on the “social model” (an approach that views disability as created by societal barriers rather than by individual impairments alone) [17] and the “minority group model” (which recognizes people with disabilities as a marginalized minority group) [18], exclusion still exists in various aspects of life. People with cognitive disabilities still face barriers to accessing higher education and vocational training because of preconceived notions about their abilities [19]. Despite relevant skills, they have difficulties securing meaningful employment and career advancement opportunities because of social stigma and prejudice [20]. Participation in political or civic decision-making processes, such as voting or community involvement, is limited by discriminatory perceptions of the competence of individuals with cognitive disabilities [21]. They are also excluded from leisure, social, and cultural activities because of a lack of access or restrictive attitudes toward their participation [22].

These exclusion examples illustrate how, as a result of conceptualizing what constitutes a person of merit, individuals with cognitive disabilities are often excluded in the deepest and broadest ways from society. This mechanism is difficult to identify because it operates through our language and the most basic organized mechanisms of any society: law, health care system, education system, and more [23].

Breaking entrenched concepts and perceptions of personhood is challenging because they are deeply embedded in societal structures and norms, but emerging technologies are beginning to challenge these long held beliefs. GenAI offers an opportunity to challenge the definition of personhood perceptions by demonstrating skills previously considered unique to humans [1,4]. Although these capabilities are not yet perfect in AI, their very existence challenges the idea that such traits belong exclusively to the “normal” cognitive function of humans and that social participation is conditional on the presence of these abilities.

The revolutionary potential of GenAI invites us to reexamine the criteria for membership in the moral community and expand them beyond limiting standards. Instead of relying on a narrow model of "correct" cognitive abilities as a prerequisite for rights and participation in society [14], we may adopt, with the assistance of GenAI, a more inclusive view that recognizes human diversity and the inherent value of all individuals, regardless of their abilities [24]. By showcasing the potential of machines to exhibit complex cognitive traits, GenAI challenges the notion that certain abilities are essential for personhood and moral status. It initiates a discourse on the need to redefine our understanding of what it means to be human and to have moral worth, moving away from a focus on cognitive benchmarks and toward a more encompassing vision of human dignity and rights [1,4].

Although AI presents opportunities to challenge our understanding of personhood, there are legitimate concerns about its potential to exacerbate exclusion and narrow definitions of "normal" human cognition. The inherent biases in AI systems, stemming from their training data and algorithmic design [25-28], risk reinforcing and amplifying existing societal prejudices [29]. As AI increasingly influences decision-making processes in areas such as employment, health care, and criminal justice, there is a danger that it could lead to more stringent and narrow criteria for what constitutes "normal" human functioning. This could inadvertently heighten barriers for individuals with cognitive differences, further marginalizing them from full societal participation [30]. Moreover, as AI systems become more sophisticated in mimicking certain human cognitive abilities, there is a risk that societal expectations of human performance might be unrealistically elevated, potentially creating an even more exclusionary standard of "normal" [31]. Thus, while AI challenges our notions of personhood, it simultaneously risks entrenching and exacerbating existing forms of exclusion, highlighting the critical need for ethical AI development and deployment considering diverse human experiences and capabilities. In the following sections, we will explore 2 key areas where GenAI has the potential to drive significant change: GenAI as a social mirror and GenAI as a cognitive partner. These 2 domains highlight the multifaceted impact that GenAI can have on reshaping perceptions, removing barriers, and promoting participation of individuals with cognitive disabilities on the one hand, and exacerbating existing biases and exclusions in society on the other.

Generative AI as a Social Mirror: Opportunity and Challenge

Overview

Vallor's [32] conceptualization of AI as a societal mirror provides a compelling framework for understanding the role of AI in reflecting and potentially amplifying societal biases, particularly concerning cognitive disabilities. This mirror metaphor can be understood as follows: just as a physical mirror reflects the image of what stands before it, AI systems reflect the data, values, and biases present in the society that created them. However, unlike a simple reflection, AI systems can amplify and distort these reflections, much as a funhouse mirror might exaggerate certain features.

This mirror effect illuminates how AI systems, trained on biased data, risk perpetuating existing prejudices against individuals with cognitive differences. AI essentially learns from and then projects back the biases inherent in its training data, potentially reinforcing and spreading these biases further. Paradoxically, this same reflective quality presents a unique opportunity to identify and address longstanding societal biases, rendering implicit prejudices explicit and subject to scrutiny. By closely examining what the AI "reflects back" to us, we can gain insights into biases that might otherwise remain hidden or unacknowledged in society.

Vallor [32] posits that AI systems in general, and GenAI systems in particular, are not merely neutral technological tools but mirrors reflecting the values, norms, and biases prevalent in human society. Given that these systems are constructed upon data and content created by humans, they inherently risk replicating and perpetuating prejudices and discrimination against marginalized groups, including people with cognitive disabilities [27,33].

A study by Gadiraju et al [34] demonstrated this mirroring effect in action. They conducted 19 focus groups with 56 participants with various disabilities who interacted with a dialog model based on a large language model. The researchers found that the model frequently perpetuated harmful stereotypes and narratives about disability. For example, the model often fixated on physical disabilities, particularly wheelchairs, while neglecting other types of disabilities. It also tended to portray people with disabilities as passive, sad, and lonely, reinforcing the misconception that disability is inherently negative. Additionally, the model sometimes produced what participants referred to as "inspiration porn," objectifying people with disabilities as sources of inspiration for nondisabled people.

For example, if the information used to train AI systems contains stereotypical or derogatory expressions toward people with cognitive disabilities, there is a significant risk that these systems might "learn" to adopt discriminatory attitudes. The potential consequences are severe: AI systems could rank individuals with cognitive disabilities as having lower potential in employment or educational contexts, limit their access to certain services, or make biased decisions about them in critical areas such as insurance or credit [35].

When we look into the societal mirror reflected by AI, several possible human responses can be identified. One metaphorical response is "breaking the mirror," representing human resistance to AI use and the insights it presents [36]. While this approach attempts to avoid the uncomfortable truths AI exposes, it risks missing out on the potential benefits and insights AI can offer. Another metaphorical strategy is "cleaning the mirror," where humans attempt to eliminate biases through AI alignment processes [37]. This approach aims to create AI systems aligned with human values and intentions, striving for a bias-free environment. However, it risks producing an artificially sterile system that fails to reflect the complexities of human cognition and interaction, potentially making AI less relevant and less capable of addressing real-world complexities.

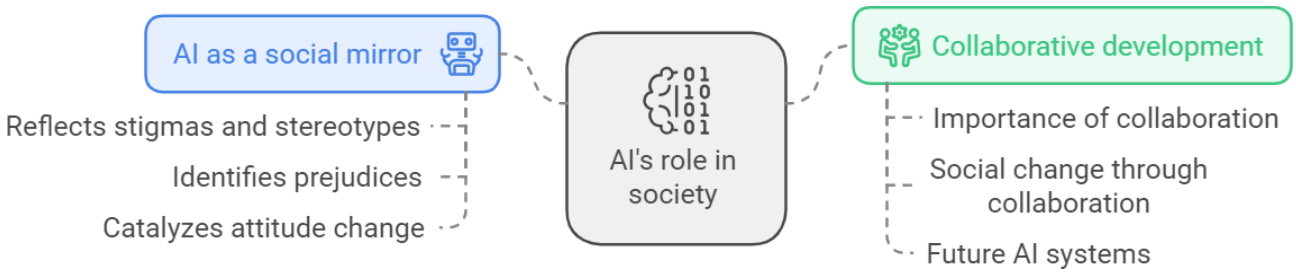
The third and most promising approach involves using reflection as a call to action in the real world. This method requires humans to acknowledge the biases reflected by AI and use this awareness as a catalyst for societal change. It demands active engagement and concrete actions from us as humans to address these issues, both in our AI systems and in society at large [38]. This approach recognizes that if such action is taken, over time, the reflection in the AI mirror itself can change, not as a result of erasing biases in the machine as in the second option, but as a consequence of real societal change that is then differently reflected in the AI mirror.

To implement this approach specifically within the realm of AI development and deployment, we must adopt advanced techniques and ensure inclusive human involvement. As contemporary AI systems increasingly incorporate vast datasets populated from the internet, traditional methods of addressing biases through direct data manipulation, such as the "datasheets" approach proposed by Gebru et al [39], while still valuable in certain contexts, have become more challenging to implement comprehensively. This shift has led to the adoption of complementary techniques that can handle the scale and complexity of modern AI systems such as self-supervised learning [40] and reward modeling [41]. Crucially, these techniques still require human decision-making at key junctures.

To truly address biases and create more equitable AI systems, particularly regarding cognitive disabilities, we must ensure that people with cognitive disabilities are actively involved in these decision-making processes. This collaborative approach aligns with our third strategy, emphasizing real-world action and societal change. By critically examining the biases revealed in AI outputs and involving diverse perspectives in the development process, we can work toward creating more inclusive AI systems. This approach not only helps in developing fairer algorithms and more representative models but also contributes to broader societal change [1,4]. In this way, the AI mirror becomes not just a reflection of our current culture, but a catalyst for the more inclusive society we aspire to create [16,42].

In conclusion, as illustrated in Figure 1, GenAI has the potential to promote social justice and shift perceptions regarding cognitive disabilities. To harness this potential, collaborative work and ongoing effort are required to embed values of accessibility, inclusion, and respect for diversity at the core of technological development. These steps can transform the "reflection in the mirror" into a positive and inclusive image for people with cognitive disabilities, potentially leading to broader societal changes in perception and inclusion.

Figure 1. GenAI as a social mirror: collaborative development for societal change. AI: artificial intelligence; GenAI: generative artificial intelligence.



While this mirror metaphor provides valuable insights, it is important to recognize its limitations. Vallor's conceptualization, though powerful, doesn't fully capture the multifaceted potential of AI, particularly for people with disabilities. It overlooks its capability to actively solve previously intractable problems and enhance accessibility. To provide a more comprehensive understanding, we must expand our view beyond the perception of AI as a mere reflective tool. In the following section, we propose considering AI not only as a mirror but also as a cognitive partner for people with disabilities, emphasizing its potential to actively support and empower individuals with cognitive differences in navigating the world.

Generative AI as a Cognitive Partner for People With Disabilities

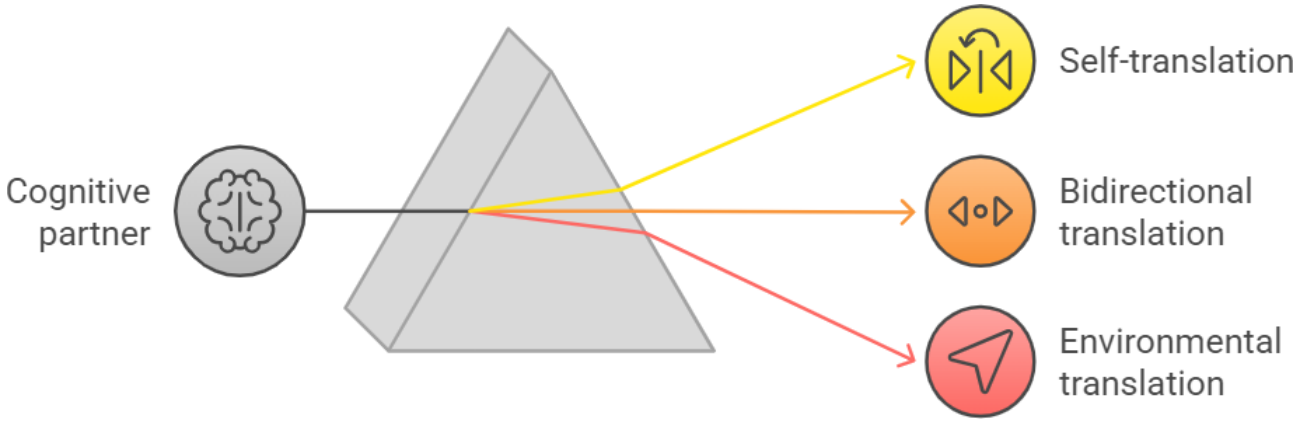
Beyond Vallor's mirror metaphor for AI and its contingent inference on social change for people with cognitive disabilities, a significant potential of GenAI lies in its ability to serve as a "cognitive partner," empowering participation of these people in life domains that were previously blocked or limited for them [43-45]. This partnership can be metaphorically described as a "cognitive copilot" (an AI assistant for complex cognitive tasks), assisting and empowering the individual with tasks requiring complex cognitive functions. For example, GenAI can help a person with cognitive disabilities manage daily tasks such as scheduling, budgeting, or navigating urban spaces by providing personalized reminders, recommendations, and guidance [46,47]. Additionally, it can serve as an advisor in complex social situations, such as interpreting body language [48], suggesting appropriate responses to expressions of anger or mockery from others, or assisting in decision-making [1,49]. In this way, GenAI may act as a kind of "social copilot," providing real-time support and feedback, allowing persons with cognitive disability to expand their circle of social interactions, inclusion, and activities.

One of the outstanding strengths of GenAI is its ability to function as a translator and mediator between languages, concepts, and realities. For people with cognitive disabilities, translation and mediation pose a central challenge in daily life, both in understanding the environment and in expressing themselves in a way others can understand [50]. With its natural learning and processing capabilities, GenAI can bridge these gaps and make information and communication more accessible.

The application of GenAI as a cognitive copilot can focus on 3 main areas (Figure 2):

1. Translating and making the inner world of people with cognitive disability accessible to themselves: GenAI can help people with cognitive disabilities better understand themselves, their thoughts, emotions, and needs. This is achieved by providing explanations and conceptualizations in clear and accessible language, identifying and interpreting emotional states, and suggesting strategies for coping with challenges [50]. GenAI can serve as an "internal translator" that through a process of assistive conceptual scaffolding and cognitive structuring [51] assists individuals in accurate self-understanding and self-expression.
2. Bidirectional translation and mediation in interpersonal communication: By analyzing interpersonal and social information, GenAI can mediate interactions with other people, making it possible to negotiate the complexities inherent in human communication more successfully. The unique contribution of GenAI in this area lies in its ability to bridge the communication gap in both directions, helping the person with cognitive disability understand the social environment, the intentions of others, and the implicit messages in discourse, and making the person's wants, needs, and emotions more accessible to the social environment [1]. For example, on one hand, GenAI can offer interpretations of social cues and recommend appropriate responses, and on the other assist individuals in articulating their thoughts more clearly and presenting their unique perspectives. The technology can serve as a "two-way social translator," enabling people with disability and their environment to better understand each other and promote respectful and equitable communication.
3. Making the physical environment and public spaces accessible: GenAI can act as an "environmental translator," converting complex information about the world into a clear and disability-friendly format. This can include, for example, simplifying official texts, graphically converting numeric data, or creating interactive guides for navigating public spaces [52]. Thus, GenAI models that are open to the public can "see" and "understand" photos and videos and describe their content [1], so that people with cognitive disabilities may gain greater access and independence in managing their lives.

Figure 2. Three main areas of GenAI application as a cognitive partner. GenAI: generative artificial intelligence.



The goal is not to "normalize" individuals with cognitive disabilities or to erase their disability. The cognitive partner metaphor, similar to Vallor's mirror metaphor, can show how the use of AI might exacerbate exclusionary attitudes and further marginalize individuals with disabilities. Therefore, using AI for social change in our attitude toward people with cognitive disabilities means that the aim of this technology should be to enable access to environments and spaces that were previously closed or socially inaccessible to them, while also facilitating the accessibility of these environments to the individuals themselves. The approach should be person-centered, respecting diversity, and tailored to the unique aspirations and needs of everyone, rather than imposing a uniform standard of "proper" functioning.

Serious consideration must be given to the ethical implications of such a close integration between humans and machines, particularly in the areas of autonomy and responsibility. Questions of privacy, data security, and people's ownership of decisions made by AI systems need to be thoroughly examined [52,53]. Robust oversight and regulatory mechanisms must be in place to ensure the responsible and ethical use of AI, safeguarding the rights and well-being of users. This is especially critical when working with vulnerable populations such as people with cognitive disabilities, where protecting individual autonomy is important [27,33].

In conclusion, although AI-based "cognitive copilot" applications for people with cognitive disabilities have the potential to remove barriers, increase participation, and promote equal opportunities across various domains of life, it is essential to proceed with caution. This technology must function as a "translator" to contribute to a more inclusive and equitable society, and we must remain vigilant to its risks. Ensuring that AI development is person-centered, ethically sound, and involves active participation from the disability community is crucial for harnessing its benefits without worsening existing biases and systemic barriers.

Implication for AI Developers and Technologists

GenAI has immense potential to promote inclusion and equality for people with cognitive disabilities but to realize this potential requires a perceptual shift on the part of developers, engineers, researchers, and product managers. Instead of focusing narrowly on "fixing" certain impairments, they must adopt a more holistic approach that views technology as a lever for social integration

and broad improvement in quality of life [54-56]. This involves a transition from regarding GenAI as a mere technical solution to perceiving it as a tool for effecting social change for the population with cognitive disabilities.

In practice, close and ongoing collaboration with people with cognitive disabilities throughout all stages of development is important [57]. Development teams must learn from the unique experiences and needs of individuals with cognitive disability and meaningfully integrate them into the design and construction of GenAI systems and prompts.

Recent research has demonstrated the feasibility and importance of this approach. For example, Newbutt et al [58] conducted a systematic review of studies involving autistic individuals in the design of extended reality technologies. They found that out of 20 studies published between 2002 - 2022, several successfully engaged autistic individuals as active co-designers and cocreators, allowing them to shape the final products according to their needs and preferences. This highlights the growing trend and importance of including the target users in the design process.

This requires a joint definition of goals, adapting user interfaces and user experience to their modes of thinking and communication, and clearly formulating principles of cognitive accessibility from the earliest planning stages [59]. The aspiration is for the empowerment and inclusion of people with cognitive disabilities to be embedded in the core of the technology and in the layer of its use.

Bircanin et al [60] presented a practical approach to including adults with severe intellectual disabilities in co-design through active support. They demonstrated how principles such as “every moment has potential,” “graded assistance,” “little and often,” and “maximizing choice and control” can be applied in design contexts to ensure meaningful participation of individuals with severe cognitive disabilities. This approach provides concrete strategies for AI developers to engage with this population during the development process.

For example, it is important to examine how the prompt-based user interface can be made accessible and adapted to the cognitive and communication characteristics of people with different types of cognitive disabilities. Consideration should be given to whether the development of dedicated products is the right direction or whether personal adaptation at the level of the individual user is preferable [61]. Answering such questions requires ongoing discourse and feedback from the community itself.

Dirks [57] explored the ethical challenges in inclusive software development projects with people with cognitive disabilities. The study emphasized the importance of maximizing choice and control for participants, using a graded assistance approach, and ensuring every moment has potential for meaningful engagement. These principles can guide AI developers in creating more inclusive design processes.

To assist developers and researchers in implementing the principles presented in this paper, we propose a working protocol specifically tailored to the development challenges of GenAI technologies aimed at people with cognitive disabilities. The protocol (Table 1) is based on the model developed by Amershi et al [62], which was formulated following comprehensive research, including a review of academic and industry literature, interviews with experts, and an examination of a wide range of AI-based products. The original model defines 18 general guidelines for designing human-AI interactions across different time frames and stages of interaction. In practice, these guidelines serve as a framework for developing human-centered AI systems, focusing on aspects such as transparency, fairness, reliability, safety, privacy, security, and accountability. Developers and designers use these guidelines to enhance human-AI interaction by implementing practices such as explaining AI decisions to users, designing interfaces that enable user control and feedback, and incorporating mechanisms to identify and mitigate biases [63].

Table . Protocol for designing artifical intelligence (AI) interactions for people with cognitive disabilities.[a]

Stage and dimension		Guidelines for AI interaction with people with cognitive disabilities	Implementation examples
Initial			
	Personal	I1. Identify and adapt to the user's unique cognitive and emotional needs.	I1. Create a personal profile including preferences, abilities, and challenges.
	Interpersonal	I2. Show awareness of the social and cultural context of system use.	I2. Consider the human environment (eg, caregivers or family members) as part of system definition.
During interaction			
	Personal	D1. Provide custom-tailored, gradual, and structured responses to personal needs during use.	D1. Identify difficulties and adapt the level of assistance and feedback in real time.
	Interpersonal	D2. Promote positive and reciprocal communication with the human environment.	D2. Mediate social interactions by simplifying and explaining social cues.
	Environmental	D3. Assist in orientation, navigation, and independent functioning in complex spaces.	D3. Provide detailed instructions and cues on proper conduct in different places.
When the system errs			
	Personal	E1. Handle errors respectfully and in an empowering way, with emphasis on learning and progress.	E1. Provide repeated opportunities to try again, together with verbal encouragement.
	Interpersonal	E2. Involve support persons in the process of learning and correction.	E2. Provide a possibility for a caregiver to assist in problem-solving or making necessary adjustments.
	Environmental	E3. Avoid placing responsibility on the user in complex or unexpected situations.	E3. Make human backup available by default in case of significant problems.
Over time			
	Personal	T1. Continually adapt to the pace of development, learning, and changes in personal needs.	T1. Track progress and adapt tasks and goals accordingly.
	Interpersonal	T2. Show sensitivity to changes in relationships and roles within the support circle.	T2. Update user profiles and access settings based on feedback from the environment.
	Environmental	T3. Show flexibility and adaptability to changing environments and transitions between contexts.	T3. Automatically detect location changes and provide relevant recommendations.
	Collaboration	T4. Actively involve users and stakeholders in the ongoing development of the system.	T4. Provide mechanisms for receiving feedback and involving users in decisions about updates and improvements.

[a]The model for this protocol by Amershi et al [62] is based on extensive research and analysis of a range of artificial intelligence products and defines 18 general guidelines across different stages of interaction. We adapted and extended this model to address specifically the needs and challenges of designing artificial intelligence technologies for people with cognitive disabilities. The protocol incorporates 4 key dimensions: personal, interpersonal, environmental, and collaborative, and provides concrete examples of how these considerations can be integrated throughout the life cycle of the artificial intelligence system. By implementing this protocol, developers can create artificial intelligence tools that empower and enhance the lives of individuals with cognitive disabilities.

Building on the analysis presented in this paper, we expand the model of Amershi et al [62] and adapt it to the 4 central dimensions in which AI systems can assist people with cognitive disabilities: the personal, the interpersonal, the environmental, and the collaborative. For each of these dimensions, we propose guidelines and offer practical examples of how the relevant considerations can be embedded at different stages of the system life cycle, from defining the initial requirements, through ongoing interaction, to continuous adaptation and improvement. The proposed protocol serves as a foundation that requires further development, testing, and investigation, but it can serve as a starting point for discourse and the advancement of best

practices in designing AI systems for individuals with cognitive disabilities.

Conclusion

The emergence of GenAI technologies represents a pivotal moment in reconceptualizing disability and personhood. We suggest that the advent of GenAI challenges assumptions about what qualifies an individual as a "person" and questions the notion that cognitive abilities are the sole determinant of one's rights and societal participation.

In this paper, we explored the transformative potential of GenAI in reshaping perceptions, dismantling barriers, and empowering individuals with cognitive disabilities. By serving as a social mirror [32], AI systems can expose and challenge deeply ingrained biases and prejudices, compelling us to confront the ways we have historically marginalized and excluded the population with cognitive disabilities. Simultaneously, by functioning as a cognitive partner, GenAI may provide unprecedented opportunities for individuals with cognitive disabilities to participate in society.

Realizing this vision requires more than technological innovation, however. It demands a gradual shift in societal attitudes and a sincere effort to involve people with cognitive disabilities in the AI development process, granting them autonomy and recognizing and valuing their abilities. This is where the role of technology professionals and GenAI developers becomes crucial.

The importance of designing AI thoughtfully lies in the understanding that whether we consider AI as a mirror or as a cognitive partner, both metaphors indicate that AI will increasingly mediate how we perceive the world, ourselves, and others, confirming once again McLuhan's [64] statement that "the medium is the message." This means that the significant effect of AI lies not merely in the content we explore through it but in how its very use changes us. Therefore, the design and development of AI tools will profoundly influence the future of human society, how we perceive individuals with disabilities, as well as the rights and social positions they will attain. Therefore, how AI is being shaped now will determine its role in reinforcing existing biases or promoting a more inclusive and equitable society.

The proposed protocol, based on the work by Amershi et al [62], offers a practical framework for implementing these principles as part of GenAI development for people with cognitive disabilities. This paper marks only the beginning of the discussion about GenAI and developmental disabilities, therefore we must remain vigilant regarding the ethical and social implications of GenAI and continue to engage in open, multidisciplinary dialogue about how to harness its potential for the greater good.

The path ahead is complex and challenging, but it is also filled with immense possibilities. As we look toward the future, the evolution of AI from reactive, prompt-based systems to proactive, autopilot models promises to further expand these possibilities, particularly for individuals with cognitive disabilities. These advanced systems, capable of learning user needs and initiating interactions without explicit prompts, could provide more seamless and intuitive support, potentially revolutionizing the way we approach cognitive assistance.

Technological progress also involves an ongoing need for ethical and inclusive development. We must prioritize user autonomy and privacy while maximizing the benefits of technological assistance. This balance is important not only for protecting individual rights but also for ensuring that AI serves the needs of those it aims to support.

By embracing the potential of GenAI while remaining vigilant regarding its ethical implications, researchers, developers, and policy makers can create technologies that not only uplift those who have been historically marginalized but enrich the human experience for us all. In doing so, we may take a step toward a future where technology serves as a platform for inclusivity and empowerment.

Acknowledgments

The authors would like to express their gratitude to the Artificial Third community for promoting multidisciplinary discourse on artificial intelligence in mental health. This community has made possible valuable interactions between researchers in the fields of psychology, disability studies, and artificial intelligence, contributing to the development of this theoretical study.

Conflicts of Interest

The author TS is the chief scientist of R&D at Microsoft Israel. The views and opinions expressed here are those of the authors and do not reflect the official policy or position of Microsoft. TS received no financial compensation for his contribution to this work.

References

1. Elyoseph Z, Shoval DH, Levkovich I. Beyond personhood: ethical paradigms in the generative artificial intelligence era. Am J Bioeth 2024 Jan;24(1):57-59. [doi: 10.1080/15265161.2023.2278546] [Medline: 38236857]
2. Holm S, Lewis J. The ends of personhood. Am J Bioeth 2024 Jan;24(1):30-32. [doi: 10.1080/15265161.2023.2278561] [Medline: 38236874]
3. Blumenthal-Barby J. The end of personhood. Am J Bioeth 2024 Jan;24(1):3-12. [doi: 10.1080/15265161.2022.2160515] [Medline: 36635972]

4. Haber Y, Levkovich I, Hadar-Shoval D, Elyoseph Z. The artificial third: a broad view of the effects of introducing generative artificial intelligence on psychotherapy. JMIR Ment Health 2024 May 23;11:e54781. [doi: 10.2196/54781] [Medline: 38787297]
5. Sutcliffe MS, Radonovich K. Cognitive disabilities in children and adolescents. In: Halpern-Felsher B, editor. Encyclopedia of Child and Adolescent Health: Academic Press; 2023:11-21. [doi: 10.1016/B978-0-12-818872-9.00136-9]
6. Houwen S, Visser L, van der Putten A, Vlaskamp C. The interrelationships between motor, cognitive, and language development in children with and without intellectual and developmental disabilities. Res Dev Disabil 2016 Jun;53-54:19-31. [doi: 10.1016/j.ridd.2016.01.012]
7. Dev P. Introduction to different kinds of cognitive disorders. In: Gupta S, editor. Bio-Inspired Algorithms and Devices for Treatment of Cognitive Diseases Using Future Technologies: IGI Global; 2022:39-55. [doi: 10.4018/978-1-7998-9534-3.ch003]
8. Peltokorpi J, Hoedt S, Colman T, Rutten K, Aghezzaf EH, Cottyn J. Manual assembly learning, disability, and instructions: an industrial experiment. Int J Prod Res 2023 Nov 17;61(22):7903-7921. [doi: 10.1080/00207543.2023.2195957]
9. Heuser P, Letmathe P, Vossen T. Skill development in the field of scheduling: a structured literature review. Eur J Oper Res 2025 Mar;321(3):697-716. [doi: 10.1016/j.ejor.2024.04.005]
10. Rojas M, Balderas DC, Maldonado J, Ponce P, Lopez-Bernal D, Molina A. Lack of verified inclusive technology for workers with disabilities in Industry 4.0: a systematic review. Int J Sustainable Eng 2024 Dec 31;17(1):1-21. [doi: 10.1080/19397038.2024.2328711]
11. Litwin P, Antonelli D, Stadnicka D. Employing disabled workers in production: simulating the impact on performance and service level. Int J Prod Res 2024 Jun 17;62(12):4530-4545. [doi: 10.1080/00207543.2023.2266066]
12. Young G. Personhood across disciplines: applications to ethical theory and mental health ethics. Ethics Med Public Health 2019 Jul;10:93-101. [doi: 10.1016/j.jemep.2019.100407]
13. Rosfort R. Personhood. In: Stanghellini G, Broome M, Fernandez AV, FusarPoli P, Raballo A, Rosfort R, editors. The Oxford Handbook of Phenomenological Psychopathology: Oxford University Press; 2019:335-343. [doi: 10.1093/oxfordhb/9780198803157.013.38]
14. Asch A. Disability, bioethics, and human rights. In: Albrecht GL, Seelman KD, Bury M, editors. Handbook of Disability Studies: Sage Publications; 2001. [doi: 10.4135/9781412976251.n12]
15. Shakespeare T. The social model of disability. In: Davis LJ, editor. The Disability Studies Reader: Routledge; 2006:197-204.
16. Davis LJ, editor. The Disability Studies Reader, 5th edition: Routledge; 2016. [doi: 10.4324/9781315680668]
17. Iacovou M. A contribution towards a possible re-invigoration of our understanding of the social model of disability's potential. Disabil Soc 2021 Aug 9;36(7):1169-1185. [doi: 10.1080/09687599.2020.1787816]
18. Brinkman AH, Rea-Sandin G, Lund EM, et al. Shifting the discourse on disability: moving to an inclusive, intersectional focus. Am J Orthopsychiatry 2023;93(1):50-62. [doi: 10.1037/ort0000653] [Medline: 36265035]
19. Cinquin PA, Guitton P, Sauzéon H. Designing accessible MOOCs to expand educational opportunities for persons with cognitive impairments. Behav Inf Technol 2021 Aug 18;40(11):1101-1119. [doi: 10.1080/0144929X.2020.1742381]
20. Fuentes K, Hsu S, Patel S, Lindsay S. More than just double discrimination: a scoping review of the experiences and impact of ableism and racism in employment. Disabil Rehabil 2024 Feb;46(4):650-671. [doi: 10.1080/09638288.2023.2173315] [Medline: 36724368]
21. Schidel R. Philosophy universal enfranchisement for citizens with cognitive disabilities – a moral-status argument disabilities – a moral-status argument. Crit Rev Int Soc Polit Philos 2023;26(5):658-679. [doi: 10.1080/13698230.2022.2062212]
22. Zhu D, Al Mahmud A, Liu W. Social connections and participation among people with mild cognitive impairment: barriers and recommendations. Front Psychiatry 2023;14:1188887. [doi: 10.3389/fpsyt.2023.1188887] [Medline: 37476544]
23. Foucault M. Discipline and Punish: The Birth of the Prison: Vintage; 1995.
24. Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. Science 2023 Jul 14;381(6654):187-192. [doi: 10.1126/science.adh2586] [Medline: 37440646]
25. Chan CKY, Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. Int J Educ Technol High Educ 2023;20(1):43. [doi: 10.1186/s41239-023-00411-8]
26. Yusuf A, Pervin N, Román-González M. Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives. Int J Educ Technol High Educ 2024;21(1):21. [doi: 10.1186/s41239-024-00453-6]
27. Hadar-Shoval D, Asraf K, Mizrachi Y, Haber Y, Elyoseph Z. Assessing the alignment of large language models with human values for mental health integration: cross-sectional study using Schwartz's theory of basic values. JMIR Ment Health 2024 Apr 9;11:e55988. [doi: 10.2196/55988] [Medline: 38593424]
28. Hadar-Shoval D, Asraf K, Shinan-Altman S, Elyoseph Z, Levkovich I. Embedded values-like shape ethical reasoning of large language models on primary care ethical dilemmas. Heliyon 2024 Sep 30;10(18):e38056. [doi: 10.1016/j.heliyon.2024.e38056] [Medline: 39381244]
29. Cath C, Wachter S, Mittelstadt B, Taddeo M, Floridi L. Artificial Intelligence and the "Good Society": the US, EU, and UK approach. Sci Eng Ethics 2018 Apr;24(2):505-528. [doi: 10.1007/s11948-017-9901-7] [Medline: 28353045]

30. Whittaker M, Alper M, Bennett CL, et al. Disability, bias, and AI. AI Now Institute. 2019. URL: https://ainowinstitute.org/publication/disabilitybiasai-2019 [accessed 2024-12-24]
31. Cave S, Dihal K. The whiteness of AI. Philos Technol 2020 Dec;33(4):685-703. [doi: 10.1007/s13347-020-00415-6]
32. Vallor S. The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking: Oxford University Press; 2024.
33. Hadar-Shoval D, Elyoseph Z, Lvovsky M. The plasticity of ChatGPT's mentalizing abilities: personalization for personality structures. Front Psychiatry 2023;14:1234397. [doi: 10.3389/fpsyt.2023.1234397] [Medline: 37720897]
34. Gadiraju V, Kane S, Dev S, et al. "I wouldn't say offensive but...": disability-centered perspectives on large language models. Presented at: FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency; Jun 12-15, 2023; Chicago, IL, United States of America. [doi: 10.1145/3593013.3593989]
35. Kidd C, Birhane A. How AI can distort human beliefs. Science 2023 Jun 23;380(6651):1222-1223. [doi: 10.1126/science.adi0248] [Medline: 37347992]
36. Vallor S. The AI mirror: reclaiming our humanity in an age of machine thinking. Presented at: AIES '22: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society; May 19-21, 2022; Oxford, United Kingdom. [doi: 10.1145/3514094.3539567]
37. Sætra HS, Coeckelbergh M, Danaher J. The AI ethicist's dirty hands problem. Commun ACM 2023 Jan;66(1):39-41. [doi: 10.1145/3529751]
38. Chien J, Danks D. Beyond behaviorist representational harms: a plan for measurement and mitigation. Presented at: FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency; Jun 3-6, 2024; Rio de Janeiro, Brazil. [doi: 10.1145/3630106.3658946]
39. Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for datasets. Commun ACM 2021 Dec;64(12):86-92. [doi: 10.1145/3458723]
40. Kommrusch S, Monperrus M, Pouchet LN. Self-supervised learning to prove equivalence between straight-line programs via rewrite rules. IEEE Trans Softw Eng 2023;49(7):3771-3792. [doi: 10.1109/TSE.2023.3271065]
41. Da Costa L, Sajid N, Parr T, Friston K, Smith R. Reward maximization through discrete active inference. Neural Comput 2023 Apr 18;35(5):807-852. [doi: 10.1162/neco_a_01574] [Medline: 36944240]
42. Canton E, Hedley D, Spoor JR. The stereotype content model and disabilities. J Soc Psychol 2023 Jul 4;163(4):480-500. [doi: 10.1080/00224545.2021.2017253] [Medline: 35037596]
43. Ferland L, Li Z, Sukhani S, Zheng J, Zhao L, Gini M. Assistive AI for coping with memory loss. Presented at: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18); Feb 2-7, 2018; New Orleans, Louisiana, USA URL: https://cdn.aaai.org/ocs/ws/ws0528/17360-76000-1-PB.pdf [accessed 2024-12-24]
44. Pednekar S, Dhirawani P, Shah R, Shekokar N, Ghag K. Voice-based interaction for an aging population: a systematic review. Presented at: 2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT); Jan 19-20, 2023; Jaipur, India. [doi: 10.1109/ICCT56969.2023.10075801]
45. Robledo-Castro C, Castillo-Ossa LF, Corchado JM. Artificial cognitive systems applied in executive function stimulation and rehabilitation programs: a systematic review. Arab J Sci Eng 2023;48(2):2399-2427. [doi: 10.1007/s13369-022-07292-5] [Medline: 36185593]
46. Huq SM, Maskeliūnas R, Damaševičius R. Dialogue agents for artificial intelligence-based conversational systems for cognitively disabled: a systematic review. Disabil Rehabil Assist Technol 2024 Apr;19(3):1059-1078. [doi: 10.1080/17483107.2022.2146768] [Medline: 36413423]
47. Hueso M, Álvarez R, Marí D, Ribas-Ripoll V, Lekadir K, Vellido A. Is generative artificial intelligence the next step toward a personalized hemodialysis? RIC 2023 Dec 20;75(6):309-317. [doi: 10.24875/RIC.23000162]
48. Elyoseph Z, Refoua E, Asraf K, Lvovsky M, Shimoni Y, Hadar-Shoval D. Capacity of generative AI to interpret human emotions from visual and textual data: pilot evaluation study. JMIR Ment Health 2024 Feb 6;11:e54369. [doi: 10.2196/54369] [Medline: 38319707]
49. Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. Front Psychol 2023;14:1199058. [doi: 10.3389/fpsyg.2023.1199058] [Medline: 37303897]
50. Kelly C, Cornwell P, Hewetson R, Copley A. The pervasive and unyielding impacts of cognitive-communication changes following traumatic brain injury. Int J Lang Commun Disord 2023;58(6):2131-2143. [doi: 10.1111/1460-6984.12923] [Medline: 37424402]
51. Asman O, Segal-Reich M. Supported decision making "adaptive suit" for non-dominating mental scaffolding. Am J Bioeth - Neurosci 2023;14(3):238-240. [doi: 10.1080/21507740.2023.2243878]
52. Tal A, Elyoseph Z, Haber Y, et al. The artificial third: utilizing ChatGPT in mental health. Am J Bioeth 2023 Oct;23(10):74-77. [doi: 10.1080/15265161.2023.2250297] [Medline: 37812102]
53. Elyoseph Z, Gur T, Haber Y, et al. An ethical perspective on the democratization of mental health with generative AI. JMIR Ment Health. 2024 Oct 17. [doi: 10.2196/58011] [Medline: 39417792]
54. Johansson S, Gulliksen J, Lantz A. User participation when users have mental and cognitive disabilities. Presented at: The 17th International ACM SIGACCESS Conference on computers and accessibility; Oct 26-28, 2015; Lisbon, Portugal. [doi: 10.1145/2700648.2809849]

55. Donati M, Pacini F, Baldanzi L, Turturici M, Fanucci L. Remotely controlled electronic goalkeeper: an example of improving social integration of persons with and without disabilities. Appl Sci (Basel) 2023;13(11):6813. [doi: 10.3390/app13116813]
56. Harb B, Sidani D. Smart technologies challenges and issues in social inclusion – case of disabled youth in a developing country. JABS 2022 Mar 23;16(2):308-323. [doi: 10.1108/JABS-10-2020-0389]
57. Dirks S. Ethical challenges in inclusive software development projects with people with cognitive disabilities: ethische herausforderungen in inklusiven softwareentwicklungsprojekten mit menschen mit kognitiven beeinträchtigungen. Presented at: MuC '22: Proceedings of Mensch und Computer 2022; Nov 4-7, 2022; Darmstadt, Germany. [doi: 10.1145/3543758.3547575]
58. Newbutt N, Glaser N, Francois MS, Schmidt M, Cobb S. How are autistic people involved in the design of extended reality technologies? a systematic literature review. J Autism Dev Disord 2024 Nov;54(11):4232-4258. [doi: 10.1007/s10803-023-06130-3] [Medline: 37715864]
59. Shastri K, Boger J, Marashi S, et al. Working towards inclusion: creating technology for and with people living with mild cognitive impairment or dementia who are employed. Dem (Lon) 2022 Feb;21(2):556-578. [doi: 10.1177/14713012211051885] [Medline: 34749536]
60. Bircanin F, Brereton M, Sitbon L, Ploderer B, Azaabanye Bayor A, Koplick S. Including adults with severe intellectual disabilities in co-design through active support. CHI Conf Hum Factor Comput Syst 2021:1-12. [doi: 10.1145/3411764.3445057]
61. Moreno L, Petrie H, Martínez P, Alarcon R. Designing user interfaces for content simplification aimed at people with cognitive impairments. Univ Access Inf Soc 2023 Mar 24;23(1):1-19. [doi: 10.1007/s10209-023-00986-z] [Medline: 37361672]
62. Amershi S, Weld D, Vorvoreanu M, et al. Guidelines for human-AI interaction. Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland, United Kingdom. [doi: 10.1145/3290605.3300233]
63. Bingley WJ, Curtis C, Lockey S, et al. Where is the human in human-centered AI? Insights from developer priorities and user experiences. Comput Hum Behav 2023 Apr;141:107617. [doi: 10.1016/j.chb.2022.107617]
64. McLuhan M. Understanding Media: The Extensions of Man: McGraw Hill; 1964.

Abbreviations

AI: artificial intelligence
GenAI: generative artificial intelligence

Edited by P Kubben; submitted 10.07.24; peer-reviewed by A Yusuf, C Wang; revised version received 22.10.24; accepted 06.11.24; published 15.01.25.

Please cite as:
Hadar Souval D, Haber Y, Tal A, Simon T, Elyoseph T, Elyoseph Z
Transforming Perceptions: Exploring the Multifaceted Potential of Generative AI for People With Cognitive Disabilities
JMIR Neurotech 2025;4:e64182
URL: https://neuro.jmir.org/2025/1/e64182
doi:10.2196/64182

Exploring Speech Biosignatures for Traumatic Brain Injury and Neurodegeneration: Pilot Machine Learning Study

Rahmina Rubaiat[1*], MSc; John Michael Templeton[2], PhD; Sandra L Schneider[3], PhD; Upeka De Silva[4], MEng; Samaneh Madanian[4], PhD; Christian Poellabauer[1*], PhD

[1]Knight Foundation School of Computer and Information Sciences, Florida International University, Miami, FL, United States

[2]Department of Computer Science and Engineering, University of South Florida, Tampa, FL, United States

[3]Department of Communicative Sciences & Disorders, Saint Mary's College, Notre Dame, IN, United States

[4]Department of Computer Science and Software Engineering, Auckland University of Technology, Auckland, New Zealand

[*]these authors contributed equally

Corresponding Author:
Rahmina Rubaiat, MSc
Knight Foundation School of Computer and Information Sciences, Florida International University, Miami, FL, United States

Abstract

Background: Speech features are increasingly linked to neurodegenerative and mental health conditions, offering the potential for early detection and differentiation between disorders. As interest in speech analysis grows, distinguishing between conditions becomes critical for reliable diagnosis and assessment.

Objective: This pilot study explores speech biosignatures in two distinct neurodegenerative conditions: (1) mild traumatic brain injuries (eg, concussions) and (2) Parkinson disease (PD) as the neurodegenerative condition.

Methods: The study included speech samples from 235 participants (97 concussed and 94 age-matched healthy controls, 29 PD and 15 healthy controls) for the PaTaKa test and 239 participants (91 concussed and 104 healthy controls, 29 PD and 15 healthy controls) for the Sustained Vowel (/ah/) test. Age-matched healthy controls were used. Young age-matched controls were used for concussion and respective age-matched controls for neurodegenerative participants (15 healthy samples for both tests). Data augmentation with noise was applied to balance small datasets for neurodegenerative and healthy controls. Machine learning models (support vector machine, decision tree, random forest, and Extreme Gradient Boosting) were employed using 37 temporal and spectral speech features. A 5-fold stratified cross-validation was used to evaluate classification performance.

Results: For the PaTaKa test, classifiers performed well, achieving F_1-scores above 0.9 for concussed versus healthy and concussed versus neurodegenerative classifications across all models. Initial tests using the original dataset for neurodegenerative versus healthy classification yielded very poor results, with F_1-scores below 0.2 and accuracy under 30% (eg, below 12 out of 44 correctly classified samples) across all models. This underscored the need for data augmentation, which significantly improved performance to 60% - 70% (eg, 26 - 31 out of 44 samples) accuracy. In contrast, the Sustained Vowel test showed mixed results; F_1-scores remained high (more than 0.85 across all models) for concussed versus neurodegenerative classifications but were significantly lower for concussed versus healthy (0.59 - 0.62) and neurodegenerative versus healthy (0.33 - 0.77), depending on the model.

Conclusions: This study highlights the potential of speech features as biomarkers for neurodegenerative conditions. The PaTaKa test exhibited strong discriminative ability, especially for concussed versus neurodegenerative and concussed versus healthy tasks, whereas challenges remain for neurodegenerative versus healthy classification. These findings emphasize the need for further exploration of speech-based tools for differential diagnosis and early identification in neurodegenerative health.

(JMIR Neurotech 2025;4:e64624) doi:10.2196/64624

KEYWORDS

speech biosignatures; speech feature analysis; amyotrophic lateral sclerosis; ALS; neurodegenerative disease; Parkinson's disease; detection; speech; neurological; traumatic brain injury; concussion; mobile device; digital health; machine learning; mobile health; diagnosis; mobile phone

Introduction

Overview

The fields of health care and medical diagnostics have witnessed a significant shift toward noninvasive and accessible methods for early detection, assessment, and monitoring of medical conditions. This shift has been driven by technological advancements and growing research interest in digital health solutions [1]. Among these, speech analysis has emerged as a promising avenue, with studies identifying speech as a potential biosignature for a variety of neurodegenerative conditions [2,3]. The ability to reliably distinguish between conditions or detect coexisting disorders is critical for accurate diagnosis, tracking disease progression, and evaluating treatment effectiveness [4].

This pilot study investigates speech-based biosignatures of 2 distinct neurodegenerative conditions, that are, neurodegenerative diseases and mild traumatic brain injuries (mTBIs), specifically concussions. Speech patterns often reflect neurodegenerative health, with specific speech features showing promise for distinguishing between these conditions. The dataset includes individuals with concussions, patients with Parkinson disease (PD), and age-matched healthy controls for both groups (15 samples for each test). These groups were selected to ensure demographic compatibility while addressing the unique speech patterns associated with each condition.

Neurodegenerative diseases, such as PD, are characterized by the progressive loss of neurons in the brain and spinal cord, leading to impairments in motor and cognitive functions [5,6]. PD involves the degeneration of dopaminergic neurons, resulting in clinical symptoms such as tremors, rigidity, bradykinesia, and postural instability [7]. These symptoms worsen over time and lack curative treatments, necessitating reliable diagnostic tools for early intervention [8]. On the other hand, concussions, a form of mTBI, result from sudden trauma to the brain, causing temporary cognitive impairments, disruptions in brain function, and neurochemical changes. Repeated concussions are associated with a heightened risk of neurodegenerative disorders, such as dementia, later in life [9]. Despite their prevalence, approximately 90% of concussions go unreported, leading to inadequate medical attention and potentially catastrophic consequences [10].

Traditional diagnostic methods for neurodegenerative diseases and concussions often rely on observable motor symptoms, such as tremors, gait disturbances, or muscle rigidity, as well as subjective assessments of cognitive impairments [11]. However, emerging research has identified speech as a valuable biomarker for neurodegenerative health. Dysarthria and dysphonia, characterized by changes in articulation and motor speech production, are prevalent in both concussions and neurodegenerative conditions like PD [12-14]. Speech features, such as mel frequency cepstral coefficients (MFCCs), jitter, shimmer, harmonics-to-noise ratio (HNR), and other temporal and spectral attributes, have been shown to correlate with underlying neurodegenerative conditions.

In this study, we analyzed speech data from 2 well-established medical speech tasks, the PaTaKa task and the Sustained Vowel task. These tasks are widely used in clinical settings for assessing speech impairments. The objective of this study is to explore the potential of speech features in differentiating between concussions and neurodegenerative conditions, as well as their respective healthy controls, and to assess the feasibility of using these features as biomarkers for diagnosis. By addressing this objective, we aim to contribute to the development of speech-based diagnostic tools for early and accurate identification of neurodegenerative health conditions.

This study evaluated 37 speech-based features (25 temporal and 12 spectral), applying machine learning models such as support vector machine (SVM), decision tree (DT), random forest (RF), and Extreme Gradient Boosting (XGBoost) to classify between the groups.

The remainder of this paper describes our methodology, feature extraction and analysis, machine learning approaches, and results for the binary classification tasks across the 2 speech tests.

Related Work

Diagnosing brain injuries and neurodegenerative diseases can be challenging; for instance, concussions may present subtle features that are difficult to detect, including using third person witness accounts of the injury, clinical examinations, and laboratory testing, where diagnostic accuracy is not always perfect [15]. Recent work has explored the diagnosis of concussions in athletes using mobile technologies [16] and speech analysis [17,18], while digital assessments, coupled with speech analysis, are also increasingly being used for individuals with neurodegenerative diseases [19]. In a study by Tsanas [19], various speech tasks have been used to distinguish between healthy people and individuals with PD, with relatively high accuracy. Other previous research has investigated the overall symptom severity of individuals with a neurodegenerative condition [11,20], the effectiveness of voice rehabilitation [21], and how to distinguish PD from other conditions such as essential tremor or atypical parkinsonism [22].

The choice of speech task is critical to obtaining speech samples that can be used for subsequent feature extraction and analysis. One commonly used speech task is to ask an individual to produce sustained phonation of vowels. For instance, the study by Mallela et al [23] presents an automatic voice assessment approach for separating healthy individuals from patients with amyotrophic lateral sclerosis (ALS). Although our study focuses exclusively on PD as the representative neurodegenerative condition, references to ALS studies are included to highlight the broader research landscape on neurodegenerative speech biosignatures and their diagnostic significance. Linear discriminant analysis is used to classify phonation, with the most successful model achieving more than 90% accuracy. Similarly, a study by Rueda and Krishnan [24] obtained sustained vowel data from 57 PD patients and 57 healthy individuals, and the study used 5 hierarchical and 1 partition-based clustering techniques to compare and cross-check PD patients at different phases. In some cases, researchers have relied on existing voice recordings, for example, obtained through the Parkinson's Voice Initiative project (the largest speech-PD dataset so far) to analyze voice impairment due to PD [25].

Daudet et al [18] developed a mobile app to diagnose concussions, using data from 47 high-schools and colleges in the Midwest. The study used several speech tasks such as repetition of a sequential motion rate, alternating motion rate, multisyllabic words (words with 4 syllables containing front, middle, and back vowels, and bilabial, alveolar, velar, and glide consonants). The work by Vashkevich et al [26] presented features for detecting pathological changes in acoustic speech signals for ALS diagnosis. It used recordings from 48 people (26 with ALS) and investigated vowel harmony. The features obtained an 88% correct classification performance using linear discriminant analysis. Various speech-based indicators, such as shimmer, jitter, HNR, and other temporal and spectral indicators, have also been explored as dysphonia measures in individuals with neurodegenerative diseases [27]. Finally, in a study by Benba et al [22], the authors investigated the most effective acoustic elements for accurately identifying symptoms of PD, combining shimmer, jitter, pitch, harmonicity, pulses, and voicing by using K-Nearest Neighbor classifiers with different types of kernels (ie, radial basis functions, linear, polynomial, and multilayer perceptron).

Machine learning–based solutions have become the standard for most health care decision-making processes, for example, most previous works focus on differentiating diseased individuals from healthy controls. For example, the work by Tsanas and Arora [28] evaluated 2289 individuals (2023 healthy controls and 246 PD patients) and analyzed 15,227 voice tasks (9994 for healthy controls and 5233 for PD patients). Similarly, the work Bongioanni [29] compared speech-based automatic classification of patients with ALS and healthy people using sustained phoneme generation, diadochokinetic task, and spontaneous speech. They classified voice samples from 25 patients with ALS and 25 healthy participants using SVMs and deep neural networks. More recently, more focus has been given to multiclass scenarios, for example, the study by Benba et al [22] used a Convolutional Neural Network Long Short-term Memory to categorize ALS, PD, and healthy controls. The study analyzed speech data from 60 people, focusing on sentence reading, sound repetition, and sustained vowels.

Though there are studies that had investigate speech features pertaining to neurodegenerative disorders or acquired neurodegenerative disorders like mTBI, there are not many studies exploring speech feature variations between those populations which might co-occur and impact speech production differently.

The aim of this study is to investigate whether distinct speech-based biomarkers, derived from commonly used tasks like the PaTaKa and Sustained Vowel tests, can effectively differentiate between concussed individuals, neurodegenerative conditions (focused on PD), and healthy controls.

Methods

Data Collection

This study focused on 2 widely used speech tasks, the sequential motion rate task (PaTaKa test) and the Sustained Vowel test. The PaTaKa test evaluates speech-motor function by asking participants to take a deep breath and repeatedly articulate “Pa-Ta-Ka” as steadily as possible in 1 breath, providing insights into the rate and precision of sequential articulatory actions.

In the Sustained Vowel test, participants were instructed to sustain the vowel sound “ah” for as long as possible, offering valuable information about voice quality and potential vocal tremor. Both tasks were assigned to four participant groups, that are (1) individuals with concussions, (2) individuals with neurodegenerative conditions (specifically PD), (3) healthy controls age-matched to the concussed group, and (4) healthy controls age-matched to the neurodegenerative group.

Individuals diagnosed with a concussion were evaluated by physicians or athletic trainers using standardized neurocognitive assessment tools, such as ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) by ImPACT Applications, Inc, SCAT (Sport Concussion Assessment Tool), an open-access tool , and SAC (Standardized Assessment of Concussion) by researchers at the University of North Carolina’s Sports Medicine Research Laboratory, within 48 hours of the suspected injury. Individuals with neurodegenerative conditions (ie, PD) were diagnosed by licensed neurologists or family physicians. All participants with PD were in the early stages of disease progression (Hoehn and Yahr stage 1 - 2) and were assessed using tools such as the MDS-UPDRS (Movement Disorder Society - Unified Parkinson’s Disease Rating Scale) and Hoehn and Yahr Scale.

Healthy controls were divided into two groups: (1) young healthy individuals age-matched to the concussed group and (2) older healthy individuals age-matched to the neurodegenerative group. This separation ensures more accurate comparisons between the groups, minimizing the confounding effects of age-related speech differences.

Participants completed the speech tasks using a mobile app (smartphone or tablet) that provided both visual and auditory instructions. The app also recorded the audio samples digitally for subsequent analysis. Audio data were collected from a total of 235 and 239 participants for the PaTaKa and Sustained Vowel tests, respectively, as shown in Table 1.

Table . Description of collected samples.

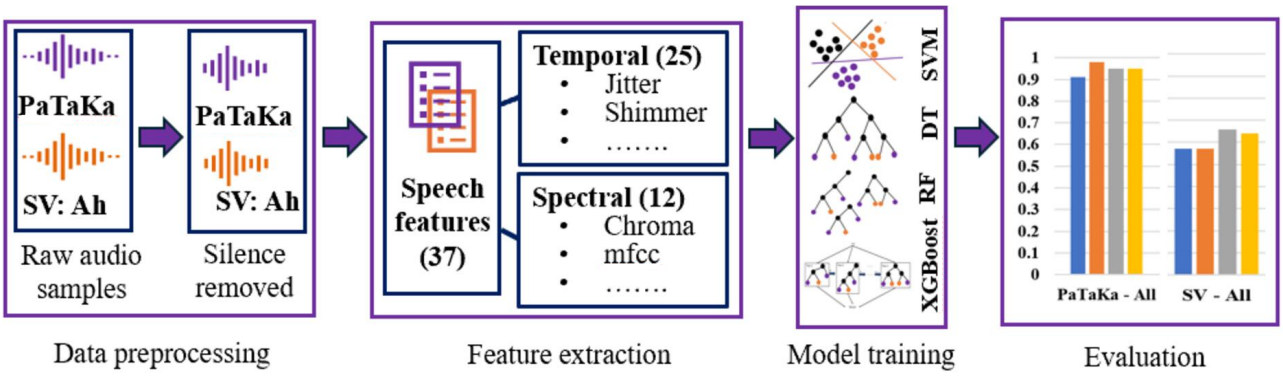
Test name and population		Samples, n	Sex		Age (years), mean (SD)
			Male, n	Female, n	
PaTaKa					
	Concussed	97	86	11	17 (3)
	Healthy control (young)	94	81	13	17 (3)
	Neurodegenerative (PD[a])	29	17	12	63.67 (4.95)
	Healthy control (older)	15	5	10	63.67 (4.95)
Sustained Vowel					
	Concussed	91	82	9	17 (3)
	Healthy control (young)	104	90	14	17 (3)
	Neurodegenerative (PD)	29	17	12	63.67 (4.95)
	Healthy control (older)	15	5	10	63.67 (4.95)

[a]PD: Parkinson disease.

The PaTaKa test dataset includes speech samples from 97 concussed participants, 29 participants with neurodegenerative conditions (ie, PD), 97 age-matched young healthy controls, and 15 age-matched older healthy controls. Similarly, the Sustained Vowel dataset consists of speech samples from 91 concussed participants, 29 participants with neurodegenerative conditions (ie, PD), 91 age-matched young healthy controls, and 15 age-matched older healthy controls.

In the remainder of this section, we describe the 4 key components of the proposed analysis methodology, shown in Figure 1, that are data preprocessing, feature extraction, model training, and evaluation.

Figure 1. Overall visualization of the 4 methodological steps: data preprocessing, feature extraction, model training, evaluation. DT: decision tree; mfcc: mel frequency cepstral coefficient; RF: random forest; SV: Sustained Vowel; SVM: support vector machine; XGBoost: Extreme Gradient Boosting.



Data Preprocessing

The voiced portions of speech signals typically carry the most critical information for analysis. Therefore, to enhance the quality and efficiency of feature extraction, it is essential to eliminate unnecessary components, such as silence intervals and extraneous noise, during the preprocessing phase. In this study, silence intervals were removed at 2 points in each speech recording using the free software developed by Muse group named "Audacity". Specifically, silence was cut from the beginning of the recording to the onset of vocalization and from the offset of vocalization to the end of the recording.

In addition, recordings that did not meet the study's requirements, such as those where participants failed to produce the expected utterances (eg, "PaTaKa" in 1 continuous breath or sustained vowel production without interruptions), were excluded from further analysis. This step ensured a high-quality dataset for feature extraction and classification, thereby improving the reliability of the results.

Data Augmentation

To address the challenges of imbalanced datasets and improve classification performance, data augmentation was applied to specific data subsets, particularly those with limited samples, such as the neurodegenerative (ie, PD) and age-matched healthy datasets. The augmentation process involved adding Gaussian noise to the raw audio signals. The noise factor was set to 0.005 to ensure that the original speech characteristics were preserved

while introducing subtle variations to increase sample diversity. For each audio file, a noise vector was generated using a Gaussian distribution, scaled by the specified noise factor, and added to the original signal. The augmented audio signals were then normalized to ensure they remained within the acceptable amplitude range for further processing.

This step increased the dataset size from 29 PD and 15 healthy samples to 58 PD and 30 healthy samples, resulting in a notable improvement in classification accuracy from under 30% (original data) to 60% - 70% (augmented data).

Feature Extraction

Feature extraction is the process of transforming raw audio data into numerical features while retaining the critical information embedded within the original signal. Among various methods for converting speech into numerical data, temporal and spectral features are widely used in speech-processing research [22,26,27,30]. In these studies, both types of features were extracted using Python's Librosa library [31].

Temporal features describe the changes in an audio signal over time, such as amplitude and pitch variation. This study extracted 25 temporal features, including 4 fundamental frequency measures (eg, mean and SD of F0), 5 jitter measures, 6 shimmer measures, and the HNR. These features provide insights into voice quality and stability, commonly associated with motor speech dysfunctions. The full list and descriptions of these temporal features are provided in Multimedia Appendix 1.

Spectral features analyze the frequency components of the speech signal and are commonly used in applications such as speech recognition and speaker identification. This study extracted 12 spectral features, including MFCC, spectral centroid, chroma features, and spectral flatness. These features capture frequency-domain characteristics that are sensitive to articulation and vocal tract configurations. Detailed descriptions of these spectral features are presented in Multimedia Appendix 1.

All 37 extracted features (25 temporal and 12 spectral) were included in the training and evaluation of machine learning models. By retaining the full feature set, we ensured that potentially valuable information was preserved, particularly given the small sample size. Data augmentation techniques, such as adding noise to the audio samples, were used to improve the robustness of the models and enhance performance, especially for the classification between neurodegenerative and healthy controls, where the original dataset resulted in poor classification performance.

Model Training

In recent years, the trend in digital health care has been to use machine learning models to classify input data (speech samples) into 2 or more classes based on extracted features. In this work, we employed several popular machine learning techniques, such as SVM, DT, RF, and XGBoost [18]. These models were chosen due to their interpretability, robustness, and ability to handle small datasets effectively, which is essential for clinical applications.

SVM, a supervised learning algorithm proposed by Boser et al [32], is grounded in statistical learning theory and is particularly effective for high-dimensional data [33]. It uses hyperplanes and margins to separate data into classes, with its performance being highly dependent on data scaling and the choice of kernel functions. DTs, on the other hand, divide feature space into regions by recursively splitting data and assigning classes to leaf nodes [34]. Despite their simplicity, DTs are prone to overfitting, especially on small datasets.

RFs mitigate this issue by employing an ensemble of DTs trained on bootstrapped datasets, with each tree built using a random subset of features [35]. The final class prediction is based on a majority vote across all trees, which reduces variance and enhances model robustness. Finally, XGBoost, a gradient boosting implementation, constructs DTs sequentially, optimizing performance by correcting errors from previous iterations [36]. It is known for its computational efficiency and scalability, making it a popular choice for structured datasets. For a given sample, the final prediction can be calculated by summing up the scores of overall leaves, which is illustrated in Multimedia Appendix 2.

Given the limited size of our dataset, we prioritized traditional machine-learning models over deep learning methods. While deep learning algorithms have demonstrated exceptional performance on large datasets, their effectiveness diminishes with smaller datasets due to overfitting and computational requirements. Traditional machine learning models, such as SVM and RF, offer superior interpretability, which is critical for clinical decision-making [28]. For instance, the study by Pishgar et al [37] found that on a small voice disorder dataset, SVM outperformed a deep neural network in terms of sensitivity and specificity.

In this study, all 37 extracted features (25 temporal and 12 spectral) were used without any feature selection or filtering. Data augmentation was applied to address the limited sample size, particularly for the neurodegenerative versus healthy dataset, where the augmented dataset improved model performance.

To train and evaluate the machine learning models, we applied a 75 - 25 stratified split of the dataset into training and test sets, ensuring that class distributions were preserved. Stratified 5-fold cross-validation was used to evaluate model performance more reliably, and Grid Search was used to fine-tune hyperparameters for all algorithms.

Evaluation

In this study, we assessed the performance of our classification models using multiple evaluation metrics, with a particular focus on the F_1-score due to its robustness in handling unbalanced datasets. The F_1-score is particularly well-suited for situations where there is an imbalance in the class distribution, as it provides a harmonic mean of precision and recall, balancing the trade-off between these 2 metrics. The F_1-score is defined as follows in Multimedia Appendix 2.

Both precision and recall are crucial in medical applications, where the consequences of false positives or false negatives can

be severe. The F_1-score offers a balanced view of a model's performance when neither precision nor recall can be prioritized over the other. A higher F_1-score (ranging from 0 to 1) indicates a better-performing model.

In addition to the F_1-score, we evaluated our models using precision, recall, and accuracy to provide a comprehensive view of model performance. These metrics helped compare the performance of models across different speech tasks (PaTaKa and Sustained Vowel) and combinations (eg, concussed vs healthy, concussed vs neurodegenerative, neurodegenerative vs healthy). The results section discusses these findings in detail, highlighting the implications of our model's performance for clinical applications.

Ethical Considerations

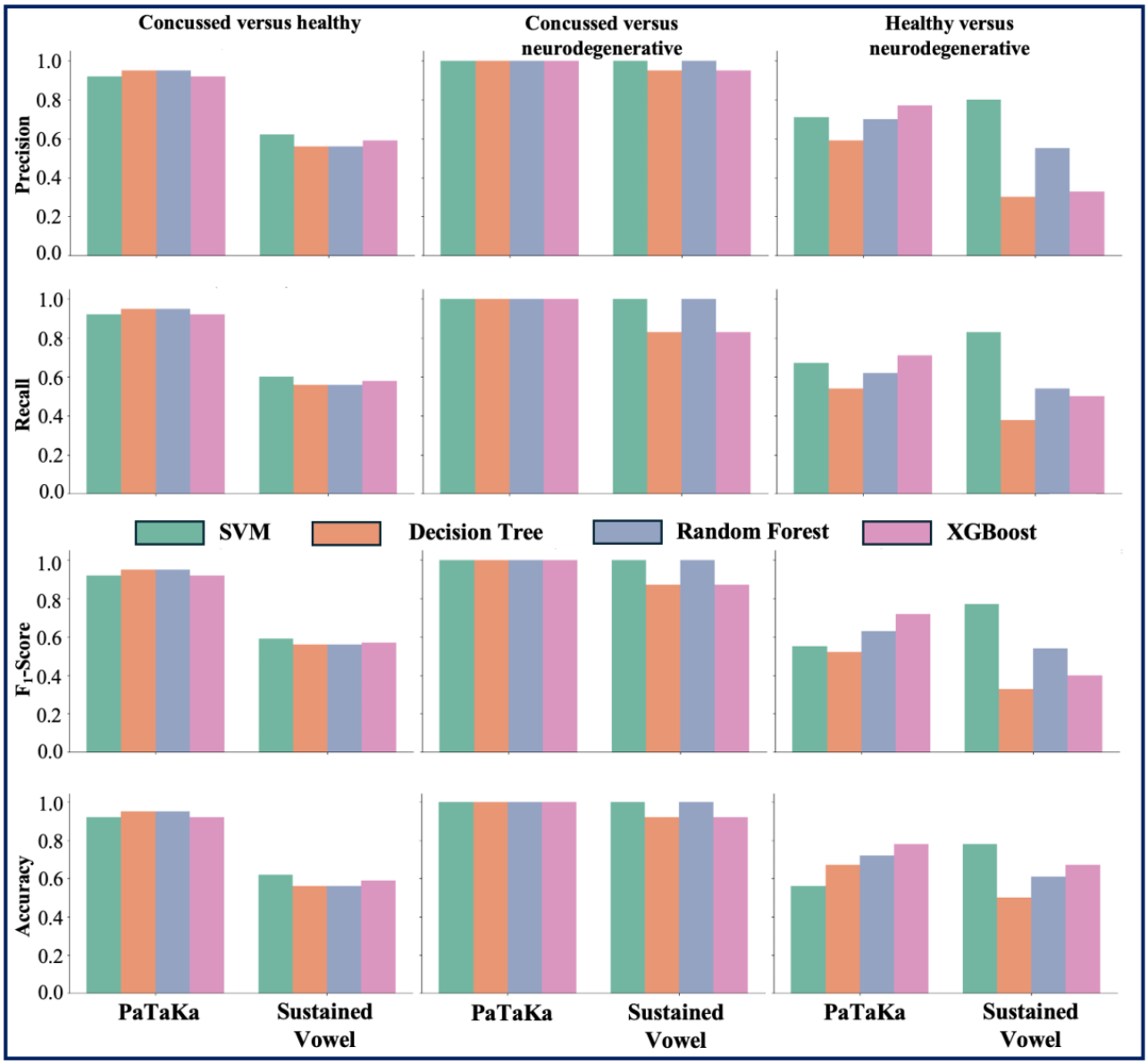
This research was conducted in compliance with ethical standards and approved by the Institutional Review Board at the University of Notre Dame. The approval numbers for this study are 18-01-4338 and 18-01-4340 for PD and concussion, respectively. All participants provided informed consent (Multimedia Appendices 3 and 4), and their confidentiality was ensured throughout the study.

Results

Overview

The performance of the models was evaluated using precision, recall, F_1-score, and accuracy across 3 participant combinations (ie, concussed vs healthy, concussed vs neurodegenerative, and neurodegenerative vs healthy) for 2 widely used speech tasks, PaTaKa and Sustained Vowel. The results provide insights into the discriminative ability of each test and highlight the comparative effectiveness of different classifiers in distinguishing between participant groups. While the PaTaKa task generally demonstrated robust performance across all combinations, the Sustained Vowel test showed varying levels of accuracy, particularly for certain groups and classifiers. The performance for each combination and test, along with discussions on their implications are illustrated in Figure 2.

Figure 2. Performance metrics by test type, model, and combination. SVM: support vector machine; XGBoost: Extreme Gradient Boosting.



Concussed Versus Healthy

PaTaKa Test

The models performed exceptionally well, achieving near-perfect precision, recall, F_1-score, and accuracy across all classifiers. DT and RF slightly outperformed SVM and XGBoost, consistently achieving 0.95. There are no sources in the current document across all metrics. These results highlight the PaTaKa test's robustness in distinguishing between concussed and healthy participants.

Sustained Vowel Test

Performance dropped significantly compared with the PaTaKa test. SVM and XGBoost achieved slightly higher metrics, with F_1-scores around 0.59 - 0.62. DT and RF had the lowest performance, with metrics around 0.56. The reduced performance might indicate that sustained vowels are less effective for distinguishing concussed participants from healthy individuals.

Concussed Versus Neurodegenerative

PaTaKa Test

All models performed perfectly, achieving precision, recall, F_1-score, and accuracy of 1.0. This demonstrates the effectiveness of the PaTaKa test for differentiating concussed participants from those with neurodegenerative conditions. Consistency across all classifiers reinforces the reliability of this task for this combination.

Sustained Vowel Test

Similar to the PaTaKa test, most models achieved perfect scores across all metrics. However, DT and XGBoost showed slightly reduced performance, with F_1-scores of 0.87 and accuracy of 0.92. Despite slight variability, the Sustained Vowel test remains a strong indicator for distinguishing these groups.

Neurodegenerative Versus Healthy

PaTaKa Test

Results varied significantly across classifiers. RF and XGBoost outperformed others, achieving F_1-scores of 0.63 and 0.72, respectively. DT and SVM performed poorly, with F_1-scores around 0.52 - 0.55. These results indicate that the PaTaKa test has moderate effectiveness for this group but requires careful classifier selection.

Sustained Vowel Test

Similar trends were observed. XGBoost achieved the highest F_1-score (0.40) and accuracy (0.67), while other models showed significantly lower performance. This underscores the challenge of distinguishing neurodegenerative participants from healthy controls using sustained vowel tasks.

Feature Set Analysis

Understanding the importance of individual features in classification tasks is crucial for interpreting the predictive power of machine learning models. In this study, we examined feature importance across all tests and combinations to identify the most influential speech features contributing to the classification of concussed, neurodegenerative, and healthy individuals. Feature importance was calculated for each model (SVM, DT, RF, and XGBoost) using a combination of metrics, such as Gini importance, SHAP values, or permutation importance, depending on the model.

To identify globally significant features, we analyzed the frequency of features ranked among the top 5 across all 24 tests. A summary of the top 10 most frequent features is presented in Table 2, while Table 3 provides combination-specific feature importance values. The most frequently identified features were temporal and spectral characteristics, which are known to capture both short-term and long-term speech patterns.

Table . Top 10 most frequent features across all tests.

Rank	Feature	Frequency	Mean importance
1	duration	15	0.29
2	zero_crossing_rate	13	0.33
3	spectral_flatness	12	0.30
4	mfcc[a]	11	0.25
5	spectral_bandwidth	7	0.42
6	spectral_centroid	6	0.07
7	spectral_contrast	5	0.07
8	chroma_stft	5	0.32
9	HNR[b]	4	0.06
10	f4_median	4	0.04

[a]mfcc: mel frequency cepstral coefficient.

[b]HNR: harmonics-to-noise ratio.

Table . Combination specific feature importance value.

Combination and test		Feature	Value
Concussed versus healthy			
	PaTaKa	Duration	1.9
	PaTaKa	Zero-crossing rate	0.47
	Sustained Vowel	Spectral flatness	0.12
Concussed versus neurodegenerative			
	PaTaKa	Spectral bandwidth	1.3
	Sustained Vowel	MFCC[a]	2.9
Neurodegenerative versus healthy			
	PaTaKa	HNR[b]	0.43
	Sustained Vowel	Spectral flatness	0.76

[a]MFCC: mel frequency cepstral coefficient.

[b]HNR: harmonics-to-noise ratio.

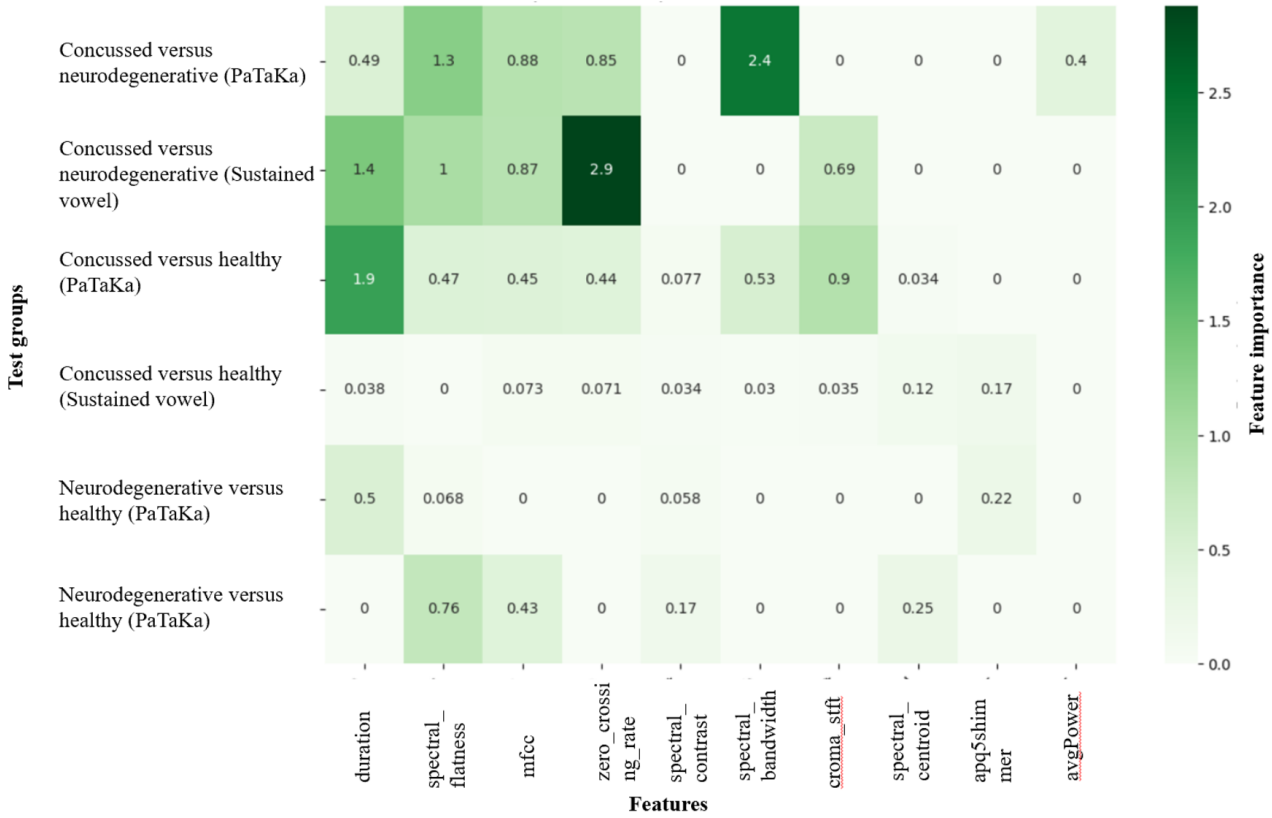
Among the top 10 features, duration, zero-crossing rate, and spectral_flatness were the most influential, appearing consistently across multiple tests and combinations. These features reflect critical aspects of speech production, including articulation rate, periodicity, and frequency smoothness. For instance:

- Duration: This feature provides insights into motor control and speech articulation by measuring the length of utterances.
- Zero-crossing rate: Indicative of voice signal periodicity, this feature is particularly significant in distinguishing voiced and unvoiced speech segments.
- Spectral_flatness: This feature quantifies the uniformity of the speech spectrum, distinguishing between harmonic and noise-like components.

Combination-specific patterns further highlight the variability in feature importance depending on the test (PaTaKa or Sustained Vowel) and the target classification task (concussed vs healthy, concussed vs neurodegenerative, and neurodegenerative vs healthy). For example, (1) in the concussed versus healthy classification, features like mfcc and spectral bandwidth were highly impactful, particularly in the PaTaKa test, (2) in the Concussed concussed versus neurodegenerative classification, spectral_centroid and chroma_stft played a significant role in distinguishing between the 2 groups, and (3) for the neurodegenerative versus healthy classification, features such as f4_median and HNR were key discriminators, particularly in the Sustained Vowel test.

The distribution of feature importance values across combinations and tests is visualized in Figure 3, while the detailed numerical values for each combination and test are available in Table 3. These findings emphasize the variability of feature contributions across different tasks and highlight the importance of task-specific feature analysis for robust classification.

Figure 3. Top 10 most frequent features across all tests. mfcc: mel frequency cepstral coefficients.



Discussion

Principal Findings

The findings of this study provide valuable insights into the use of speech-based features for differentiating between neurodegenerative conditions, particularly mTBI (concussions) and neurodegenerative diseases (eg, PD). By leveraging 2 commonly used speech tasks, the PaTaKa test and the Sustained Vowel test, and a variety of machine learning models, we achieved classification accuracies ranging from 60% to 90%, with RF and XGBoost models consistently outperforming others. In addition, we identified key speech features, such as duration, zero-crossing rate, and spectral flatness, as critical biomarkers for distinguishing between these conditions. These results underscore the potential of speech features as noninvasive biomarkers for neurodegenerative health assessment and highlight the complementary roles of the PaTaKa and Sustained Vowel tests in revealing task-specific and globally significant features.

Key Observations

First, task-specific performance. The PaTaKa test consistently outperformed the Sustained Vowel test across all combinations. This may be attributed to the sequential articulatory movements required in the PaTaKa test, which can better capture subtle motor and speech deficits. For example, in the concussed versus healthy classification, F_1-scores for PaTaKa exceeded 0.9 across all models, whereas the Sustained Vowel test achieved F_1-scores below 0.6 for the same classification. These findings highlight the importance of task selection in speech analysis and suggest that diadochokinetic tasks may provide richer diagnostic information.

Second, model-specific trends. Among the machine learning models, RF and XGBoost consistently performed well, demonstrating their ability to handle complex, nonlinear relationships in speech data. This aligns with previous research highlighting the robustness of ensemble learning methods in biomedical and speech signal processing tasks [38].

Third, the high interpretability of DTs also provides an advantage for clinical applications, particularly in scenarios where transparency is critical for adoption in health care settings.

Fourth, despite its slightly lower performance in some scenarios, DT models remain valuable due to their simplicity and ease of implementation.

Fifth, interestingly, SVMs displayed strong performance in balanced datasets, particularly in the concussed versus neurodegenerative classification, where precision and recall consistently reached 1.0 for the PaTaKa test. This finding is consistent with previous studies showing that SVMs are effective for high-dimensional data, especially when datasets are carefully preprocessed and balanced [39]. The performance of SVM in this classification task further underscores its utility in distinguishing nuanced differences between distinct neurodegenerative conditions using speech features.

Finally, feature importance. The analysis of feature importance revealed that a small subset of features consistently played a dominant role across tests and combinations. Temporal features such as duration and zero-crossing rate were particularly influential, likely reflecting disruptions in motor control and

speech rhythm caused by both concussions and neurodegenerative conditions. Spectral features, including spectral_flatness, mfcc, and spectral_bandwidth, were also critical, highlighting their utility in capturing frequency-domain variations associated with speech pathologies. These results align with previous research, which has emphasized the role of both temporal and spectral features in detecting neurodegenerative impairments.

Comparison With Previous Studies

Our findings corroborate and extend existing literature on speech-based biomarkers for neurodegenerative conditions. Previous research has demonstrated the utility of features such as MFCC and jitter for detecting PD [4], as well as features like zero-crossing rate and duration for identifying concussions [19]. However, this study uniquely emphasizes the differentiation between neurodegenerative diseases like PD and mild traumatic brain injury (eg, concussions), a task that remains relatively underexplored in existing literature.

Furthermore, the inclusion of both PaTaKa and Sustained Vowel tests enables a more comprehensive analysis of task-specific feature relevance. While previous studies have evaluated the diagnostic utility of individual speech tasks (eg, sustained phonation for ALS in studies by Allison et al [13] and Tsanas et al [27]), this work highlights how combining multiple tasks can reveal unique and complementary insights into speech biosignatures associated with diverse neurodegenerative conditions.

In addition to confirming the significance of widely used features such as spectral flatness and zero-crossing rate, our study identifies new combinations of features, including spectral contrast and chroma-based features, as being critical for distinguishing between these groups. These results align with recent advancements in the field, where ensemble learning models, such as RF and XGBoost, are increasingly used to capture the intricate, nonlinear relationships within speech data [23].

By addressing age-related variability and introducing data augmentation to mitigate the challenges of limited datasets, this study not only validates previously established findings but also sets the stage for future research aimed at improving the diagnostic accuracy of speech-based assessments across distinct but potentially overlapping neurodegenerative conditions.

Implications for Clinical Practice

The results of this study highlight several practical implications for clinical applications.

First, noninvasive diagnostics. The reliance on speech features, which can be collected using readily available devices like smartphones, opens up possibilities for remote and noninvasive diagnostics. This is particularly valuable in resource-constrained settings where access to advanced imaging or neurophysiological tests may be limited.

Second, early detection. The ability to detect subtle speech impairments associated with neurodegenerative conditions could enable earlier diagnosis, allowing for timely interventions.

Finally, task selection. The superior performance of the PaTaKa test suggests that it should be prioritized in future speech-based diagnostic protocols, particularly for distinguishing between concussions and neurodegenerative conditions.

Limitations

Despite the promising results, there are several limitations to this study.

First, small dataset—the dataset size, particularly for neurodegenerative diseases, was relatively small. This may limit the generalizability of the findings to larger, more diverse populations.

Second, demographic differences—the age gap between the concussed (younger) and neurodegenerative (older) populations poses a potential confounding factor. While age-matched healthy controls were included, the results could be influenced by inherent age-related differences in speech production.

Third, feature engineering and contextual factors—while the study identified important features, the reliance on manual feature extraction may overlook nuanced patterns. Advanced techniques, such as deep learning–based feature discovery, could reveal hidden characteristics in speech data. Future research should also account for comorbidities and age-related factors, as these can influence speech biosignatures and potentially confound results. Age-normalized datasets and statistical adjustments can further enhance the robustness of classification models.

Future Directions

This study demonstrates the potential of speech-based features to differentiate between concussed, neurodegenerative, and healthy individuals. While promising, the findings also highlight several areas for improvement and expansion, which we aim to address in future work.

First, dataset expansion and diversity. The current dataset includes limited samples from each group, particularly for neurodegenerative diseases. Future studies will expand the dataset to include larger and more diverse populations, ensuring broader generalizability of the results. In addition, we aim to achieve a more balanced age distribution across all participant groups, enabling more robust analyses and minimizing potential biases.

Second, age-related effects. While we mitigated some confounding effects of age by including 2 distinct healthy control groups (age-matched for concussed and neurodegenerative participants), future studies will incorporate more advanced strategies to address age-related variations in speech features. These include (1) explicitly including age as a covariate in statistical models to control its effects and quantify its influence on the results, (2) conducting age-matched subgroup analyses to validate that classification performance is not driven by age-related biases but by the underlying neurodegenerative conditions, and (3) expanding the dataset to improve the representation of younger and older age groups across all conditions.

Third, feature engineering and discovery. While this study focused on predefined temporal and spectral features, advanced deep learning models such as autoencoders or transformer-based models could uncover latent features that may better distinguish between neurodegenerative conditions. In addition, further exploration of task-specific feature relevance could reveal complementary insights into speech patterns for different health conditions.

Fourth, longitudinal data analysis. Future work should explore longitudinal data to track changes in speech biosignatures over time. This would help identify temporal patterns associated with disease progression and recovery, providing valuable insights for monitoring treatment efficacy and early diagnosis.

Fifth, integration with clinical practice. To enhance the clinical utility of this research, future efforts should focus on integrating speech-based diagnostic tools into real-world health care settings. This includes (1) developing user-friendly mobile apps or web applications for noninvasive speech analysis and (2) collaborating with clinicians to validate the models and evaluate their effectiveness in clinical decision making processes.

Finally, evaluation metrics and benchmarking. Expanding the evaluation metrics to include area under the receiver operating characteristic curve and precision-recall curves would provide a more comprehensive understanding of model performance. In addition, benchmarking against existing speech-based models or alternative diagnostic tools could further contextualize the findings and demonstrate the added value of the proposed methods.

By addressing these areas, future research can build upon the findings of this study to further advance the field of speech analysis in neurodegenerative health, improve diagnostic accuracy, and pave the way for noninvasive, scalable diagnostic tools.

Conclusion

This study demonstrates the potential of speech features, particularly those derived from the PaTaKa test, as effective biomarkers for distinguishing between concussed, neurodegenerative, and healthy individuals. By identifying task-specific and globally important features, the findings lay the groundwork for developing noninvasive, speech-based diagnostic tools that can be readily implemented in clinical practice. Further research addressing the study's limitations could pave the way for broader applications of speech analysis in neurodegenerative health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Feature description.
[DOCX File, 17 KB - neuro_v4i1e64624_app1.docx]

Multimedia Appendix 2
Equations.
[DOCX File, 15 KB - neuro_v4i1e64624_app2.docx]

Multimedia Appendix 3
Consent form for participants with neurodegenerative conditions.
[PDF File, 76 KB - neuro_v4i1e64624_app3.pdf]

Multimedia Appendix 4
Consent form for participants with concussions.
[PDF File, 76 KB - neuro_v4i1e64624_app4.pdf]

References

1. Templeton JM, Poellabauer C, Schneider S. Enhancement of neurocognitive assessments using smartphone capabilities: systematic review. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15517. [doi: 10.2196/15517] [Medline: 32442150]
2. Ramanarayanan V, Lammert AC, Rowe HP, Quatieri TF, Green JR. Speech as a biomarker: opportunities, interpretability, and challenges. Perspect ASHA SIGs 2022 Feb 11;7(1):276-283. [doi: 10.1044/2021_PERSP-21-00174]
3. Robin J, Harrison JE, Kaufman LD, Rudzicz F, Simpson W, Yancheva M. Evaluation of speech-based digital biomarkers: review and recommendations. Digit Biomark 2020;4(3):99-108. [doi: 10.1159/000510820] [Medline: 33251474]
4. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 2009 Apr;56(4):1015-1022. [doi: 10.1109/TBME.2008.2005954] [Medline: 21399744]
5. Gan L, Cookson MR, Petrucelli L, La Spada AR. Converging pathways in neurodegeneration, from genetics to mechanisms. Nat Neurosci 2018 Oct;21(10):1300-1309. [doi: 10.1038/s41593-018-0237-7] [Medline: 30258237]

6. Kiaei M. New hopes and challenges for treatment of neurodegenerative disorders: great opportunities for young neuroscientists. Basic Clin Neurosci 2013;4(1):3-4. [Medline: 25337322]
7. Doherty C. ALS vs Parkinson's: what are the differences? Verywell Health. 2024. URL: https://www.verywellhealth.com/als-vs-parkinsons-6826333 [accessed 2025-02-06]
8. Youdim MB, Riederer P. Understanding Parkinson's disease. Sci Am 1997 Jan;276(1):52-59. [doi: 10.1038/scientificamerican0197-52] [Medline: 8972618]
9. Shively S, Scher AI, Perl DP, Diaz-Arrastia R. Dementia resulting from traumatic brain injury: what is the pathology? Arch Neurol 2012 Oct;69(10):1245-1251. [doi: 10.1001/archneurol.2011.3747] [Medline: 22776913]
10. Delaney JS, Abuzeyad F, Correa JA, Foxford R. Recognition and characteristics of concussions in the emergency department population. J Emerg Med 2005 Aug;29(2):189-197. [doi: 10.1016/j.jemermed.2005.01.020] [Medline: 16029831]
11. O'Sullivan SB, Schmitz TJ, Fulk G. Physical Rehabilitation: FA Davis; 2019.
12. Penttilä N, Korpijaakko-Huuhka AM, Kent RD. Disfluency clusters in speakers with and without neurogenic stuttering following traumatic brain injury. J Fluency Disord 2019 Mar;59:33-51. [doi: 10.1016/j.jfludis.2019.01.001] [Medline: 30641458]
13. Allison KM, Yunusova Y, Campbell TF, Wang J, Berry JD, Green JR. The diagnostic utility of patient-report and speech-language pathologists' ratings for detecting the early onset of bulbar symptoms due to ALS. Amyotroph Lateral Scler Frontotemporal Degener 2017 Aug;18(5-6):358-366. [doi: 10.1080/21678421.2017.1303515] [Medline: 28355886]
14. Aich S, Kim HC, younga K, Hui KL, Al-Absi AA, Sain M. A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of Parkinson's disease. Presented at: 2019 21st International Conference on Advanced Communication Technology (ICACT); Feb 17-20, 2019; PyeongChang, South Korea. [doi: 10.23919/ICACT.2019.8701961]
15. Duffy JR. Motor Speech Disorders: Clues to Neurologic Diagnosis: Humana Press; 2000:35-53. [doi: 10.1007/978-1-59259]
16. Smith R, Chepisheva M, Cronin T, Seemungal BM. Chapter 16 - diagnostic approaches techniques in concussion/mild traumatic brain injury: where are we. In: Hoffer ME, Balaban CD, editors. Neurosensory Disorders in Mild Traumatic Brain Injury: Academic Press; 2019:247-277. [doi: 10.1016/B978-0-12-812344-7.00016-9]
17. Patel S, Grabowski C, Dayalu V, Cunningham M, Testa AJ. Fluency changes due to sports-related concussion. medRxiv. Preprint posted online on Sep 23, 2021. [doi: 10.1101/2021.09.19.21263791]
18. Daudet L, Yadav N, Perez M, Poellabauer C, Schneider S, Huebner A. Portable mTBI assessment using temporal and frequency analysis of speech. IEEE J Biomed Health Inform 2017 Mar;21(2):496-506. [doi: 10.1109/JBHI.2016.2633509] [Medline: 27913365]
19. Tsanas A. New insights into Parkinson's disease through statistical analysis of standard clinical scales quantifying symptom severity. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3412-3415. [doi: 10.1109/EMBC.2019.8856559] [Medline: 31946612]
20. Julien JP. ALS: astrocytes move in as deadly neighbors. Nat Neurosci 2007 May;10(5):535-537. [doi: 10.1038/nn0507-535] [Medline: 17453052]
21. Magrinelli F, Picelli A, Tocco P, et al. Pathophysiology of motor dysfunction in Parkinson's disease as the rationale for drug treatment and rehabilitation. Parkinsons Dis 2016;2016:9832839. [doi: 10.1155/2016/9832839] [Medline: 27366343]
22. Benba A, Jilbab A, Hammouch A. Voice analysis for detecting patients with Parkinson's disease using the hybridization of the best acoustic features. Int J Electr Eng Inform 2016 Mar 30;8(1):108-116. [doi: 10.15676/ijeei.2016.8.1.8]
23. Mallela J, Illa A, Suhas BN, et al. Voice based classification of patients with amyotrophic lateral sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning. Presented at: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 4-8, 2020; Barcelona, Spain. [doi: 10.1109/ICASSP40776.2020.9053682]
24. Rueda A, Krishnan S. Clustering Parkinson's and age-related voice impairment signal features for unsupervised learning. Adv Data Sci Adapt Data Anal 2018 Apr;10(2):1840007. [doi: 10.1142/S2424922X18400077]
25. Tsanas A, Arora S. Biomedical speech signal insights from a large scale cohort across seven countries: the Parkinson's voice initiative study. 2019 Presented at: 11th International Workshop, Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA); Dec 17-19, 2019; Firenze, Italy.
26. Vashkevich M, Azarov E, Petrovsky A, Rushkevich Y. Features extraction for the automatic detection of ALS disease from acoustic speech signals. Presented at: 2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA); Sep 19-21, 2018; Poznan, Poland. [doi: 10.23919/SPA.2018.8563414]
27. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271. [doi: 10.1109/TBME.2012.2183367] [Medline: 22249592]
28. Tsanas A, Arora S. Assessing Parkinson's Disease Speech Signal Generalization of Clustering Results across Three Countries: Findings in the Parkinson's Voice Initiative Study: SCITEPRESS - Science and Technology Publications; 2021:124-131. [doi: 10.5220/0010383001240131]
29. Bongioanni P. Paralysis in ALS patients: an overview on assessment and treatment of motor impairment. Int J Med Biol Front 2012;18(2/3):147.

30. Wall C, Powell D, Young F, et al. A deep learning-based approach to diagnose mild traumatic brain injury using audio classification. PLoS One 2022;17(9):e0274395. [doi: 10.1371/journal.pone.0274395] [Medline: 36170287]
31. McFee B, Raffel C, Liang D, et al. librosa: audio and music signal analysis in Python. Presented at: Python in Science Conference; Jul 6-12, 2015; Austin, Texas. [doi: 10.25080/Majora-7b98e3ed-003]
32. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Presented at: COLT92; Jul 27-29, 1992; Pittsburgh, PA, United States. [doi: 10.1145/130385.130401]
33. Vapnik V. The Nature of Statistical Learning Theory: Springer Science Business Media; 2013.
34. Rokach L, Maimon O. Data Mining with Decision Trees: Theory and Applications, 2nd edition 2014, Vol. 81.
35. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
36. Abdurrahman G, Sintawati M. Implementation of XGBoost for classification of Parkinson's disease. J Phys: Conf Ser 2020 May 1;1538(1):012024. [doi: 10.1088/1742-6596/1538/1/012024]
37. Pishgar M, Karim F, Majumdar S, Darabi H. Pathological voice classification using mel-cepstrum vectors and support vector machine. arXiv. Preprint posted online on Dec 19, 2018. [doi: 10.48550/arXiv.1812.07729]
38. Difference between Random Forest and XGBoost. Geeks for Geeks. 2024. URL: https://www.geeksforgeeks.org/difference-between-random-forest-vs-xgboost/ [accessed 2025-02-06]
39. Jain M, Narayan S, Balaji P, P BK. Speech emotion recognition using support vector machine. arXiv. Preprint posted online on Feb 3, 2020. [doi: 10.48550/arXiv.2002.07590]

Abbreviations

ALS: amyotrophic lateral sclerosis
DT: decision tree
HNR: harmonics-to-noise ratio
ImPACT: Immediate Post-Concussion Assessment and Cognitive Testing
MDS-UPDRS: Movement Disorder Society - Unified Parkinson's Disease Rating Scale
MFCC: mel frequency cepstral coefficient
mTBI: mild traumatic brain injury
PD: Parkinson disease
RF: random forest
SAC: Standardized Assessment of Concussion
SCAT: Sport Concussion Assessment Tool
SVM: support vector machine
XGBoost: Extreme Gradient Boosting

Edited by P Kubben; submitted 22.07.24; peer-reviewed by H Rajaguru, M Gasmi, R Norel, S Mao; revised version received 26.11.24; accepted 08.01.25; published 12.02.25.

Please cite as:
Rubaiat R, Templeton JM, Schneider SL, De Silva U, Madanian S, Poellabauer C
Exploring Speech Biosignatures for Traumatic Brain Injury and Neurodegeneration: Pilot Machine Learning Study
JMIR Neurotech 2025;4:e64624
URL: https://neuro.jmir.org/2025/1/e64624
doi:10.2196/64624

Exploring Remote Monitoring of Poststroke Mood With Digital Sensors by Assessment of Depression Phenotypes and Accelerometer Data in UK Biobank: Cross-Sectional Analysis

Stephanie J Zawada[1], PhD, MS; Ali Ganjizadeh[2], MD; Gian Marco Conte[2], MD, PhD; Bart M Demaerschalk[1,3], MD, MSc; Bradley J Erickson[2,4], MD, PhD

[1]Mayo Clinic College of Medicine and Science, 5777 E. Mayo Boulevard, Phoenix, AZ, United States

[2]AI Laboratory, Department of Radiology, Mayo Clinic, Rochester, MN, United States

[3]Department of Neurology, Division of Cerebrovascular Diseases, Mayo Clinic, Phoenix, AZ, United States

[4]Mayo Clinic College of Medicine and Science, Rochester, MN, United States

Corresponding Author:
Stephanie J Zawada, PhD, MS
Mayo Clinic College of Medicine and Science, 5777 E. Mayo Boulevard, Phoenix, AZ, United States

Abstract

Background: Interest in using digital sensors to monitor patients with prior stroke for depression, a risk factor for poor outcomes, has grown rapidly; however, little is known about behavioral phenotypes related to future mood symptoms and if patients with and without previously diagnosed depression experience similar phenotypes.

Objective: This study aimed to assess the feasibility of using digital sensors to monitor mood in patients with prior stroke with a prestroke depression diagnosis (DD) and controls. We examined relationships between physical activity behaviors and self-reported depression frequency.

Methods: In the UK Biobank wearable accelerometer cohort, we retrospectively identified patients who had previously suffered a stroke (N=1603) and conducted cross-sectional analyses with those who completed a subsequent depression survey follow-up. Sensitivity analyses assessed a general population cohort excluding previous stroke participants and 2 incident cohorts: incident stroke (IS) and incident cerebrovascular disease (IC).

Results: In controls, the odds of being in a higher depressed mood frequency category decreased by 23% for each minute spent in moderate - to - vigorous physical activity (odds ratio 0.77, 95% CI 0.69 - 0.87; *P*<.001). This association persisted in both general cohorts and in the IC control cohort.

Conclusions: Although moderate - to - vigorous physical activity was linked with less frequent depressed mood in patients with prior stroke without DD, this finding did not persist in DDs. Thus, accelerometer-mood monitoring may provide clinically useful insights about future mood in patients with prior stroke without DDs. Considering the finding in the IC cohort and the lack of findings in the IS cohorts, accelerometer-mood monitoring may also be appropriately applied to observing broader cerebrovascular disease pathogenesis.

(JMIR Neurotech 2025;4:e56679) doi:10.2196/56679

KEYWORDS

depression; cerebrovascular disease; remote monitoring; stroke; accelerometers; mobile phone

Introduction

Overview

Depression is an established risk factor for poor outcomes after a stroke and transient ischemic attack (TIA), including subsequent stroke and other cerebrovascular diseases (CeVDs) [1,2]. Although poststroke depression (PSD) affects roughly one-third of patients with stroke, screening for depression in patients after a stroke is not routine, with less than 10% of patients with stroke screened [3]. Furthermore, it remains unclear when follow-up PSD screening should occur, as current research suggests that not all patients will experience PSD symptoms immediately after a stroke and, for those who do, the majority will experience recurrent depression episodes in the years after a stroke [4]. A reason for this gap in screening is the shortage of neurologists, particularly those with diagnostic training in identifying PSD [5]. Accelerated by the widespread adoption of personal mobile devices, from computers to smartwatches, it is critical to investigate the potential of such devices to collect meaningful data outside of clinical settings, aiding clinicians in identifying depressed mood in patients with

stroke—and, potentially, those most at risk for subsequent stroke and CeVDs [6].

Background

The prevalence of PSD remains unknown, partly due to its heterogeneous nature, spanning unique somatic, behavioral, cognitive, motivational, and emotional components [7]. The severity of its manifestation ranges from mild symptoms to clinical-grade depression, the former of which relies on self-reported scoring methods inherently subject to bias, especially in patients with cognitive impairment for whom self-reported surveys may not be reliable [8]. Although clinician-administered assessments, like the Montgomery–Åsberg Depression Rating Survey (MADRS), offer gold-standard assessments of symptoms, nurse and physician shortages complicate the routine administration of such instruments [9].

In some survivors, depression may emerge alongside the incipient pathogenesis of cerebrovascular dysfunction, while for others, depression may be a reaction to being conscious of cognitive impairment or the putative manifestation of silent cerebral infarcts [10,11]. As such, individual depression phenotypes may vary greatly across survivors with identical survey summary scores. Although investigations into the associations between stroke location within the brain and self-reported depression survey scores have yielded inconclusive results, a recent cross-sectional study of patients with prior stroke (n=200) found that symptoms assessed by MADRS correlate with specific macrostructural characteristics [12]. Considering that clinician-administered assessments, like MADRS, are more accurate than self-reported survey scores in patients with prior stroke, the need for a modified approach to monitoring patients with stroke for depression emerges.

In recent years, objective data from portable and wearable sensors have demonstrated the feasibility of augmenting self-reported mood surveys outside of clinics, a promising approach for monitoring patients with symptomatic and asymptomatic deteriorating brain health outside of standardized, clinical environments [13-20]. In addition, accelerometer measures of behavior have established a difference in PA engagement stratified by depression severity, highlighting the need for a thoughtful approach to PSD screening and monitoring that ensures patients with emerging or mild depression symptoms, unlike those with previous documented depressive episodes, are not neglected [21].

While triaging patients with PSD for preventative intervention could yield clinically meaningful functional recovery outcomes, the potential of such an approach for preventing future CeVD diagnoses remains to be seen. Numerous studies have found that depressive symptoms are associated with an increased risk of subsequent CeVD, from acute CeVDs, like stroke and TIA, to more chronic conditions, like cerebral arterial stenosis and vascular dementia [11,22-24]. Furthermore, recent research suggests daily functioning and cognitive changes may be observable up to 10 years before some types of CeVD [25]. Thus, particular attention should be paid to behavioral patterns in patients with PSD to elucidate phenotypes with predictive potential for functional outcomes and neurologic disorders.

Previous Work

Blending self-reported assessments of phenomena, like mood, recorded through web browsers and smartphone apps, with passive sensor data, like that from wearable accelerometers, is gaining popularity in real-world settings [26,27]. Numerous pilot studies have demonstrated the potential for wearable and minimally invasive sensors to detect neurologic conditions; however, these tools have neither been validated in population cohorts nor combined with survey sampling of mood [28].

Early-stage evidence suggests that monitoring lifestyle behavior and mood in PSD is feasible [29-31]. The results of a small longitudinal study (n=40) suggest that self-reported moderate-to-vigorous physical activity (MVPA) before stroke is associated with improved mobility and self-care as well as decreased discomfort after stroke [32]. While the study did not sample mood outside of clinical environments, Reinholdsson et al [33] used self-report surveys to expand on the above findings, demonstrating that patients who engage in higher levels of prestroke physical activity (PA) experienced less severe PSD compared with patients who were physically inactive .

In addition, current literature on accelerometers in PSD suggests that distinct behavioral patterns may identify patients with depression within the first year after a stroke. In a 2022 prospective observational study of recently discharged patients with minor ischemic stroke (n=76), participants wore accelerometers in-hospital for 1 week. Analyses revealed that only increased sedentary behavior (SB) and reduced light physical activity (LPA) were linked with more intense depression, assessed through a written Geriatric Depression Scale survey, 3 months after hospitalization in this older adult cohort [34]. In a small pilot study (n=40) of stroke survivors, MVPA was linked with positive mood [35]. Although extensive research has confirmed links between sleep disorders and both depression and incident CeVD (IC), no research has observed both depressive symptoms and objectively measured sleep after stroke [36,37]. Furthermore, no previous accelerometer research into PSD beyond the first year of stroke recovery has been published.

Goal of This Study

The goal of this study is twofold: first, to investigate potential associations between objectively measured behavior and future depression frequency in patients with prior stroke assessed by a remote approach and second, to explore whether that association varies between patients with prior stroke with a prestroke depression diagnosis (DDs) and those without (controls).

We conducted a cross-sectional analysis with the UK Biobank (UKBB), the most extensive lifestyle and mood cohort to date, assessing the relationships between accelerometer-measured sleep, SB, LPA, and MVPA and a subsequent depression descriptor (depressed mood frequency). Given that depression before stroke may yield behavioral phenotypes distinct from those emergent in participants without a prestroke depression diagnosis, we created 2 cohorts of patients with prior stroke: those with a clinical depression diagnosis before stroke and those without. As this analysis focuses on participants who may

develop or have undiagnosed PSD, participants whose PSD diagnosis was recorded were excluded. Adjusting for age, sex, ethnicity, multiple relevant comorbidities, and time elapsed between accelerometer monitoring and depression survey submission, we hypothesized that increased LPA and MVPA time would be associated with a reduction in the odds of being in a more frequent depressed mood category while increased SB time would be associated with a rise in the odds of being in a more frequent depressed mood category. Considering the established relationship between sleep and depressed mood, we created a binary variable (yes or no) for guideline-recommended sleep (7 - 9 h/d). We hypothesized that guideline-recommended sleep would be associated with a reduction in the odds of being in a more frequent depressed mood category.

Methods

Recruitment

The UKBB enrolled middle-aged (40 - 69 y) participants (N=502,364) at 22 assessment centers across the United Kingdom at a baseline assessment (2006 - 2010), which included in-person interviews, touchscreen surveys, and physical examinations to extract lifestyle and environmental data used in this study. Although all baseline participants (n=502,151) were invited, only 72,652 enrolled in the 1-week accelerometer study (2013 - 2015) and completed the depression frequency survey (2016 - 2017). Hospital and other diagnostic registries were linked to enrolled participants.

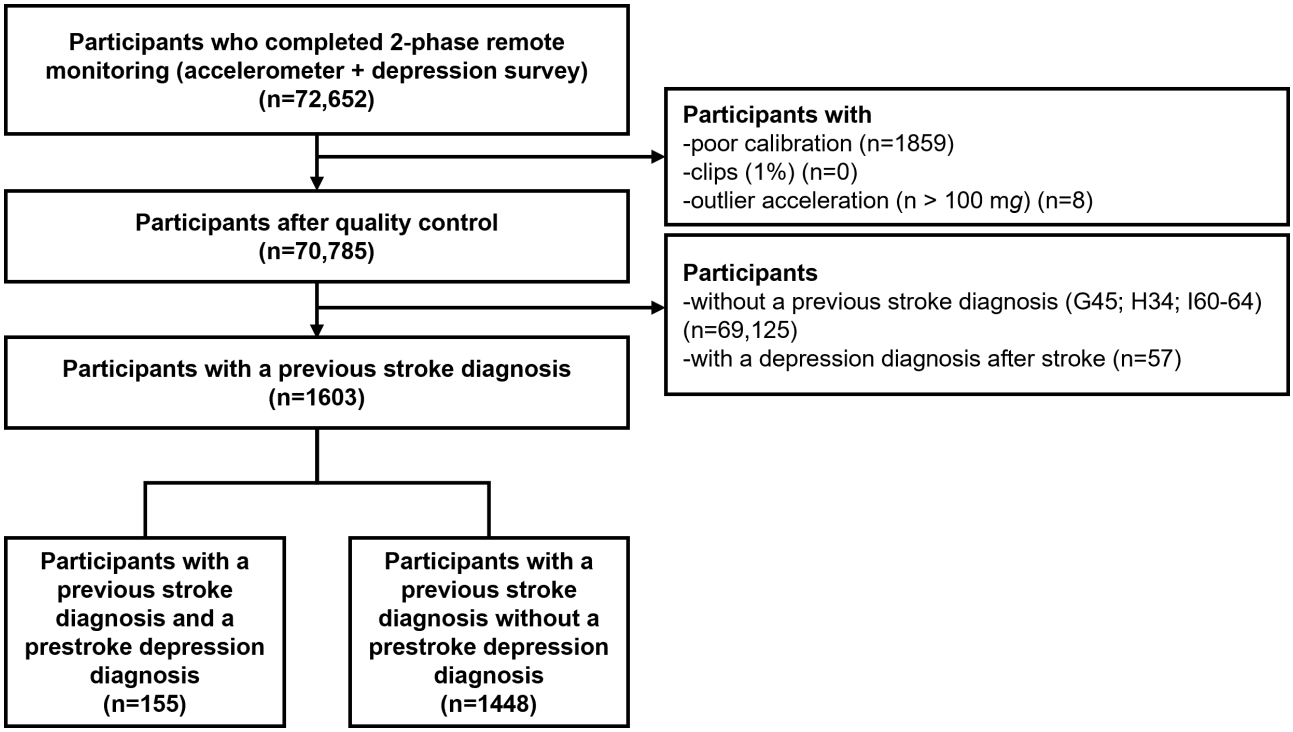
Participant Cohorts

Among participants who completed both remote monitoring components, those with dementia (n=23) were excluded. Quality control filtering demonstrated by Madjedi et al [38] was applied (n=70,785), which excluded those with outlier acceleration (>100 mg), more than 1% of readings exceeding ±8 *g* (clips), accelerometer wear time less than 3 days, and missing data for at least one 60-minute interval throughout 24-hour periods. Only participants with a previous stroke, including ischemic stroke, hemorrhagic stroke, and TIA (G45), were included (n=1660). Retinal artery occlusion (H34) was included as a stroke, as it is now considered a type of acute ischemic stroke [39]. Participants who were diagnosed with depression after stroke but before the accelerometer study (n=57) were excluded.

Among those meeting the inclusion criteria (n=1603), participants were divided into two cohorts: (1) those with a prestroke depression diagnosis at accelerometer study commencement (n=155) and (2) controls, that is, those without a prestroke depression diagnosis (n=1448) (Figure 1). No participants were diagnosed with depression between the accelerometer study and the follow-up depression survey.

Participants with a history of depression (*International Classification of Diseases, Tenth Revision* [*ICD-10*] codes F32-39) comprised the depression diagnosis (DDs) cohort. Definitions (*ICD-10* codes) used for inclusion and exclusion criteria as well as diagnostic classification are available in Multimedia Appendix 1.

Figure 1. Classification algorithm for participant cohorts.



Data Collection

Accelerometer study participants were instructed to wear the Axivity AX3 commercial accelerometer wristwatch continuously on their dominant arm for 1 week. The depressed mood frequency question was administered through a link accessible on smartphone, tablet or PC browsers as part of the standardized Patient Health Questionnaire-2 (PHQ-2) survey: “Over the past two weeks, how often have you felt down, depressed, or hopeless?” Responses were ordinal scores indicating the

frequency of depressed mood, with 1="Not at all"; 2="Several days"; 3="More than half of days"; and 4="Nearly every day."

Permanent covariables were obtained at baseline visit, including sex and ethnicity. For each participant, age at the time of accelerometer study was calculated. Time-to-assessment was individually calculated by subtracting the accelerometer start date from the date of submitting the depressed mood survey. Comorbidity diagnoses before the accelerometer study were obtained from linked patient and hospital databases.

Statistical Analysis

To compare continuous and categorical covariables, the Mann-Whitney U test and χ^2 test, respectively, were used. A cross-sectional analysis using ordinal logistic regression to investigate the association between objective behavior predictors and the ordinal outcome variable, depressed mood frequency over the past 2 weeks, was conducted on data obtained at the accelerometer study and remote follow-up survey.

For both DD and control cohorts, separate models were fitted to evaluate whether the role of objective behavior predictors in depressed mood frequency differed between cohorts.

Analyses were performed in R (R Foundation for Statistical Computing), using *polr* from the library MASS. The effect sizes of objective behavior predictors, adjusted for confounders, on depressed mood frequency were plotted as odds ratios with 95% CIs. The Likelihood Ratio Test was used to obtain all *P* values and associated CIs. $P<.05$ was statistically significant.

Sensitivity Analysis

Three sensitivity analyses (also using ordinal logistic regression models), each considering DDs and controls, were performed using UKBB data. First, a general population dataset wasf generated. This included all participants eligible for inclusion in the accelerometer study and follow-up depression frequency survey who did not have a previous stroke diagnosis.

Next, participants with an initial IC diagnosis (after the depression frequency survey) were filtered into a separate dataset. Ordinal logistic regression models were fitted to assess the relationships between objective behavior predictors and depressed mood frequency. Finally, participants in the IC cohort who had an IS diagnosis were filtered into a separate dataset, and ordinal logistic regression models were fitted to assess the target relationship. The investigation of IC as a composite end point reflects updated understanding of stroke as sharing etiology with other neurologic rather than circulatory system disorders, as defined in the most recent *International Classification of Diseases, Eleventh Revision* (*ICD-11*) [37].

For each filtered cohort, sample characteristics were obtained for review.

Ethical Considerations

National Health Service Research Ethics Committee (11/NW/0382) granted ethical approval for the UKBB population cohort study. Informed consent was obtained from all UK Biobank participants under National Health Service National Research Ethics Service (Ref 11/NW/0382). All UKBB data are deidentified.

Results

Study Characteristics

For participants in the 2-stage remote monitoring study (Table 1), the DDs had a higher proportion of women compared with controls (58.7% vs 40.7%). On average, DDs were younger (64 vs 66 y), slept slightly longer (9.2 vs 9.0 h/d), spent slightly less time in MVPA (29.3 vs 37.3 min/d) and SB (580.1 vs 583.6 min/d), and spent slightly more time in LPA (281.4 vs 278.2 min/d).

Table . Baseline characteristics of patients with previous stroke.

		Prestroke depression	Controls	*P* value
Number of participants, n		155	1448	
Age, mean (SD)		64 (7)	66 (6.5)	<.001
Gender, n (%)				
	Men	64 (41.3)	859 (59.3)	<.001
Race, n (%)				
	White	153 (98.7)	1418 (97.9)	.68
Sleep, mean (SD)		9.2 (1.8)	9.0 (1.8)	<.001
Sleep (7 - 9 h/d), n (%)		71 (45.8)	736 (50.8)	.27
SB[a], mean (SD)		580.1 (114.4)	583.6 (112.8)	<.001
LPA[b], mean (SD)		281.4 (106.9)	278.2 (102.4)	<.001
MVPA[c], mean (SD)		29.3 (31.2)	37.3 (33.0)	<.001
Time-to-assessment, mean (SD)		1.8 (0.7)	1.8 (0.6)	<.001
Diabetes, n (%)		21 (13.5)	130 (9.0)	.09
Hyperlipidemia, n (%)		76 (49)	615 (42.5)	.14
Hypertension, n (%)		155 (100)	1448 (100)	1
Multiple strokes, n (%)		40 (25.8)	348 (24.0)	.70
Time since most recent stroke, mean (SD)		7.8 (6.4)	9.8 (8.8)	<.001

[a]SB: sedentary behavior.

[b]LPA: light physical activity.

[c]MVPA: moderate-to-vigorous physical activity.

All participants had a hypertension diagnosis. The average time between accelerometer study start and depressed mood survey submission (time-to-assessment) was 1.8 years for both cohorts.

The average time from the initial stroke to the accelerometer study commencement was less for DDs than controls (7.8 vs 9.8 y).

Among DDs, 9 participants slept less than 7 hours while 75 slept more than 9 hours. In the control group, 79 participants slept less than 7 hours while 633 slept more than 9 hours.

Cross-Sectional Analysis

No significant association persisted in both the DD and control cohorts (Table 2). In controls, for each minute spent in MVPA per day, the odds of being in a higher depressed mood frequency category decreased by 23% (*P*<.001).

Table . Ordinal logistic regression assessing objective behavior predictors and depressed mood frequency.

Previous stroke participants	Prestroke depression		Controls	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Sleep (7 - 9 hr/d)	0.49 (0.23 - 1.03)	.06	0.88 (0.66 - 1.19)	.41
SB[b] (min/d)	1.00 (1.00 - 1.01)	.10	1.00 (1.00 - 1.00)	.63
LPA[c] (min/d)	1.00 (0.99 - 1.00)	.20	1.00 (1.00 - 1.00)	.35
MVPA[d] (min/d)	0.86 (0.64 - 1.17)	.33	0.77 (0.69 - 0.87)	<.001

[a]OR: odds ratio.

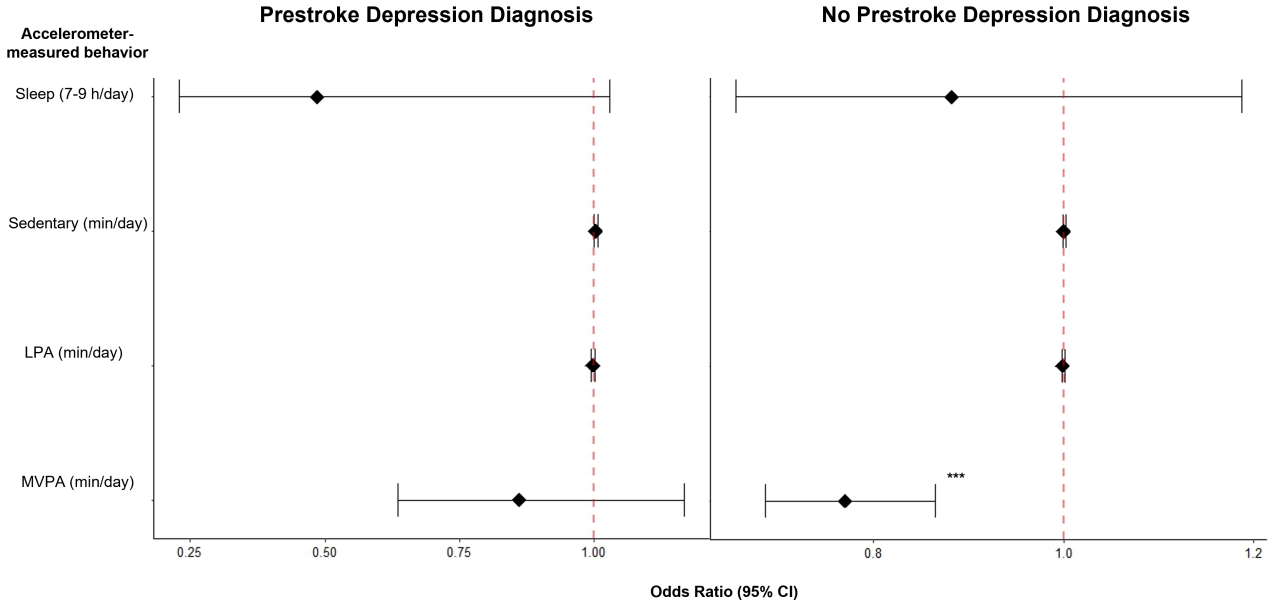
[b]SB: sedentary behavior.

[c]LPA: light physical activity.

[d]MVPA: moderate-to-vigorous physical activity.

Models were adjusted for age, sex, ethnicity, time-to-assessment, hyperlipidemia diagnosis, and diabetes diagnosis. Odds ratios (ORs) with 95% CIs for frequency of depressed mood are reported (Figure 2). ORs above 1 correspond to an increase in the accelerometer-measured behavior associated with increased depressed mood frequency.

Figure 2. Forest plot of odds ratios for depressed mood frequency by accelerometer-measured behavior comparing participants with prestroke depression diagnosis (DDs) and control cohorts. LPA: light physical activity; MVPA: moderate-to-vigorous physical activity. *** denotes statistical significance.



Sensitivity Analysis

Study Characteristics

In each filtered cohort (Multimedia Appendix 2), DDs were younger than controls (general cohort: 60 vs 62 y; IS: 64 vs 66 y; IC: 65 vs 67 y) and had a greater proportion of females (69.4% vs 56.8%; 60.0% vs 45.6%; 61.8% vs 45.1%). In the general population cohort, DDs had a greater proportion of White participants (97.7% vs 97%). On average, DDs also spent less time across cohorts in MVPA (35.0 vs 42.9 min/d; 31.0 vs 39.5 min/d; 32.0 vs 38.4 min/d), less time in LPA (295.0 vs 300.3 min/d; 286.7 vs 291.0 min/d; 287.5 vs 287.7 min/d), and more time asleep (9.1 vs 8.9 h/d; 9.1 vs 9.0 h/d; 9.04 vs 8.98 h/d).

While DDs in the general cohort spent slightly less time, on average, in SB than controls (564.3 vs 564.4 min/d), DDs in the IS and IC cohorts spent more time sedentary on average (577.1 vs 569.9 min/d; 578.2 vs 575.3 min/d).

In the general cohort, DDs had a higher proportion of diabetes (4.4% vs 2.9%) and hyperlipidemia (17.9% vs 14.9%) diagnoses and a lower proportion of participants with optimal sleep duration per day (49.5% vs 55.4%).

For the IS cohort, the average time from the completion of the depression survey to first stroke diagnosis was slightly more for DDs (1.9, SD 0.7 y) than controls (1.8, SD 0.6 y). In the IC cohort, the average time from the completion of the depression survey to first CeVD diagnosis was similarly more for DDs (1.9, SD 0.7 y) than controls (1.8, SD 0.7 y).

Cross-Sectional Analysis

In the general model (Multimedia Appendix 3), for each minute spent in MVPA, the odds of being in a higher depressed mood frequency category decreased by 18.4% (P<.001) and 13.5% (P<.001) for DDs (n=6096) and controls (n=62,589), respectively.

Also in the general model, specific only to controls, getting guideline-recommended sleep hours (7 - 9 h) each day was associated with a decreased odds of being in a higher depressed mood frequency category (5.3%; P=.02).

No significant associations were identified for those in the IS-only cohort (Multimedia Appendix 4).

For the final sensitivity analysis (Multimedia Appendix 5), assessing only those participants with an IC diagnosis, including stroke, the odds of being in a higher depressed mood frequency category decreased by 12.2% for each minute increase in MVPA (P=.03), only in controls (n=1526).

Discussion

Principal Findings

This investigation partially supports the hypothesis that objective behavior predictors would be associated with future depressed mood frequency. Although we found no significant associations between depressed mood frequency and SB, LPA or sleep for patients with prior stroke, regardless of prestroke depression diagnosis, we did find that the odds of being in a higher depressed mood frequency category decreased for each minute spent in MVPA; however, this association was only observed in participants without a prestroke depression diagnosis. This finding supports the exploratory aim of this manuscript, suggesting that participants with prestroke depression may experience different behavioral patterns compared to those without a prestroke depression diagnosis. Such a finding can potentially help clinicians tailor programs monitoring patients at risk of PSD.

The sensitivity analysis in the general cohort corroborates established findings that MVPA confers a protective effect on

mood, regardless of previous depression diagnosis. The lack of findings for the sensitivity analysis including only IS cases may be driven by the small sample sizes; however, the lack of findings also brings into question the potential for accelerometers to capture clinically actionable aberrations in patients before a stroke. Given that the protective effect of MVPA on depressed mood frequency was observed in the control cohort of patients with IC, accelerometer monitoring may be more appropriately directed to assess a broader range of neurologic changes, not just those linked with strokes.

Overall, the results suggest that accelerometer-based monitoring of behavior linked to depressed mood frequency may help clinicians identify patients who would benefit from resource-intensive screening, like the MADRS assessment. The sensitivity analyses support a separate approach for monitoring patients with a previous depression diagnosis, or more severe depression, compared to those with no documented depression or mild undiagnosed depression. When applied to predictive monitoring, a remote accelerometer-mood survey approach may be useful in cohorts of patients without a previous depression diagnosis, considering that patients with IC without clinical depression may experience observable behavior and mood changes before a CeVD diagnosis while their clinically depressed counterparts may not.

Limitations

A chief limitation of this study is that self-report data, like the depressed mood frequency survey, are subject to inaccuracies. Self-reported bias in survey responses may lead to misclassification of depressive symptom frequency and could influence different time-dependent results in our cohorts. Furthermore, the frequency of depression measures was not obtained by a clinician-graded protocol but, rather, by a survey questionnaire. Also, as the accelerometer study was only administered for one week and, on average, over a year before the follow-up mood survey, the impact of time between the objective measures and follow-up could have introduced substantial changes. The lack of associations observed for DDs may be due to the small sample size of participants with a previous depression diagnosis across cohorts. Moreover, the accelerometer study was only 1-week long and, therefore, may not generalize well to accurately represent busier or less busy weeks for patients. Accelerometer data collected on weekends versus weekdays may be distinct; however, this was not considered in this study.

The dichotomous investigation of clinically depressed and control patients are study strengths. In addition, UKBB participants were primarily White, limiting the generalizability of our findings outside of European populations. This UKBB study also primarily included participants aged 60 years and older and, as such, may not generalize well to young or middle-age adult populations. The majority of DDs were female across cohorts, a frequent finding in studies; however, male patients are less likely to seek out mental health resources, and the cohort stratification may be impacted by this.

Also, in the main analysis of previous stroke patients, participants diagnosed with clinical-grade depression after first stroke were excluded from this analysis. Considering the long gap in time from initial stroke to accelerometer study commencement, participants with a more immediate PSD diagnosis may either exhibit more intense symptoms or experience an underlying pathogenesis distinct from participants whose PSD symptoms are mild or emerge in the years after stroke.

Combining stroke types together as a single end point, as was done in the main analysis as well as the IS sensitivity analysis, may not consider unique characteristics of each stroke type and, as such, generated no significant results. Sleep was also assessed as a daily composite value, without consideration for time spent in a nap or broken sleep throughout the day. Together, these 2 limitations may have introduced confounding effects when considering sleep and depressed mood frequency, as previous research has shown short and long sleep to be associated with increased risk of intracerebral hemorrhage and ischemic stroke, respectively [40]. Furthermore, considering that all participants in our cohorts were hypertensive, MVPA's protective effect on depressive mood frequency may occur through improved cardiovascular health, rather than by conferring direct cerebral effects.

Comparison With Previous Work

No previous study assessed objective behavior measures and self-reported depressed mood frequency in patients with prior stroke years after their initial diagnosis. A key problem inherent in accelerometer research is that adherence to study designs is less-than-satisfactory for most studies [41]. This study also excluded participants with a more immediate PSD diagnosis, considering only those with prestroke depression diagnoses or those with no or mild depression after stroke. A self-report survey study of recent patients with prior stroke found that patients with high levels of PA before a stroke experienced less severe PSD [33]. Although our study could not confirm this analysis due to the design of the UKBB study, we extended those results by confirming that MVPA confers a protective effect on mood before a CeVD diagnosis in patients without a previous depression diagnosis, but not before a stroke-only diagnosis.

One plausible explanation for the lack of association between MVPA and depressed mood frequency in DDs may be that stroke survivors with a previous depression diagnosis have persistently deficient levels of brain-derived neurotropic factor (BDNF), a trophic factor released after exercise that is linked with improved mood benefits. It is well established that stroke patients in general have lower levels of BDNF, a marker of poor functional recovery [42]. The lack of a link between improved mood and MVPA in DDs may be driven by a less intense "exercise high" due to reduced or impaired BDNF function. In addition, other contributing factors, such as time spent in MVPA or neuroinflammation, may play a role in modulating BDNF expression in DDs. Of note, the lack of a significant association between MVPA and depressed mood frequency in participants with a previous depression diagnosis may be attributed to less time spent in MVPA compared with controls across all cohorts (patients with prior stroke, general population, IS, and IC). Time spent in MVPA may need to exceed a time threshold in

participants with previous depression diagnoses to improve mood.

The significant findings for IC cases, compared with the lack of findings for IS-only cases, are consistent with the updated *ICD-11* classification of CeVDs as a type of brain disease with shared etiology, rather than circulatory system disorders [37]. The protective effect of guideline-recommendation sleep (7 - 9 h/d) only observable in controls in the general cohort corroborates established work; however, the lack of associations across other cohorts may be explained by high levels of individual variability in sleep patterns, that is, nighttime disturbances, insomnia, and so on, previously identified in patients with depression as well as those at high risk of stroke [43,44].

A small pilot study of patients with minor ischemic stroke that found SB was positively associated with depression intensity and LPA was inversely associated with depression intensity [34]. Considering that this accelerometer study was conducted within the first 3 months after hospital discharge, our results extend these findings to look at mood in the years after a stroke. For instance, SB and LPA may be significant to monitor in the months after a stroke, while MVPA may be appropriate to monitor in the years after a stroke. Alternatively, MVPA may be less useful to monitor in minor ischemic stroke cases.

Using a larger dataset, our study builds on the feasibility demonstration of a small real-world study with patients with prior stroke, years after diagnosis, collecting one week of accelerometer data and ecological momentary assessments [45]. The results of our general cohort analysis considering participants without a previous depression diagnosis align with those from Sarris et al [46], who found that self-reported optimal sleep and PA were linked with decreased frequency of depressed mood in UKBB participants.

Conclusions

Our results highlight the importance of encouraging MVPA in patients with prior stroke without a depression diagnosis. Patients with prior strokes may be able to minimize short- and long-term disability and improve outcomes by proactively managing depressive symptoms. Applying MVPA to improve mood provides the added benefits of exercise-induced inflammation reduction and enhanced vascular elasticity while simultaneously reducing the risk of developing comorbidities and arterial stenosis or occlusion [47].

Considering that the only significant associations in the main analysis and incident sensitivity analyses were those that involved MVPA, it calls into question whether using accelerometer and depressed mood frequency survey data together can help clinicians identify patients who would benefit from remote monitoring, that is, this approach may generate more noise than signal over time. This study only considered a brief (1-week) accelerometer study, and over a year, on average, eclipsed between the *in situ* accelerometer study and the remote mood follow-up survey. Since neither the main analysis (previous stroke cohort) nor the incident sensitivity analyses resulted in significant associations for participants with a previous depression diagnosis, this underscores the need for additional research to determine whether this type of monitoring strategy can generate clinically actionable insights in participants with a previous depression diagnosis. Behavioral monitoring with accelerometer data and self-report surveys may not be helpful in patients with severe, or clinical-grade, depression. Future research should consider large sample sizes, longitudinal study designs, and analyze results stratified by time-to-diagnosis. Relevant to remote monitoring researchers, our findings highlight behavioral differences for those developing exploratory programs and clinically meaningful digital endpoints.

Overall, the cross-sectional analyses offer a robust perspective into the appropriateness of depression monitoring by digital sensors, using accelerometer wristwatches and smartphone, tablet, or PC-linked sensors. These insights offer clinical teams a strategy for translating digital health data, in this case, objective and subjective behavior measures, into scientifically valid frameworks for investigation. Future monitoring of patients at risk of different CeVD types, including those with a previous stroke diagnosis, should expand on our strategy and use both active and passive data to investigate relationships between objective digital sensor data and subsequent mood reports in patients diagnosed with and screened for depression. Based on our exploratory analysis, the potential for longitudinal data from objective sensors to predict mood appears feasible. In addition, PSD researchers should aim to characterize behavior measures linked with depressed mood across defined and clinically meaningful time periods, such as in the 3-month routine monitoring period after a stroke or TIA, considering that observable behaviors may evolve as CeVD or other neurologic disorder pathogenesis progresses.

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 91159. This research was funded by Mayo Clinic. SJZ was supported by a predoctoral fellowship in value assessment from the PhRMA Foundation and a National Center for Advancing Translational Sciences TL1 training grant (5TL1TR002380-05).

Disclaimer

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data Availability

All data are publicly available, upon research approval access, from UK Biobank [48]. The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request. Analysis code is available [49].

Authors' Contributions

SJZ completed the study design and manuscript drafting. BJE, BMD, and AG provided clinical expertise and contributed to manuscript editing. GMC provided expertise for obtaining data access and designing the study. AG conducted statistical review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions for classifying patients.
[DOCX File, 17 KB - neuro_v4i1e56679_app1.docx]

Multimedia Appendix 2
Sample characteristics across sensitivity cohorts.
[DOCX File, 18 KB - neuro_v4i1e56679_app2.docx]

Multimedia Appendix 3
Ordinal logistic regression assessing objective behavior predictors and depressed mood frequency in the general cohort.
[DOCX File, 15 KB - neuro_v4i1e56679_app3.docx]

Multimedia Appendix 4
Ordinal logistic regression assessing objective behavior predictors and depressed mood frequency in incident stroke cohorts.
[DOCX File, 15 KB - neuro_v4i1e56679_app4.docx]

Multimedia Appendix 5
Ordinal logistic regression assessing objective behavior predictors and depressed mood frequency in incident cerebrovascular disease cohorts.
[DOCX File, 15 KB - neuro_v4i1e56679_app5.docx]

References

1. Broomfield NM, Quinn TJ, Abdul-Rahim AH, Walters MR, Evans JJ. Depression and anxiety symptoms post-stroke/TIA: prevalence and associations in cross-sectional data from a regional stroke registry. BMC Neurol 2014 Oct 1;14:198. [doi: 10.1186/s12883-014-0198-8] [Medline: 25269762]
2. Medeiros GC, Roy D, Kontos N, Beach SR. Post-stroke depression: a 2020 updated review. Gen Hosp Psychiatry 2020;66:70-80. [doi: 10.1016/j.genhosppsych.2020.06.011] [Medline: 32717644]
3. Selvaraj S, Arora T, Casameni Montiel T, et al. Early screening for post-stroke depression, and the effect on functional outcomes, quality of life and mortality: a protocol for a systematic review and meta-analysis. BMJ Open 2021 Aug 17;11(8):e050451. [doi: 10.1136/bmjopen-2021-050451] [Medline: 34404715]
4. Liu L, Marshall IJ, Pei R, et al. Natural history of depression up to 18 years after stroke: a population-based South London Stroke Register study. Lancet Reg Health Eur 2024 May;40:100882. [doi: 10.1016/j.lanepe.2024.100882] [Medline: 38745986]
5. Dall TM, Storm MV, Chakrabarti R, et al. Supply and demand analysis of the current and future US neurology workforce. Neurology (ECronicon) 2013 Jul 30;81(5):470-478. [doi: 10.1212/WNL.0b013e318294b1cf] [Medline: 23596071]
6. Brambilla M, Cerasetti M, Pepe F, et al. Comparison of Oxford Cognitive Screen and Montreal Cognitive Assessment feasibility in the stroke unit setting. a pilot study. Cereb Circ Cogn Behav 2021;2:100021. [doi: 10.1016/j.cccb.2021.100021] [Medline: 36324706]
7. Capaldi II VF. Emerging strategies in the treatment of poststroke depression and psychiatric distress in patients. PRBM 2010;3:109. [doi: 10.2147/PRBM.S10035]
8. Nakling AE, Aarsland D, Næss H, et al. Cognitive deficits in chronic stroke patients: neuropsychological assessment, depression, and self-reports. Dement Geriatr Cogn Disord Extra 2017 Sep 14;7(2):283-296. [doi: 10.1159/000478851]
9. Ahmed NN, Reagu S, Alkhoori S, Cherchali A, Purushottamahanti P, Siddiqui U. Improving mental health outcomes in patients with major depressive disorder in the Gulf States: a review of the role of electronic enablers in monitoring residual symptoms. J Multidiscip Healthc 2024;17:3341-3354. [doi: 10.2147/JMDH.S475078] [Medline: 39010931]

10. Lyketsos CG, Kozauer N, Rabins PV. Psychiatric manifestations of neurologic disease: where are we headed? Dialogues Clin Neurosci 2007;9(2):111-124. [doi: 10.31887/DCNS.2007.9.2/clyketsos] [Medline: 17726911]
11. van Sloten TT, Sigurdsson S, van Buchem MA, et al. Cerebral small vessel disease and association with higher incidence of depressive symptoms in a general elderly population: the AGES-Reykjavik study. Am J Psychiatry 2015 Jun;172(6):570-578. [doi: 10.1176/appi.ajp.2014.14050578] [Medline: 25734354]
12. Krick S, Koob JL, Latarnik S, et al. Neuroanatomy of post-stroke depression: the association between symptom clusters and lesion location. Brain Commun 2023;5(5):fcad275. [doi: 10.1093/braincomms/fcad275] [Medline: 37908237]
13. Moses JC, Adibi S, Shariful Islam SM, Wickramasinghe N, Nguyen L. Application of smartphone technologies in disease monitoring: a systematic review. Healthcare (Basel) 2021 Jul 14;9(7):889. [doi: 10.3390/healthcare9070889] [Medline: 34356267]
14. Wouters F, Gruwez H, Vranken J, et al. The potential and limitations of mobile health and insertable cardiac monitors in the detection of atrial fibrillation in cryptogenic stroke patients: preliminary results from the REMOTE trial. Front Cardiovasc Med 2022;9:848914. [doi: 10.3389/fcvm.2022.848914] [Medline: 35498000]
15. Phienphanich P, Tankongchamruskul N, Akarathanawat W, et al. Automatic stroke screening on mobile application: features of gyroscope and accelerometer for arm factor in FAST. Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 23-27, 2019; Berlin, Germany p. 4225-4228. [doi: 10.1109/EMBC.2019.8857550]
16. Capela NA, Lemaire ED, Baddour N, Rudolf M, Goljar N, Burger H. Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. J Neuroeng Rehabil 2016 Jan 20;13:5. [doi: 10.1186/s12984-016-0114-0] [Medline: 26792670]
17. Beerten SG, Proesmans T, Vaes B. A heart rate monitoring app (FibriCheck) for atrial fibrillation in general practice: pilot usability study. JMIR Form Res 2021 Apr 7;5(4):e24461. [doi: 10.2196/24461] [Medline: 33825692]
18. Forster SD, Gauggel S, Loevenich R, et al. A microanalysis of mood and self-reported functionality in stroke patients using ecological momentary assessment. Front Neurol 2022;13:854777. [doi: 10.3389/fneur.2022.854777] [Medline: 35665036]
19. Forster SD, Gauggel S, Petershofer A, Völzke V, Mainz V. Ecological momentary assessment in patients with an acquired brain injury: a pilot study on compliance and fluctuations. Front Neurol 2020;11:115. [doi: 10.3389/fneur.2020.00115] [Medline: 32194494]
20. Lenaert B, van Kampen N, van Heugten C, Ponds R. Real-time measurement of post-stroke fatigue in daily life and its relationship with the retrospective Fatigue Severity Scale. Neuropsychol Rehabil 2022 Jul;32(6):992-1006. [doi: 10.1080/09602011.2020.1854791] [Medline: 33297839]
21. Helgadóttir B, Forsell Y, Ekblom Ö. Physical activity patterns of people affected by depressive and anxiety disorders as measured by accelerometers: a cross-sectional study. PLoS One 2015;10(1):e0115894. [doi: 10.1371/journal.pone.0115894] [Medline: 25585123]
22. Murphy RP, Reddin C, Rosengren A, et al. Depressive symptoms and risk of acute stroke: INTERSTROKE case-control study. Neurol (ECronicon) 2023 Apr 25;100(17):e1787-e1798. [doi: 10.1212/WNL.0000000000207093] [Medline: 36889922]
23. Everson-Rose SA, Roetker NS, Lutsey PL, et al. Chronic stress, depressive symptoms, anger, hostility, and risk of stroke and transient ischemic attack in the multi-ethnic study of atherosclerosis. Stroke 2014 Aug;45(8):2318-2323. [doi: 10.1161/STROKEAHA.114.004815] [Medline: 25013018]
24. Avan A, Hachinski V. Stroke and dementia, leading causes of neurological disability and death, potential for prevention. Alzheimers Dement 2021 Jun;17(6):1072-1076. [doi: 10.1002/alz.12340] [Medline: 34057294]
25. Heshmatollah A, Dommershuijsen LJ, Fani L, Koudstaal PJ, Ikram MA, Ikram MK. Long-term trajectories of decline in cognition and daily functioning before and after stroke. J Neurol Neurosurg Psychiatry 2021 Nov;92(11):1158-1163. [doi: 10.1136/jnnp-2021-326043] [Medline: 34230107]
26. Johnson SA, Karas M, Burke KM, et al. Wearable device and smartphone data quantify ALS progression and may provide novel outcome measures. NPJ Digit Med 2023 Mar 6;6(1):34. [doi: 10.1038/s41746-023-00778-y] [Medline: 36879025]
27. Winer JR, Lok R, Weed L, et al. Impaired 24-h activity patterns are associated with an increased risk of Alzheimer's disease, Parkinson's disease, and cognitive decline. Alz Res Therapy 2024;16(1):35. [doi: 10.1186/s13195-024-01411-0]
28. Zawada SJ, Aissa NH, Conte GM, et al. In situ physiologic and behavioral monitoring with digital sensors for cerebrovascular disease: a scoping review. Mayo Clin Proc Dig Health 2023 Jun;1(2):139-160. [doi: 10.1016/j.mcpdig.2023.03.007]
29. Jean FAM, Swendsen JD, Sibon I, Fehér K, Husky M. Daily life behaviors and depression risk following stroke: a preliminary study using ecological momentary assessment. J Geriatr Psychiatry Neurol 2013 Sep;26(3):138-143. [doi: 10.1177/0891988713484193] [Medline: 23584854]
30. Johnson EI, Sibon I, Renou P, Rouanet F, Allard M, Swendsen J. Feasibility and validity of computerized ambulatory monitoring in stroke patients. Neurology (ECronicon) 2009 Nov 10;73(19):1579-1583. [doi: 10.1212/WNL.0b013e3181c0d466] [Medline: 19901250]
31. Sibon I, Lassalle-Lagadec S, Renou P, Swendsen J. Evolution of depression symptoms following stroke: a prospective study using computerized ambulatory monitoring. Cerebrovasc Dis 2012;33(3):280-285. [doi: 10.1159/000334663] [Medline: 22285959]

32. Tanaka H, Kitamura G, Tamura M, et al. Pre-stroke physical activity is associated with post-stroke physical activity and sedentary behavior in the acute phase. Sci Rep 2023;13(1):21298. [doi: 10.1038/s41598-023-48232-z]
33. Reinholdsson M, Palstam A, Jood K, S Sunnerhagen K. Associations between pre-stroke physical activity levels and health-related quality of life 3 months after stroke: a registry-based study (part of PAPSIGOT). Int J Stroke 2023 Dec;18(10):1178-1185. [doi: 10.1177/17474930231184367]
34. Ashizawa R, Honda H, Yoshizawa K, Kameyama Y, Yoshimoto Y. Association between physical activity levels and depressive symptoms in patients with minor ischemic stroke. J Stroke Cerebrovasc Dis 2022 Sep;31(9):106641. [doi: 10.1016/j.jstrokecerebrovasdis.2022.106641] [Medline: 35834937]
35. Lau SCL, Tabor Connor L, Baum CM. Motivation, physical activity, and affect in community-dwelling stroke survivors: an ambulatory assessment approach. Ann Behav Med 2023 Apr 12;57(4):334-343. [doi: 10.1093/abm/kaac065] [Medline: 36732938]
36. Alvaro PK, Roberts RM, Harris JK. A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. Sleep 2013 Jul 1;36(7):1059-1068. [doi: 10.5665/sleep.2810] [Medline: 23814343]
37. Ramasubbu R. Relationship between depression and cerebrovascular disease: conceptual issues. J Affect Disord 2000;57(1-3):1-11. [doi: 10.1016/s0165-0327(99)00101-9] [Medline: 10708811]
38. Madjedi KM, Stuart KV, Chua SYL, et al. The association of physical activity with glaucoma and related traits in the UK Biobank. Ophthalmology 2023 Oct;130(10):1024-1036. [doi: 10.1016/j.ophtha.2023.06.009] [Medline: 37331483]
39. Mac Grory B, Schrag M, Biousse V, et al. Management of central retinal artery occlusion: a scientific statement from the American Heart Association. Stroke 2021 Jun;52(6):e282-e294. [doi: 10.1161/STR.0000000000000366] [Medline: 33677974]
40. Lu H, Wu PF, Li RZ, Zhang W, Huang GX. Sleep duration and stroke: a Mendelian randomization study. Front Neurol 2020;11:976. [doi: 10.3389/fneur.2020.00976]
41. Barak S, Wu SS, Dai Y, Duncan PW, Behrman AL. Locomotor Experience Applied Post-Stroke (LEAPS) Investigative Team. Adherence to accelerometry measurement of community ambulation poststroke. Phys Ther 2014;94(1):101-110. [doi: 10.2522/ptj.20120473]
42. Chaturvedi P, Singh AK, Tiwari V, Thacker AK. Brain-derived neurotrophic factor levels in acute stroke and its clinical implications. Brain Circ 2020;6(3):185-190. [doi: 10.4103/bc.bc_23_20] [Medline: 33210043]
43. Lim JA, Yun JY, Choi SH, Park S, Suk HW, Jang JH. Greater variability in daily sleep efficiency predicts depression and anxiety in young adults: estimation of depression severity using the two-week sleep quality records of wearable devices. Front Psychiatry 2022;13:1041747. [doi: 10.3389/fpsyt.2022.1041747] [Medline: 36419969]
44. Ma J, Ma N, Zhang L, Xu L, Liu X, Meng G. Association of total sleep duration variability with risk of new stroke in the middle-aged and elderly Chinese population. BMC Neurol 2024 Jun 25;24(1):217. [doi: 10.1186/s12883-024-03727-8] [Medline: 38918750]
45. Lau SCL, Connor LT, King AA, Baum CM. Multimodal ambulatory monitoring of daily activity and health-related symptoms in community-dwelling survivors of stroke: feasibility, acceptability, and validity. Arch Phys Med Rehabil 2022 Oct;103(10):1992-2000. [doi: 10.1016/j.apmr.2022.06.002] [Medline: 35780826]
46. Sarris J, Thomson R, Hargraves F, et al. Multiple lifestyle factors and depressed mood: a cross-sectional and longitudinal analysis of the UK Biobank (N = 84,860). BMC Med 2020 Nov 12;18(1):354. [doi: 10.1186/s12916-020-01813-5] [Medline: 33176802]
47. Prior PL, Suskin N. Exercise for stroke prevention. Stroke Vasc Neurol 2018 Jun;3(2):59-68. [doi: 10.1136/svn-2018-000155] [Medline: 30191075]
48. UK Biobank. URL: https://www.ukbiobank.ac.uk/ [accessed 2024-12-18]
49. Szawada-mayo/jmir_56679. GitHub. URL: https://github.com/szawada-mayo/jmir_56679 [accessed 2024-12-13]

Abbreviations

BDNF: brain-derived neurotropic factor
CeVD: cerebrovascular disease
DD: depression diagnosis
IC: incident cerebrovascular disease
***ICD-10*:** *International Classification of Diseases, Tenth Revision*
***ICD-11*:** *International Classification of Diseases, Eleventh Revision*
IS: incident stroke
LPA: light physical activity
MADRS: Montgomery–Åsberg Depression Rating Survey
MVPA : moderate-to-vigorous physical activity
OR: odds ratio
PA: physical activity
PHQ-2: Patient Health Questionnaire-2
PSD: poststroke depression

SB: sedentary behavior
TIA: transient ischemic attack
UKBB: UK Biobank

Edited by P Kubben; submitted 23.01.24; peer-reviewed by E Harahsheh, N Allen; revised version received 23.10.24; accepted 14.11.24; published 10.01.25.

Please cite as:
Zawada SJ, Ganjizadeh A, Conte GM, Demaerschalk BM, Erickson BJ
Exploring Remote Monitoring of Poststroke Mood With Digital Sensors by Assessment of Depression Phenotypes and Accelerometer Data in UK Biobank: Cross-Sectional Analysis
JMIR Neurotech 2025;4:e56679
URL: https://neuro.jmir.org/2025/1/e56679
doi:10.2196/56679

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