Smart Health 41 (2026) 100685

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

NeuroLens: A multimodal patient system for early screening and real-time monitoring of Parkinson's disease☆

Zixu Geng [a], Shih-Ching Yeh [b,c], Xiaowei Xu [d], Jian Zhuang [d], Jingwei Li [a], Ruitian Wu [a], Feng Lin [e], Longxiang Pan [g], Wenyao Xu [f], Ming-Chun Huang [a,*]

[a] Department of Data and Computational Science, Duke Kunshan University, Kunshan, Jiangsu, China
[b] Department of Photonic System, National Yang Ming Chiao Tung University, Taiwan
[c] Department of Computer Science and Information Engineering, National Central University, Taiwan
[d] Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences) and Southern Medical University, Guangzhou, China
[e] State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, Zhejiang, China
[f] Department of Computer Science and Engineering, University at Buffalo, the State University of New York, USA
[g] University of Pittsburgh Medical Center, USA

ARTICLE INFO

Keywords:
Parkinson's disease
Multimodal sensing
Digital health
Real-time monitoring
Early screening

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder that requires early detection and continuous monitoring. This paper presents NeuroLens, a multimodal patient system that integrates sensor data, behavioral analysis, and real-time monitoring to support early screening and disease progression tracking. The proposed system demonstrates potential for improving clinical decision-making and enabling personalized healthcare interventions.

1. Introduction

Parkinson's disease (PD) represents a significant health challenge, especially among the elderly, leading to severe impairments in motor function and a range of psychological, social, and economic burdens on patients, caregivers, and society (Bloem et al., 2021). Parkinson's Disease manifests through a series of debilitating symptoms that profoundly impact patients' physical and psychological well-being. Common motor symptoms include resting tremor, bradykinesia, rigidity, and postural instability (Xia & Mao, 2012), which can make even simple daily tasks challenging (Armstrong & Okun, 2020). As the disease progresses, these motor symptoms worsen, and patients may eventually lose the ability to live independently, rendering them unable to walk, stand, or perform any form of self-care. This results in complete dependence on others for all daily activities, placing immense physical and psychological strain on the patient. Prolonged immobility can lead to severe complications such as pneumonia, deep vein thrombosis, and pressure ulcers, further increasing the patient's health risks. In addition to these life-threatening motor impairments, many Parkinson's Disease patients experience non-motor symptoms such as severe cognitive decline, depression, anxiety, and hallucinations (Pfeiffer, 2016). These psychological symptoms not only worsen the patient's mental state but also increase the burden on families and caregivers, significantly diminishing the quality of life for both patients and their loved ones. Cognitive decline may lead to severe memory

☆ This article is part of a Special issue entitled: 'CHASE 2026_invited only' published in Smart Health.

* Corresponding author.
E-mail addresses: zg73@duke.edu (Z. Geng), shihching.yeh@g.ncu.edu.tw (S.-C. Yeh), xiao.wei.xu@foxmail.com (X. Xu), Zhuangjian@gdph.org.cn (J. Zhuang), jl957@duke.edu (J. Li), rw219@duke.edu (R. Wu), flin@zju.edu.cn (F. Lin), panl3@upmc.edu (L. Pan), wenyaoxu@buffalo.edu (W. Xu), mh596@duke.edu (M.-C. Huang).

https://doi.org/10.1016/j.smhl.2026.100685
Received 9 June 2026; Accepted 20 June 2026
Available online 1 July 2026

loss, an inability to make simple decisions, and the inability to handle daily tasks, potentially progressing to dementia. This results in the complete loss of self-awareness and the ability to live independently.

As the disease advances, it leads to irreversible physical and cognitive impairments, affecting the patient's ability to perform basic activities such as eating, dressing, and speaking. This further emphasizes the need for timely intervention to slow disease progression and mitigate long-term disability. Additionally, although early initiation of dopaminergic therapy may not slow the biological progression of Parkinson's Disease, clinical studies have shown that early treatment significantly improves patients' quality of life and functional outcomes. For example, the LEAP study demonstrated that even in patients with minimal disability, early low-dose levodopa treatment led to better disease-related quality of life, with significantly improved PDQ-39 scores at 22 weeks, and 21% of patients in the delayed-treatment group required early switching due to emerging symptoms (de Bie et al., 2020). Similarly, the PD MED trial reported that patients starting with levodopa achieved higher mobility scores (PDQ-39: +1.8, $p = 0.005$) and better overall quality of life (EQ-5D: +0.03, $p = 0.0002$), without an increase in motor complications. These findings were consistent across different populations, including comparative studies in Ghana and Italy, suggesting that treatment timing does not affect complication onset. Together, this growing body of evidence shows that early symptomatic treatment reduces disability and improves long-term outcomes, reinforcing the critical importance of early diagnosis to ensure patients receive timely interventions before irreversible decline occurs.

Currently, the diagnosis of Parkinson's Disease relies heavily on a combination of questionnaire-based assessments and imaging-based confirmation. The Unified Parkinson's Disease Rating Scale (UPDRS) is the most widely adopted clinical instrument, offering a structured evaluation across multiple domains including motor performance, activities of daily living, and non-motor symptoms (Martinez-Martin et al., 2015). It assigns numerical scores based on clinician observations and patient-reported experiences, providing a semi-quantitative estimate of disease burden. To improve standardization and sensitivity, the Movement Disorder Society–sponsored revision of the UPDRS (MDS-UPDRS) was introduced, which refines the scale structure and scoring anchors while expanding coverage of non-motor symptoms. The MDS-UPDRS includes four parts: (I) non-motor experiences of daily living, (II) motor experiences of daily living, (III) motor examination, and (IV) motor complications, each scored on a uniform 0–4 scale. It has demonstrated strong psychometric properties and correlates well with quality-of-life outcomes across large patient cohorts (Skorvanek et al., 2018). Besides, the Hoehn and Yahr scale is frequently used in parallel to stage disease progression, categorizing patients into discrete levels based on symptom severity and physical independence (Bhidayasiri et al., 2012). Beyond questionnaires, clinicians often use dopamine transporter (DAT) scans via SPECT imaging to visualize striatal dopamine depletion, a biological hallmark of Parkinson's Disease (Kaegi et al., 2010). Structural MRI is also utilized to rule out alternative diagnoses and assess brain atrophy patterns (Mahlknecht et al., 2010). Furthermore, patient-reported non-motor symptoms — such as mood disturbances, sleep disorders, and hallucinations — are often included in assessments to provide a more holistic view of disease progression (Mei et al., 2021). Together, these tools represent the current gold standard for clinical diagnosis, combining subjective scales with objective imaging biomarkers.

While questionnaire-based scales and imaging techniques remain indispensable tools in Parkinson's Disease diagnosis, their practical application is constrained by several real-world limitations. Instruments like the UPDRS and MDS-UPDRS provide structured and validated frameworks for evaluating motor and non-motor symptoms, yet their effectiveness depends heavily on consistent administration and interpretation. In practice, questionnaire-based assessments are conducted infrequently during scheduled clinic visits—making it difficult to capture symptom fluctuations between appointments. Moreover, questionnaire-based evaluations rely significantly on patient self-reporting and clinician interpretation, both of which introduce subjectivity. Patients may underreport or misrepresent symptoms due to recall bias, misunderstanding, or communication barriers, while clinicians may vary in their scoring due to differences in training or observational emphasis (Balestrino & Schapira, 2020; Sánchez-Ferro et al., 2016). Studies have shown that inter-rater variability in UPDRS scoring can reach up to 20%, particularly during early-stage PD when symptoms are subtle and harder to detect reliably (Regnault et al., 2019). Even the MDS-UPDRS, though improved, shows a Person Separation Index (PSI) of only 0.72 for early-stage assessments, indicating modest sensitivity in this critical diagnostic window (Regnault et al., 2019). Imaging-based techniques such as PET, SPECT, and MRI add objectivity by revealing neurophysiological and structural changes, yet they are constrained by high costs, technical infrastructure requirements, and limited availability (Mortezazadeh et al., 2021). These modalities are typically not used for routine follow-up or community-based screening due to their resource-intensive nature. Therefore, while traditional methods form the clinical foundation for Parkinson's Disease diagnosis, their reliance on infrequent, subjective, or inaccessible procedures underscores the need for more scalable, objective, and high-frequency assessment tools.

To address these challenges, we developed NeuroLens, a cloud-based platform that provides a comprehensive and scalable solution for early Parkinson's Disease screening and real-time symptom assessment. Specifically, our system incorporates the following design elements:

- **Comprehensive Multimodal Data Benchmark:** NeuroLens replaces traditional lab-grade equipment with mobile-compatible sensors, including smartphone-based IMUs, wearable EEG headsets, and microphones. This setup enables scalable and low-cost data acquisition across diverse settings. The system supports eight input modalities—gait, fine motor control, breath, speech, auditory memory, sentence reading, vowel articulation, and EEG forming one of the most comprehensive datasets to date for Parkinson's Disease assessment as shown in Table 1.
- **Interactive and Mobile Trial Protocols Co-Designed with Experienced Neurologists:** All testing protocols were jointly designed with experienced neurologists from a tertiary clinical setting to ensure clinical validity and usability in mobile environments. These protocols include Straight Walking, Turning, Tremor, Join Circle, Auditory Memory, Record Breath, Record Sentence, and Record Vowel tasks. Each protocol is optimized to capture domain-specific motor and cognitive biomarkers through smartphone-based or wearable interfaces.

Table 1
Summary of Parkinson's Disease diagnostic systems.

Paper	Method type	Sensors	Dataset	Classification	Accuracy
Tong et al. (2021)	Single-modality	Gait	47 participants	Multi-level (3 stages)	98.08%
Navita et al. (2025)	Single-modality	Gait	166 participants	Multi-level (3 stages)	96.4%
Ramesh and Bilal (2022)	Single-modality	Gait	35 participants	Binary	78%
Drotár et al. (2016)	Single-modality	Fine motor	75 participants	Binary	81.3%
Galna et al. (2014)	Single-modality	Gait	19 participants	Binary	ICC 0.989
Vásquez-Correa et al. (2018)	Multi-modality	Gait + Fine motor + Speech	84 participants	Binary	97.6%
Lim et al. (2025)	Multi-modality	Gait + Fine motor + Speech	Not specified	Binary	AUC 0.82
Wang et al. (2020)	Multi-modality	Gait + EEG	17 participants	Binary	MCC 0.211
Zhang et al. (2022)	Multi-modality	Gait + Fine motor + EEG	18 participants	Binary	93%
Zhang and Gu (2019)	Multi-modality	Gait + EEG	10 participants	Binary	94.3%
Hannan (2024)	Multi-modality	MRI + UPDRS + Genetic Biomarkers	31 participants	Multi-Level (4 stages)	94%
NeuroLens	Multi-modality	Gait + Fine motor + Breath + Speech + Memory + EEG	86 participants	Multi-Level (5 stages)	91%

- **Multimodality Data Stream Processing and Fusion via Meta Classifier Integration:** NeuroLens employs task-specific models for each modality and integrates their predictions using a meta classifier based on Bayesian-optimized soft voting. This fusion approach captures cross-modal dependencies and improves classification robustness over individual models.
- **Fine-Grained Symptom Quantification Across Five Clinically Defined Stages:** NeuroLens provides five-stage symptom scoring—Normal, Mild, Moderate, Severe, and Profound aligned with clinical standards (Martínez-Martín et al., 2015). This granularity enables better disease progression tracking and personalized treatment planning.

2. Related work

Early efforts to quantify Parkinson's Disease (PD) symptoms have heavily relied on mobile health applications and digital questionnaires derived from clinical scales such as the Unified Parkinson's Disease Rating Scale (UPDRS). Several smartphone-based applications have been developed to support daily symptom logging, medication adherence tracking, and remote motor assessments (Li, Chen et al., 2023; Zhang et al., 2020). These tools offer improved accessibility and promote longitudinal data collection outside the clinic, enabling researchers to monitor patients in assisted daily living environments. However, they do not fundamentally resolve the limitations of traditional assessments. Most mobile platforms are modeled after paper-based questionnaires and still rely on subjective patient input, introducing recall bias and interpretation errors. Moreover, the temporal sparsity of self-reporting and the lack of continuous monitoring lead to missed episodes of transient symptoms or subtle progression, which are critical in early-stage PD diagnosis and treatment optimization (Evers et al., 2019).

To address the subjectivity inherent in questionnaire-based assessments, researchers have explored the use of wearable and portable sensor systems to objectively capture specific PD symptoms. Wearable motion sensors such as inertial measurement units (IMUs), accelerometers, and gyroscopes have been widely used to measure gait parameters — such as step time asymmetry, stride variability, and cadence — which are strongly correlated with motor impairments like bradykinesia and postural instability (Brognara et al., 2019; Di Biase et al., 2020). Tong et al. (2021) applied permutation-variable importance analysis to gait data from PhysioNet, using a support vector machine (SVM) classifier to distinguish multiple PD severity levels among 47 participants, achieving 98.08% accuracy. Similarly, Navita et al. (2025) performed multi-level classification using ground reaction force data from 166 participants, reporting an accuracy of 96.4%. Ramesh and Bilal (2022) proposed a CNN+GAN model using a single wearable IMU to predict postural instability and gait disorder (PIGD) scores and classify ON/OFF medication states in 70 participants, achieving up to 100% accuracy in distinguishing OFF from ON states during a 2-minute walk test. Audio-based systems analyze features like pitch variation, articulation rate, and voice jitter to detect hypophonia and dysarthria, both of which are common in PD patients. Fine motor symptoms have been studied using spiral drawing or finger tapping tasks collected through tablet-based styluses or capacitive touchscreen input, which allow for kinematic analysis of tremor amplitude, movement velocity, and coordination (Drotár et al., 2016). In addition to wearables, non-contact systems such as time-of-flight cameras and depth sensors (e.g., Microsoft Kinect) have been utilized to assess gross motor movements and gait patterns through 3D skeletal tracking (Galna et al., 2014). These single-sensor approaches offer non-invasive, objective, and low-cost alternatives to clinical observation, enabling large-scale, automated symptom tracking.

Despite their contributions, single-sensor systems face inherent limitations. Each sensor modality targets only a narrow symptom domain, limiting the comprehensiveness of the evaluation. For example, motor sensors for gait analysis can effectively quantify lower-limb movement abnormalities but cannot capture speech impairments or upper-limb tremor. Similarly, speech-based assessments may detect vocal deficits but provide no insight into axial rigidity or fine motor control. Consequently, these systems often correlate with only one or two subdomains of the UPDRS, failing to capture the full range of symptoms necessary for reliable disease staging (Tong et al., 2021). Moreover, signal variability due to sensor placement, environmental noise, and user compliance can affect measurement reliability. Most studies also utilize relatively small sample sizes under controlled conditions, limiting their ability to generalize across different populations, disease stages, or usage environments. These constraints have prompted a shift toward

Fig. 1. System architecture: This figure depicts the NeuroLens system architecture for Parkinson's Disease severity assessment through a multimodal approach. The Multi-Modal Input Module collects data from auditory, motor, and visual tasks performed via a mobile application. This data is transmitted to the Cloud Platform, where it undergoes preprocessing and feature extraction. A meta-classifier integrates results from individual protocol-specific models to estimate Parkinson's Disease severity. The Feedback Module visualizes the severity score and provides actionable insights to participants and clinicians.

multimodal sensor fusion approaches, which aim to leverage complementary data streams to enable more holistic and accurate assessments.

Multimodal systems have progressively evolved to improve the detection and monitoring of Parkinson's Disease (PD) by combining various symptom-related signals. Vásquez-Correa et al. (2018) pioneered a system that integrates gait, handwriting, and speech data using deep convolutional networks, achieving a high binary classification accuracy of 97.6%. However, their system operated in a controlled environment and only provided coarse binary diagnosis without real-time capability. Lim et al. (2025) extended this concept by developing a smartphone-based platform that collects voice recordings, finger tapping, and walking data, achieving an AUC of 0.82. While more portable, this system still lacked clinician-guided protocols and could not support continuous feedback or fine-grained staging. Building on neurophysiological data, Wang et al. (2020) employed wearable EEG and gait sensors to detect freezing of gait episodes, revealing that fusing brain and movement signals improved detection (MCC = 0.211). Yet, the system relied on cumbersome equipment and limited participant data (17 subjects), hindering scalability. Zhang et al. (2022) later incorporated EEG, fine motor, and gait data into a multimodal model, achieving 93% accuracy, but the study used only 18 participants and remained confined to binary classification. Similarly, Zhang and Gu (2019) reported 94.3% accuracy with just 10 subjects. Cai et al. (2022) proposed a multimodal deep learning system combining pressure sensors and IMUs to estimate stride length and recognize rehabilitation activities, achieving 3.89 cm mean absolute error and 97.08% classification accuracy, demonstrating the potential of lightweight mHealth systems in real-world gait monitoring. Multimodal sensing includes EEG generally enlarge the difficulty for data collection, processing, analysis, and visualization without a cloud-based infrastructure. TherapyPal (Li, Qian et al., 2023) was proposed as a privacy-preserving companion diagnostic tool that uses smartphone sensors to assess medication effectiveness among 65 PD patients. It extracts digital symptomatic phenotypes via semantic hashing and performs contrastive learning on hash representations to detect drug efficacy. While TherapyPal demonstrates strong binary classification performance (84.1% accuracy) with larger sample size, it focuses solely on before/after-medication symptom fluctuation and does not support multi-stage symptom classification. Moreover, its primary goal is medication response detection rather than comprehensive disease staging or multi-domain assessment.

To overcome these challenges, we propose NeuroLens—a cloud-based, multimodal sensing platform that integrates six input types: gait, fine motor skills, breathing, speech, auditory memory, and EEG signals. It establishes a comprehensive multimodal data benchmark using lightweight, accessible hardware suitable for home and clinical settings. By addressing the limitations of traditional assessments, single-modality systems, and existing multimodal tools, NeuroLens offers a clinically scalable, real-time solution suitable for early-stage detection and continuous monitoring. The platform establishes a comprehensive benchmark for Parkinson's Disease assessment by integrating EEG, motion, and auditory data within a unified cloud-based framework.

3. System design

3.1. System overview

The NeuroLens system is designed to assess Parkinson's Disease (PD) severity through a multimodal connected health framework that integrates motor, cognitive, and auditory assessments. The system consists of three core components: a **Multi-Modal Input Module**, a **Cloud Platform Analysis Module**, and a **Feedback Module**. This modular design enables seamless data flow from edge sensing to cloud-based analysis and clinical feedback, as illustrated in Fig. 2.

3.2. Multi-Modal Input Module

All assessment protocols were co-designed in collaboration with experienced neurologists to ensure clinical relevance and diagnostic validity. The Multi-Modal Input Module acquires complementary data streams through a mobile application and wearable sensors, covering auditory, motor, haptic, and neurophysiological domains commonly affected in Parkinson's disease.

Auditory protocols assess speech production, respiratory control, and auditory memory through sentence reading, sustained vowel phonation, and controlled breathing tasks. These capture vocal and respiratory biomarkers associated with hypophonia, monotonicity, and impaired breath control.

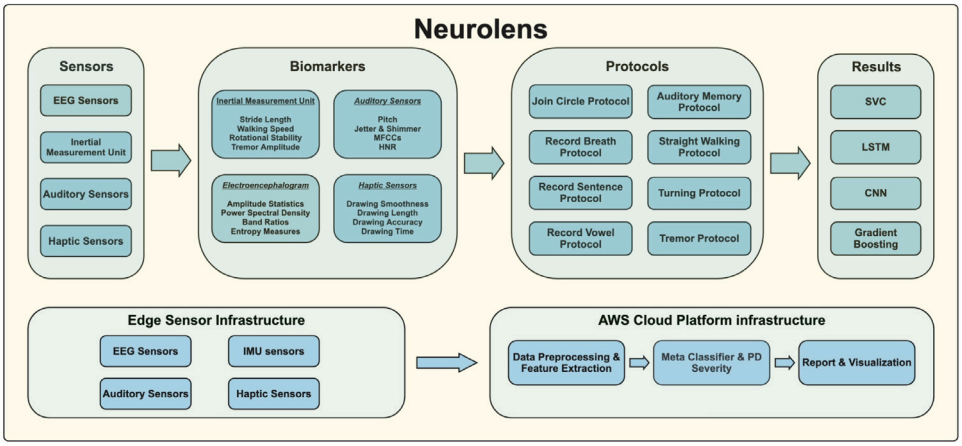


Fig. 2. NeuroLens system workflow: The diagram illustrates the integration of sensor modules, biomarker extraction, task-specific protocols, and cloud-based analysis, culminating in a personalized feedback system for Parkinson's Disease management.

Motor protocols focus on gait, balance, and tremor using smartphone-based inertial sensors during structured walking, turning, and tremor tasks, targeting hallmark PD motor symptoms including bradykinesia, postural instability, and involuntary movement patterns.

Haptic protocols evaluate fine motor control and cognitive planning through structured touchscreen tasks such as sequential circle-connecting exercises, reflecting impairments in coordination, hand-eye control, and executive function.

Neurophysiological protocols capture cortical activity via wearable EEG during resting state and task performance, providing neural-level biomarkers complementary to behavioral and motor signals.

Representative protocol illustrations are shown in Fig. 5.

3.3. Cloud Platform Analysis Module

The Cloud Platform Module transforms raw multimodal sensor data into clinically meaningful representations and integrates them for PD severity estimation. Incoming data streams undergo modality-specific preprocessing and feature extraction to characterize speech dynamics, gait and tremor behavior, neural activity patterns, and fine motor performance.

To address heterogeneity across modalities and task conditions, NeuroLens employs a meta-classifier that aggregates predictions from modality-specific models. This fusion strategy improves robustness to noisy or incomplete inputs and produces a unified PD severity score that reflects impairments across motor, cognitive, and auditory domains.

3.4. Feedback Module

The Feedback Module translates analytical results into clinically actionable insights for both patients and clinicians. Multimodal assessment outcomes are synthesized into a structured PD report summarizing symptom severity across key functional domains, including memory, reaction, auditory, and physical function.

A visual Parkinson Report, shown in Fig. 1, supports longitudinal symptom monitoring by categorizing functional performance into clinically interpretable severity levels. This design facilitates informed clinical decision-making and personalized disease management.

3.5. Edge Sensing Infrastructure

The Edge Sensing Infrastructure provides the hardware foundation for multimodal data acquisition, comprising four sensor categories that map directly to the protocol domains described in Section 3.2.

EEG sensors consist of a wearable electrode cap (shown in Fig. 3) that records multichannel neural signals. Signals are filtered using a bandpass filter and referenced to a common electrode to suppress artifacts before transmission. The EEG headset outputs ten neurophysiological indicators, including five brainwave band power values and five cognitive state indices, as summarized in Table 2.

Inertial Measurement Units (IMUs) are embedded in the participant's smartphone, providing tri-axial accelerometer and gyroscope data at 50 Hz to capture gait kinematics, postural dynamics, and tremor characteristics.

Audio sensors utilize the smartphone microphone to record speech and breathing signals at 16 kHz sampling rate, with ambient noise suppression applied locally.

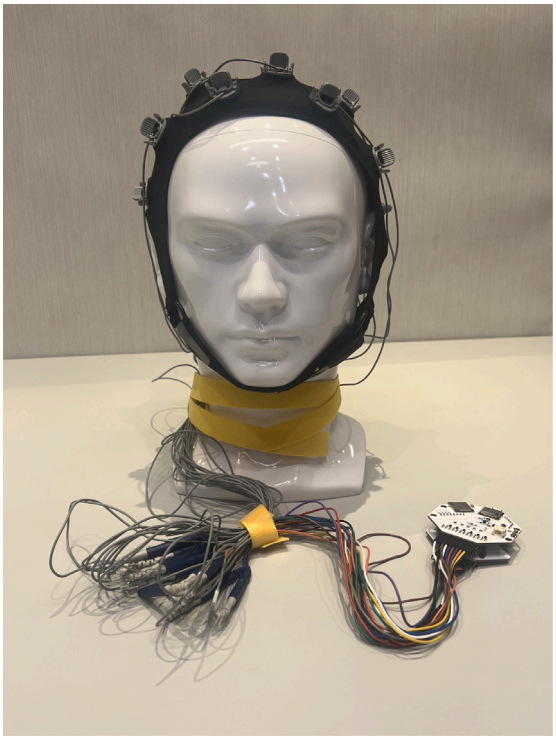


Fig. 3. EEG sensor system: A wearable head cap equipped with multiple electrodes for capturing brain activity signals.

Table 2
EEG output metrics: Brainwave band power and cognitive state indices.

No.	EEG metric	Value range
1	Alpha wave	8–25
2	Beta wave	15–70
3	Gamma wave	1–15
4	Delta wave	20–60
5	Theta wave	15–35
6	Attention	1–100
7	Stress	1–100
8	Emotion	−1, 1
9	Fatigue	1–100
10	Relaxation	1–100

Haptic interfaces leverage the capacitive touchscreen to record finger trajectory, pressure, and timing during fine motor tasks.

Lightweight preprocessing is performed on-device to reduce transmission latency, remove gross motion artifacts, and protect raw data privacy prior to cloud upload. Only preprocessed feature streams — rather than raw signals — are transmitted to the AWS platform, minimizing bandwidth overhead and supporting compliance with data privacy requirements (see Fig. 4).

3.6. Cloud platform infrastructure

The NeuroLens cloud platform is deployed on Amazon Web Services (AWS), supporting scalable data storage, efficient model inference, and real-time feedback delivery. Machine learning models are trained offline and deployed as containerized services to enable flexible updates and cost-efficient operation. Cloud-based monitoring ensures system reliability and responsiveness under real-world usage conditions.

4. Experiment

This section presents the design, data collection methods, and results of experiments conducted to evaluate the system's effectiveness in assessing Parkinson's disease symptoms. The experimental setup includes specific tasks targeting motor, cognitive, and auditory abilities, all administered through a mobile application. Data collected from these tasks are processed and analyzed using a Meta Classifier to produce an overall Parkinson's Disease severity score.

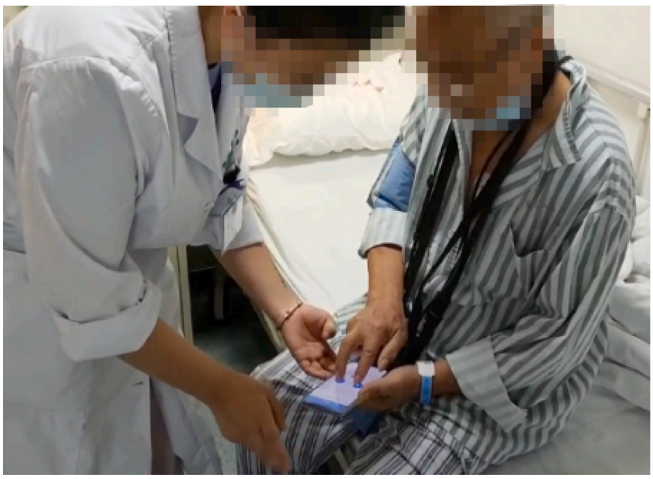


Fig. 4. A participant performing the Join Circle Protocol under the guidance of a clinician.

(a) Physical protocol. (b) Haptic protocol. (c) Auditory protocol.

Fig. 5. Illustrations of representative experimental protocols used in the Parkinson's Disease assessment system.

The dataset used in this study consists of detailed clinical and test information from 86 participants, including 64 individuals diagnosed with Parkinson's Disease and 22 healthy controls. These participants underwent comprehensive evaluations targeting motor functions and cognitive abilities. The dataset spans a wide age range (20–80 years) and includes an equal distribution of male and female participants within each group, ensuring diversity and balance in the study population. Key attributes include participant ID, age, gender, medication usage details, symptom profiles, and results from specific motor and cognitive tests.

Participants were categorized into five groups based on PD severity: 22 healthy controls (Level 0) and four patient levels representing increasing symptom severity (Level 1–4). The distribution was as follows: 34 patients at Level 1, 12 at Level 2, 10 at Level 3, and 8 at Level 4.

To ensure robust and unbiased performance estimation under the constraints of a limited sample size and pronounced class imbalance, we adopted **5-fold stratified cross-validation**. Stratified splitting was applied at the participant level to preserve the proportion of each severity class across all folds, preventing minority classes from being absent or underrepresented in any individual fold. In each iteration, four folds were used for training and one fold was held out for evaluation. All reported accuracy figures represent the mean performance averaged across the five folds. Crucially, splitting was performed at the participant level prior to any feature extraction, ensuring that no participant's data appears in both the training and evaluation subsets and thereby precluding data leakage across splits.

The experiment begins with a pre-test preparation phase where participants are required to provide information regarding their medication status to evaluate its potential impact on their test performance. They choose from four options: "Before Medication", "After Medication", "Other Times", or "I do not take Parkinson's medication". This is followed by a series of motor and cognitive evaluations, which include rhythm, auditory memory, speech, and drawing tests. Each test is designed to capture critical data related to motor coordination, cognitive function, speech clarity, and fine motor skills. Upon completing all test items, participants are asked to fill out three post-test questionnaires: the MDS-UPDRS survey (Evers et al., 2019), which assesses Parkinson's disease severity; the MMSE survey, which evaluates cognitive function; and a demographic information questionnaire that collects age, gender, and other relevant clinical information. The entire process, from pre-test preparation to post-test survey completion, ensures comprehensive data collection for subsequent analysis and supports a robust machine learning-based classification of healthy individuals and Parkinson's patients.

In this study, the cognitive ability of participants is assessed using the Mini-Mental State Examination (MMSE) (Arevalo-Rodriguez et al., 2021), and the results are categorized into labels based on their MMSE scores. Specifically, the MMSE scores are mapped to the following categories: a score between 24 and 30 is labeled as 1, indicating mild or no cognitive impairment; a score between

19 and 23 is labeled as 2, suggesting moderate cognitive impairment; a score between 10 and 17 is labeled as 3, indicating severe cognitive impairment; and a score below 10 is labeled as 4, representing very severe cognitive impairment. This categorization allows for the effective segmentation of participants based on cognitive function, supporting the development and evaluation of machine learning models designed to predict and classify motor and cognitive impairments in individuals with Parkinson's disease.

Each experiment is designed to evaluate a specific Parkinson's Disease symptom category. The mobile application captures multimodal data through auditory, motor, and haptic protocols, enabling a comprehensive assessment of Parkinson's Disease symptomatology. The primary tasks include:

- **Motor Tasks**: Straight Walking Protocol, Turning Protocol, and Tremor Protocol
- **Auditory Tasks**: Auditory Memory Protocol, Record Breath Protocol, Record Sentence Protocol, and Record Vowel Protocol
- **Haptic Task**: Join Circle Protocol, shown in Fig. 4

4.1. Mobile application test setup

All task protocols integrated into the mobile application were co-designed in close collaboration with experienced clinical experts. These experts guided the selection, structure, and duration of each task to ensure alignment with real-world diagnostic workflows and clinically relevant Parkinson's Disease (PD) symptom domains. This clinician-led design process ensures that the mobile assessments are both scientifically rigorous and suitable for early-stage PD screening and longitudinal symptom monitoring.

The mobile application evaluates motor, cognitive, and auditory functions through a set of standardized and guided tasks. Each protocol follows a structured interaction flow to ensure consistent data collection across participants while remaining intuitive for home-based use.

4.1.1. Auditory protocols

The auditory protocols assess speech production, respiratory control, and auditory memory—domains commonly affected in PD.

Auditory Memory Protocol. Participants listen to a short sequence of spoken words (3–5 items, difficulty-adjustable) and verbally recall the sequence after a brief delay. Response accuracy and recall latency provide indicators of auditory memory retention and cognitive processing speed.

Record Breath Protocol. Participants perform guided breathing cycles consisting of controlled inhalation and steady exhalation. The protocol captures respiratory duration and stability, which are associated with respiratory muscle rigidity and impaired breath control in PD.

Record Sentence Protocol. Participants read a standardized sentence aloud while the application records speech fluency, articulation, and pitch variation. This protocol targets speech impairments such as hypophonia and reduced prosody.

Record Vowel Protocol. Participants sustain a designated vowel sound for a fixed duration. Vocal steadiness, pitch stability, and sound duration are analyzed to detect vocal tremor and reduced vocal endurance.

4.1.2. IMU protocols

Motor protocols leverage smartphone inertial sensors to capture gait, balance, and tremor characteristics.

Straight Walking Protocol. Participants walk a predefined number of steps while carrying the smartphone in a stable position. Accelerometer and gyroscope data are used to extract gait-related metrics such as stride length, walking speed, and stability.

Turning Protocol. Participants perform guided turning movements following auditory prompts. Sensor data captures rotational speed, balance, and postural control, which are sensitive to PD-related instability during directional changes.

Tremor Protocol. Participants hold the smartphone steadily while seated for a fixed duration. Inertial signals are recorded to quantify tremor amplitude and frequency, enabling differentiation between resting and action tremors.

4.1.3. Haptic protocol

Join Circle Protocol. Participants draw a continuous path connecting sequential target circles on the touchscreen without lifting their finger. Path accuracy, smoothness, and deviations are analyzed to assess fine motor control, hand-eye coordination, and cognitive planning.

4.1.4. Data transmission and integration

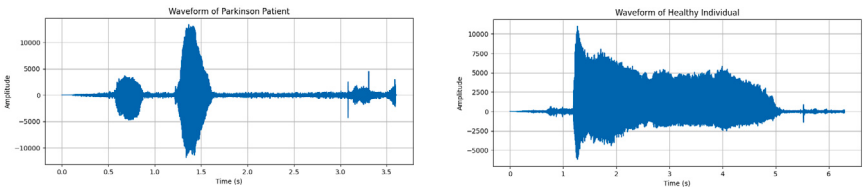
Upon completion of each task, data are securely transmitted to the cloud platform for processing and analysis. All records are timestamped and linked to participant identifiers to support traceable and longitudinal assessment. Modality-specific features are extracted server-side and integrated through a meta-classifier to produce a unified PD severity estimate.

4.2. Feature extraction and preprocessing

To assess Parkinson's Disease severity, we extracted and processed a variety of features from the auditory, motor, and haptic data collected from each test. These features aim to capture symptoms commonly associated with Parkinson's Disease, including vocal instability, motor dysfunction, and cognitive impairment.

Table 3
Definition of corresponding features and description of each variable.

Variable	Description
pitch	Fundamental frequency of the voice
jitter	Short-term frequency instability
shimmer	Short-term amplitude instability
hnr	Harmonics-to-noise ratio
mfcc	Mel Frequency Cepstral Coefficients
tremor	Low-frequency vocal tremor
dfa	Complexity of the vocal signal
rpde	Temporal stability of the signal
ppe	Phonation period variability
wavelet	Wavelet coefficients representing multi-scale features
gq	Glottal quotient of vocal fold vibration
gne	Glottal-to-noise excitation ratio
vfer	Vocal fold excitation ratio
totalTime	Total time to complete the task



(a) Parkinson's Disease vocal wave. (b) Health vocal wave.

Fig. 6. Comparison of vocal waveforms between a Parkinson's Disease patient and a healthy individual. The Parkinson's Disease patient's waveform demonstrates irregularities in vocal production, while the healthy individual's waveform shows smooth and consistent patterns. This contrast highlights the potential of vocal waveform analysis for distinguishing between Parkinson's Disease patients and healthy individuals.

4.2.1. Auditory features

Auditory features were extracted to evaluate vocal performance, respiratory control, and memory retention. Key auditory characteristics such as pitch, jitter, and shimmer provide insights into vocal stability, which is often affected in Parkinson's Disease patients. Additionally, advanced acoustic features like Mel Frequency Cepstral Coefficients (MFCCs) and wavelet coefficients capture the frequency and temporal nuances in participants' voices. The analysis includes measurements that reflect phonation quality (e.g., harmonics-to-noise ratio, glottal quotient), signal complexity (e.g., detrended fluctuation analysis), and vocal stability (e.g., phonation period variability). The full feature set is defined in Table 3.

Figs. 6(a) and 6(b) compare sustained vowel waveforms from a PD patient and a healthy individual. As shown in Fig. 6(a), the PD waveform exhibits pronounced irregularities in amplitude and frequency, indicative of vocal tremor, reduced vocal strength, and unstable phonation. In contrast, Fig. 6(b) displays a stable waveform with consistent amplitude and frequency, reflecting smooth phonation and preserved vocal control. Quantitative features such as jitter, shimmer, and HNR derived from these signals provide a robust, non-invasive basis for machine-learning-based PD differentiation.

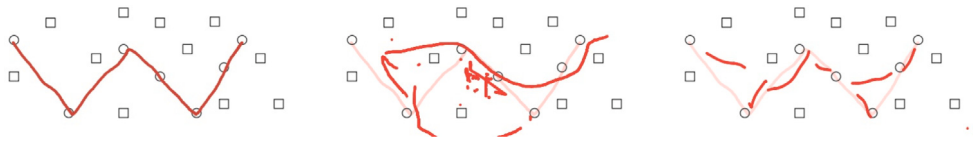
4.2.2. Motor and haptic features

Motor and haptic data were collected to analyze participants' movement patterns, balance, and fine motor skills—essential indicators of PD-related motor impairment. Motor features include stride length, walking speed, and balance stability metrics obtained from accelerometer, gyroscope, and magnetometer data. Haptic features from the Join Circle Protocol capture path accuracy, drawing smoothness, and error rate, evaluating fine motor control and cognitive planning.

Figs. 7(a), 7(b), and 7(c) illustrate progressive motor impairment across participants. The healthy subject (Fig. 7(a)) produces smooth, continuous trajectories connecting all target circles. A participant with moderate PD (Fig. 7(b)) shows frequent deviations indicative of tremor-related instability. Severe impairment (Fig. 7(c)) yields fragmented, incomplete paths revealing pronounced deficits in motor planning and coordination. Quantitative metrics such as path smoothness, deviation error, and trajectory continuity serve as discriminative features for severity classification.

4.3. Classification models and performance

Modality-specific classifiers were selected based on the temporal and statistical structure of each feature set. For motor tasks, Long Short-Term Memory (LSTM) networks were chosen due to their ability to model sequential dependencies in temporal data. Each walking sequence is represented as a multivariate time series $\mathbf{X}_i \in \mathbb{R}^{T\times d}$, where T is the number of time steps and d the number



(a) Health Join Circle. (b) PD Join Circle 1. (c) PD Join Circle 2.

Fig. 7. Comparison of Join Circle Protocol results across participants. (a) A healthy individual demonstrates smooth, continuous, and accurate lines connecting all target circles, indicating strong fine motor control and cognitive planning. (b) A Parkinson's Disease patient with moderate symptoms shows irregular paths and deviations from the target circles, reflecting reduced motor precision and tremor-induced instability. (c) A Parkinson's Disease patient with more severe symptoms exhibits fragmented and incomplete paths with poor continuity and accuracy, highlighting serious impairments in hand-eye coordination and fine motor control.

of sensor channels per step. Two stacked LSTM layers (64 units each) process the sequence; the final hidden state $\mathbf{h}_T$ is projected through a fully connected layer with softmax activation:

$$\hat{y}_i = \text{softmax}(\mathbf{W} \cdot \mathbf{h}_T + \mathbf{b}), \tag{1}$$

where $\mathbf{W}$ and $\mathbf{b}$ are trainable parameters and $\hat{y}_i \in \mathbb{R}^K$ is the predicted distribution over K severity levels. Training minimizes categorical cross-entropy:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{1}[y_i = k] \log \hat{y}_i^{(k)}. \tag{2}$$

For auditory and haptic tasks, whose feature vectors are high-dimensional and exhibit non-linear class boundaries, Support Vector Classifiers (SVCs) were selected. Given feature vectors $\{\mathbf{x}_i\}_{i=1}^N$ with severity labels $y_i \in \{0, 1, 2, 3, 4\}$, the SVC solves:

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \end{aligned} \tag{3}$$

where $\phi(\cdot)$ is the kernel-induced mapping and C is the regularization parameter. For tasks with high inter-subject variance in vocal features, Gradient Boosting was applied as an ensemble alternative. Their outputs are subsequently fused by the meta-classifier described in Section 4.4.

4.3.1. Motor symptoms

Motor symptoms were evaluated through gait, balance, and tremor tests:

- **Straight Walking Protocol**: Achieved **82%** accuracy by modeling stride length, step symmetry, and walking speed sequences. LSTMs capture the stride-to-stride temporal dependencies that rule-based methods cannot represent.
- **Turning Protocol**: Reached **96%** accuracy using gyroscope-derived rotational dynamics. The high sensitivity reflects that PD-related postural instability is strongly expressed during directional changes.
- **Tremor Protocol**: Attained **79%** accuracy. The comparatively lower figure reflects the non-stationary nature of resting tremor, which varies substantially across severity levels and individuals.

4.3.2. Auditory symptoms

Cognitive and auditory symptoms were evaluated through memory, speech, and respiratory control tasks:

- **Auditory Memory Protocol** (SVC): **73%** accuracy. Temporal stability and harmonic features capture cognitive processing speed, which degrades progressively in PD.
- **Record Breath Protocol** (SVC): **68%** accuracy. Respiratory irregularities are subtle; the SVC kernel helps separate overlapping class distributions in the feature space.
- **Record Sentence Protocol** (Gradient Boosting) and **Record Vowel Protocol** (SVC): both achieved moderate accuracy. Gradient Boosting was applied to sentence-reading features given their high inter-subject variance; ensemble averaging reduces overfitting on the relatively small dataset.

4.3.3. Haptic symptoms

The **Join Circle Protocol** (SVC) achieved **75%** accuracy. Path smoothness and deviation error exhibit non-linear relationships with severity level, making kernel-based SVC more effective than linear discriminants for this task.

Table 4
Per-class classification metrics of the meta-classifier (5-fold stratified cross-validation).

Class	n	Precision	Recall	F1
Level 0 (Healthy)	22	1.00	0.96	0.98
Level 1 (Mild)	34	0.97	0.94	0.96
Level 2 (Moderate)	12	0.85	0.92	0.88
Level 3 (Severe)	10	0.75	0.90	0.82
Level 4 (Profound)	8	0.86	0.75	0.80
Overall accuracy	–		**91.0%**	

4.4. Meta classifier integration

The meta classifier integrates predictions from individual models across all test modalities to generate a comprehensive Parkinson's Disease severity assessment. By leveraging complementary features extracted from motor, auditory, and cognitive tasks, the meta classifier enhances diagnostic performance and mitigates the limitations of single-task classifiers.

Formally, let $\hat{\mathbf{y}}_i^{(m)} \in \mathbb{R}^K$ denote the predicted probability distribution over K severity levels for the ith participant from the mth modality-specific classifier, where $m \in \{1, 2, \ldots, M\}$. The meta classifier performs soft voting by computing a weighted average over the individual predictions:

$$\hat{\mathbf{y}}_i = \sum_{m=1}^{M} \alpha_m \cdot \hat{\mathbf{y}}_i^{(m)}, \quad \text{with} \quad \sum_{m=1}^{M} \alpha_m = 1, \quad \alpha_m \geq 0 \tag{4}$$

The final predicted class label is obtained by:

$$\hat{y}_i = \arg\max_k \hat{\mathbf{y}}_i^{(k)} \tag{5}$$

To determine the optimal set of weights $\{\alpha_m\}$, we employed Bayesian optimization using the Optuna framework. Let the weight vector be $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_M] \in \Delta^{M-1}$, where Δ^{M-1} denotes the $(M-1)$-dimensional probability simplex. The goal is to maximize the validation accuracy $\mathcal{A}_{\text{val}}$ of the soft voting ensemble on a held-out set:

$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha} \in \Delta^{M-1}} \mathcal{A}_{\text{val}}(\boldsymbol{\alpha}) \tag{6}$$

Optuna builds a Tree-structured Parzen Estimator (TPE) to approximate the objective landscape. At each iteration t, it selects the next candidate $\boldsymbol{\alpha}_t$ by maximizing the acquisition function:

$$\boldsymbol{\alpha}_t = \arg\max_{\boldsymbol{\alpha}} \text{EI}(\boldsymbol{\alpha}; \mathcal{M}) \tag{7}$$

where $\text{EI}(\cdot)$ is the Expected Improvement, balancing exploration and exploitation.

This optimization proceeds iteratively until convergence or a fixed budget is reached. Compared to uniform or heuristic weighting, this principled approach allows the ensemble to adaptively emphasize more informative modalities while maintaining interpretability and simplicity.

By integrating soft voting with Bayesian-optimized weight tuning, the proposed meta-classifier achieves improved generalization and stability across varying PD severity levels. As shown in Fig. 9, this fusion strategy maintains high overall performance while substantially reducing level-wise performance fluctuations, enabling reliable severity stratification beyond binary disease detection.

Fig. 10 further illustrates accuracy distributions across PD severity levels. Although the *Turning Test* attains the highest single-task accuracy (93%), its performance varies markedly in early stages. In contrast, the multimodal classifier yields a slightly lower overall accuracy (91%) but demonstrates significantly greater consistency across all severity levels, as quantified by a balanced accuracy of **89.3%**. This stability indicates superior robustness to heterogeneous and subtle clinical manifestations, particularly in early-stage patients where unimodal signals are often weak.

To provide a rigorous evaluation under class imbalance, we report balanced accuracy, per-class precision, recall, and F1-score in addition to overall accuracy. Balanced accuracy is defined as the arithmetic mean of per-class recall:

$$\text{BA} = \frac{1}{K} \sum_{k=0}^{K-1} \frac{TP_k}{TP_k + FN_k}, \tag{8}$$

where TP_k and FN_k denote true positives and false negatives for class k, and $K = 5$. Table 4 summarizes the per-class metrics of the meta-classifier evaluated under 5-fold stratified cross-validation.

Fig. 8 presents the confusion matrix aggregated across all five folds. No errors span more than one stage boundary, and no healthy participant is misclassified as moderate or severe PD. This adjacency-constrained error pattern is clinically meaningful, as confusing neighboring stages carries substantially lower diagnostic risk than cross-boundary misclassification.

Statistical validation using one-way ANOVA across Levels 0–4, followed by Tukey's HSD test, confirms these observations. Unimodal models exhibit significant performance variation across severity levels ($p < 0.05$), whereas the multimodal classifier shows no statistically significant differences ($p > 0.05$), highlighting its robustness under real-world symptom variability.

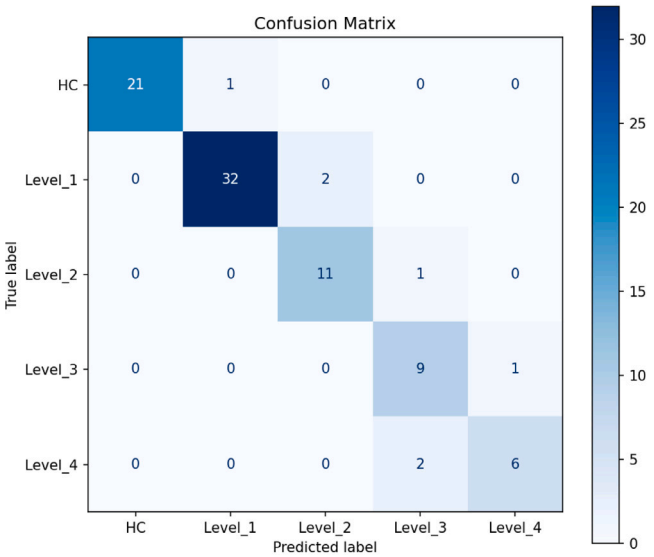


Fig. 8. Confusion matrix of the meta-classifier aggregated over 5-fold stratified cross-validation. Rows represent true severity levels; columns represent predicted levels. All misclassifications are confined to adjacent severity boundaries, with no cross-boundary errors between healthy controls and moderate-to-severe PD stages.

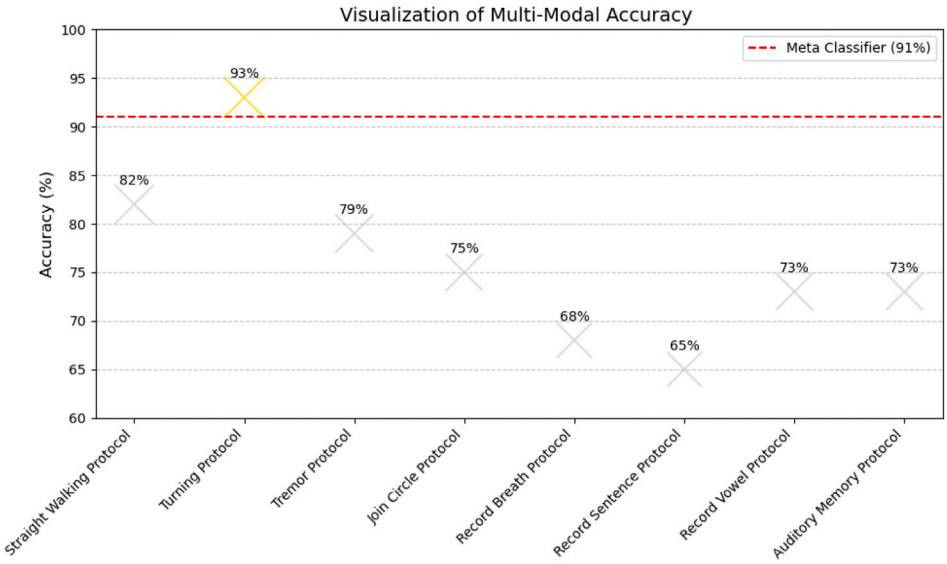


Fig. 9. Visualization of multi-modal accuracy. This plot shows the accuracy achieved by different tests, with the yellow circle representing the accuracy of the meta classifier. The dashed red line indicates the overall accuracy of the meta classifier (91%).

Error analysis reveals that unimodal models frequently misclassify early-stage subjects (Levels 1–2), reflecting limited sensitivity to subtle impairments. The multimodal classifier corrects a substantial portion of these errors, particularly in cases where gait and speech appear normal but neurological disruption is captured by complementary modalities. Residual errors are concentrated at the Level 2–3 boundary, consistent with the confusion matrix findings and known clinical symptom overlap in mid-stage PD. Notably, even weaker modalities (e.g., breathing and sustained vowel tasks) contribute complementary information when fused, reducing early-stage false negatives and strengthening overall prediction.

Compared with existing approaches, NeuroLens achieves competitive performance (91% overall, 89.3% balanced) using lightweight, non-invasive, and easily deployable sensors. While Hannan (2024) reports higher overall accuracy (~94%), their framework relies on MRI and PET imaging, limiting scalability and suitability for routine or remote screening. NeuroLens instead offers a practical multimodal alternative without compromising clinical relevance.

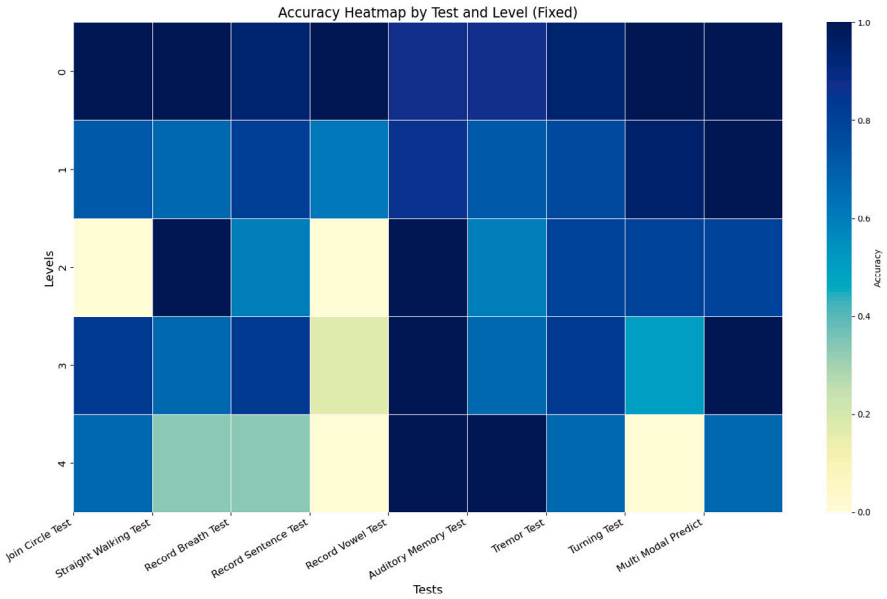


Fig. 10. Accuracy heatmap by test and level. The heatmap visualizes the accuracy achieved by different tests at various levels, with color intensity representing accuracy. Higher accuracy is shown by darker blue shades, while lower accuracy is represented by lighter shades. The analysis emphasizes the varied performance of each test across different levels, showcasing the advantages of a multimodal approach in enhancing the overall evaluation of Parkinson's Disease.

4.5. Discussion

This study presents a significant advancement in the detection and assessment of Parkinson's Disease by introducing a comprehensive, multimodal platform. The experimental results underscore the advantages of this approach, particularly in addressing limitations inherent to single-modality methods. Traditional tests are effective at identifying advanced symptoms of Parkinson's Disease, but they often lack the sensitivity needed for detecting subtle, early-stage indicators. NeuroLens fills this gap by integrating motor, auditory, and cognitive data, providing a holistic view of Parkinson's Disease symptoms and improving diagnostic accuracy.

The contribution of this work lies in its ability to leverage complementary data modalities to achieve superior diagnostic outcomes. Motor-based tasks, such as the Straight Walking Protocol, Turning Protocol, and Tremor Protocol, performed exceptionally well in detecting physical symptoms associated with mid- to late-stage Parkinson's Disease. For instance, the Turning Protocol achieved a remarkable 93% accuracy in identifying balance issues, a symptom prevalent in advanced Parkinson's Disease. However, motor tasks alone fall short in capturing cognitive impairments or subtle vocal fluctuations that are often present in the disease's early stages.

To address these limitations, NeuroLens incorporates auditory and haptic-cognitive assessments, which demonstrated the platform's capacity to identify early cognitive and vocal changes. These features, while challenging to detect in isolation, serve as critical early indicators of Parkinson's Disease. For example, the Auditory Memory Protocol and Record Breath Protocol provided essential data when integrated into the meta-classifier model, enhancing sensitivity to early cognitive and respiratory symptoms. Specifically, auditory features such as pitch stability, jitter, and shimmer were effective in detecting subtle vocal instabilities, highlighting the utility of this multimodal approach.

The development and implementation of the meta-classifier further exemplify the study's contribution. By integrating predictions from individual tasks, the meta-classifier achieved an overall accuracy of 91%, mitigating the limitations of single modalities and enabling a comprehensive assessment of motor precision, vocal clarity, and cognitive planning abilities. This integration not only enhances diagnostic accuracy but also supports early intervention strategies, with the potential to significantly improve patient outcomes.

Beyond diagnostic improvements, this study demonstrates the feasibility of an accessible, cloud-based multimodal platform for real-time Parkinson's Disease evaluation. By leveraging mobile applications for data collection and AWS cloud infrastructure for processing, NeuroLens ensures scalability and real-time feedback. This design empowers both patients and clinicians with actionable insights, enabling continuous monitoring and tailored interventions.

5. Conclusion

NeuroLens presents a novel solution for early-stage Parkinson's Disease detection by addressing the limitations of traditional diagnostic methods. Traditional diagnostic techniques, which often rely on subjective evaluations and specialized equipment, face

challenges in terms of both accuracy and accessibility, especially during the early stages of Parkinson's Disease when symptoms may be less apparent. NeuroLens overcomes these challenges by integrating EEG, IMU sensors, and audio data into a cloud-based platform, enhancing diagnostic precision and enabling the quantification of Parkinson's Disease severity. This multimodal approach not only improves detection but also facilitates the assessment of medication effectiveness over time. Achieving 91% accuracy in distinguishing Parkinson's Disease patients from healthy individuals, NeuroLens offers a non-invasive, scalable, and cost-effective tool for early detection and continuous monitoring of Parkinson's Disease symptoms.

In addition to its diagnostic capabilities, NeuroLens integrates real-time data processing via AWS cloud infrastructure, which ensures both accessibility and the generation of actionable visual diagnostic reports. This platform provides healthcare providers with immediate insights, enabling dynamic care adjustments. NeuroLens's ability to detect subtle, early-stage symptoms through its multimodal approach makes it a superior tool for early intervention, crucial for slowing disease progression and improving patient outcomes. Future work will focus on refining its sensitivity to cognitive and vocal symptoms, further enhancing the platform's diagnostic capabilities and positioning NeuroLens as a vital tool in the ongoing management and detection of Parkinson's disease.

For the future development, we plan to expand our patient dataset to include a broader range of individuals at various stages of Parkinson's Disease. This will improve the robustness of the system, ensuring its reliability in early diagnosis. Additionally, future development will focus on validating the effects of medication in real time. We aim to integrate longitudinal data to evaluate symptom changes before and after treatment, providing healthcare providers with actionable insights to adjust therapies and improve patient outcomes. These efforts will ensure that NeuroLens continues to evolve as a comprehensive tool for both diagnosing and managing Parkinson's Disease.

CRediT authorship contribution statement

Zixu Geng: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Shih-Ching Yeh:** Investigation, Writing – review & editing. **Xiaowei Xu:** Data curation, Investigation, Resources, Validation. **Jian Zhuang:** Data curation, Investigation, Resources, Validation. **Jingwei Li:** Conceptualization, Visualization, Writing – original draft. **Ruitian Wu:** Conceptualization, Methodology, Software. **Feng Lin:** Resources, Supervision. **Longxiang Pan:** Resources. **Wenyao Xu:** Resources, Supervision. **Ming-Chun Huang:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

We used generative AI tools only to improve language clarity, check grammar, and assist with formatting references. All scientific content, study design, data analysis, and conclusions are the authors' own work and were carefully verified by the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research results reported in this article were sponsored by the Kunshan Municipal Government Research Funding. Approval of all ethical and experimental procedures and protocols involving human subjects was granted by the Duke Kunshan University Institutional Review Board (DKU IRB) under IRB No. FWA00021580.

Data availability

Data will be made available on request.

References

Arevalo-Rodriguez, I., Smailagic, N., Roqué-Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Cosp, X. B., & Cullum, S. (2021). Mini-mental state examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, (7).

Armstrong, M. J., & Okun, M. S. (2020). Diagnosis and treatment of Parkinson disease: A review. *Jama*, *323*(6), 548–560.

Balestrino, R., & Schapira, A. (2020). Parkinson disease. *European Journal of Neurology*, *27*(1), 27–42.

Bhidayasiri, R., Tarsy, D., Bhidayasiri, R., & Tarsy, D. (2012). Parkinson's disease: Hoehn and yahr scale. In *Movement Disorders: A Video Atlas: A Video Atlas* (pp. 4–5). Springer.

Bloem, B. R., Okun, M. S., & Klein, C. (2021). Parkinson's disease. *The Lancet*, *397*(10291), 2284–2303.

Brognara, L., Palumbo, P., Grimm, B., & Palmerini, L. (2019). Assessing gait in Parkinson's disease using wearable motion sensors: A systematic review. *Diseases*, *7*(1), 18.

Cai, Y., Qian, X., Cao, H., Zheng, J., Xu, W., & Huang, M.-C. (2022). Mhealth technologies toward active health information collection and tracking in daily life: A dynamic gait monitoring example. *IEEE Internet of Things Journal*, *9*(16), 15077–15088.

de Bie, R. M., Clarke, C. E., Espay, A. J., Fox, S. H., & Lang, A. E. (2020). Initiation of pharmacological therapy in Parkinson's disease: When, why, and how. *The Lancet Neurology*, *19*(5), 452–461.

Di Biase, L., Di Santo, A., Caminiti, M. L., De Liso, A., Shah, S. A., Ricci, L., & Di Lazzaro, V. (2020). Gait analysis in Parkinson's disease: An overview of the most accurate markers for diagnosis and symptoms monitoring. *Sensors*, *20*(12), 3529.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artificial Intelligence in Medicine*, *67*, 39–46.

Evers, L. J., Krijthe, J. H., Meinders, M. J., Bloem, B. R., & Heskes, T. M. (2019). Measuring Parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS. *Movement Disorders*, *34*(10), 1480–1487.

Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., & Rochester, L. (2014). Accuracy of the microsoft kinect sensor for measuring movement in people with Parkinson's disease. *Gait & Posture*, *39*(4), 1062–1068.

Hannan, S. A. (2024). Advancing Parkinson's disease severity prediction using multimodal convolutional recursive deep belief networks. *International Journal of Advanced Computer Science & Applications*, *15*(2).

Kaegi, G., Bhatia, K. P., & Tolosa, E. (2010). The role of DAT-SPECT in movement disorders. *Journal of Neurology, Neurosurgery & Psychiatry*, *81*(1), 5–12.

Li, H., Chen, H., Xu, C., Li, Z., Zhang, H., Qian, X., Li, D., Huang, M.-c., & Xu, W. (2023). Neuralgait: Assessing brain health using your smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(4), 1–28.

Li, H., Qian, X., Ma, R., Xu, C., Li, Z., Li, D., Lin, F., Huang, M.-C., & Xu, W. (2023). TherapyPal: Towards a privacy-preserving companion diagnostic tool based on digital symptomatic phenotyping. In *Proceedings of the 29th annual international conference on mobile computing and networking* (pp. 1–15).

Lim, W.-S., Fan, S.-P., Chiu, S.-I., Wu, M.-C., Wang, P.-H., Lin, K.-P., Chen, Y.-M., Peng, P.-L., Jang, J.-S. R., & Lin, C.-H. (2025). Smartphone-derived multidomain features including voice, finger-tapping movement and gait aid early identification of Parkinson's disease. *Npj Parkinson's Disease*, *11*(1), 111.

Mahlknecht, P., Hotter, A., Hussl, A., Esterhammer, R., Schocke, M., & Seppi, K. (2010). Significance of MRI in diagnosis and differential diagnosis of Parkinson's disease. *Neurodegenerative Diseases*, *7*(5), 300–318.

Martinez-Martin, P., Chaudhuri, K., Rojo-Abuin, J., Rodriguez-Blazquez, C., Alvarez-Sanchez, M., Arakaki, T., Bergareche-Yarza, A., Chade, A., Garretto, N., Gershanik, O., et al. (2015). Assessing the non-motor symptoms of Parkinson's disease: MDS-UPDRS and NMS scale. *European Journal of Neurology*, *22*(1), 37–43.

Martínez-Martín, P., Rodríguez-Blázquez, C., Alvarez, M., Arakaki, T., Arillo, V. C., Chaná, P., Fernández, W., Garretto, N., Martínez-Castrillo, J. C., Rodríguez-Violante, M., et al. (2015). Parkinson's disease severity levels and MDS-unified Parkinson's disease rating scale. *Parkinsonism & Related Disorders*, *21*(1), 50–54.

Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: A review of literature. *Frontiers in Aging Neuroscience*, *13*, Article 633752.

Mortezazadeh, T., Seyedarabi, H., Mahmoudian, B., & Islamian, J. P. (2021). Imaging modalities in differential diagnosis of Parkinson's disease: opportunities and challenges. *Egyptian Journal of Radiology and Nuclear Medicine*, *52*, 1–12.

Navita, Mittal, P., Sharma, Y. K., Rai, A. K., Simaiya, S., Lilhore, U. K., & Kumar, V. (2025). Gait-based Parkinson's disease diagnosis and severity classification using force sensors and machine learning. *Scientific Reports*, *15*(1), 328.

Pfeiffer, R. F. (2016). Non-motor symptoms in Parkinson's disease. *Parkinsonism & Related Disorders*, *22*, S119–S122. http://dx.doi.org/10.1016/j.parkreldis.2015.09.004, URL https://www.sciencedirect.com/science/article/pii/S135380201500379X, Proceedings of XXI World Congress on Parkinson's Disease and Related Disorders, December 6-9, 2015, Milan, Italy.

Ramesh, V., & Bilal, E. (2022). Detecting motor symptom fluctuations in Parkinson's disease with generative adversarial networks. *NPJ Digital Medicine*, *5*(1), 138.

Regnault, A., Boroojerdi, B., Meunier, J., Bani, M., Morel, T., & Cano, S. (2019). Does the MDS-UPDRS provide the precision to assess progression in early Parkinson's disease? Learnings from the Parkinson's progression marker initiative cohort. *Journal of Neurology*, *266*, 1927–1936.

Sánchez-Ferro, Á., Elshehabi, M., Godinho, C., Salkovic, D., Hobert, M. A., Domingos, J., Van Uem, J. M., Ferreira, J. J., & Maetzler, W. (2016). New methods for the assessment of Parkinson's disease (2005 to 2015): A systematic review. *Movement Disorders*, *31*(9), 1283–1292.

Skorvanek, M., Martinez-Martin, P., Kovacs, N., Zezula, I., Rodriguez-Violante, M., Corvol, J.-C., Taba, P., Seppi, K., Levin, O., Schrag, A., et al. (2018). Relationship between the MDS-UPDRS and quality of life: A large multicenter study of 3206 patients. *Parkinsonism & Related Disorders*, *52*, 83–89.

Tong, J., Zhang, J., Dong, E., & Du, S. (2021). Severity classification of Parkinson's disease based on permutation-variable importance and persistent entropy. *Applied Sciences*, *11*(4), 1834.

Vásquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J. R., Eskofier, B., Klucken, J., & Nöth, E. (2018). Multimodal assessment of Parkinson's disease: A deep learning approach. *IEEE Journal of Biomedical and Health Informatics*, *23*(4), 1618–1630.

Wang, Y., Beuving, F., Nonnekes, J., Cohen, M. X., Long, X., Aarts, R. M., & Van Wezel, R. (2020). Freezing of gait detection in Parkinson's disease via multimodal analysis of EEG and accelerometer signals. In *2020 42nd annual international conference of the IEEE engineering in medicine & biology society* (pp. 847–850). IEEE.

Xia, R., & Mao, Z.-H. (2012). Progression of motor symptoms in Parkinson's disease. *Neuroscience Bulletin*, *28*, 39–48.

Zhang, Y., & Gu, D. (2019). A deep convolutional-recurrent neural network for freezing of gait detection in patients with Parkinson's disease. In *2019 12th international congress on image and signal processing, bioMedical engineering and informatics* CISP-bMEI, (pp. 1–6). IEEE.

Zhang, H., Song, C., Rathore, A. S., Huang, M.-C., Zhang, Y., & Xu, W. (2020). Mhealth technologies towards Parkinson's disease detection and monitoring in daily life: A comprehensive review. *IEEE Reviews in Biomedical Engineering*, *14*, 71–81.

Zhang, W., Yang, Z., Li, H., Huang, D., Wang, L., Wei, Y., Zhang, L., Ma, L., Feng, H., Pan, J., et al. (2022). Multimodal data for the detection of freezing of gait in Parkinson's disease. *Scientific Data*, *9*(1), 606.