



## Survey paper

## Parkinson's disease detection based on artificial intelligence: Methodologies, datasets, clinical applications, challenges and future directions

Xingkai Fu<sup>a</sup>, Sike Ni<sup>a, b</sup>, Mohammed A.A. Al-qaness<sup>a, b</sup>,\*<sup>a</sup> College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321004, China<sup>b</sup> Zhejiang Institute of Optoelectronics, Jinhua 321004, China

## ARTICLE INFO

## Keywords:

Parkinson's disease  
Artificial intelligence  
Deep learning  
Wearable sensors  
Clinical translation  
Multimodal data

## ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder, and its early diagnosis is crucial for delaying disease progression and improving patients' quality of life. In recent years, artificial intelligence (AI) technologies, particularly machine learning (ML), including deep learning (DL), have offered innovative approaches for the early detection and monitoring of PD by analyzing multimodal data such as voice, gait, electroencephalogram (EEG), and medical imaging. This paper provides a comprehensive review of AI-based techniques for PD diagnosis, covering wearable device-based motion monitoring, voice analysis, automated image interpretation, and multimodal data fusion. It also summarizes the application of traditional ML algorithms and DL architectures in PD-related studies. Furthermore, the study analyzes publicly available datasets, discusses challenges in data preprocessing and model generalization, and highlights the potential of AI to enhance clinical diagnostic support and disease monitoring in PD management.

## 1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder of the central nervous system that affects millions of older adults worldwide and substantially diminishes quality of life (ul Haq et al., 2022). Hallmark features include bradykinesia, resting tremor, rigidity, and abnormalities in posture and gait (Diaz et al., 2021; Cao et al., 2021). These motor impairments often lead to disability (Wang et al., 2024; Silva et al., 2023).

Fig. 1 delineates the cardinal motor and nonmotor manifestations of PD, which are pathophysiologically associated with progressive degeneration of nigrostriatal dopaminergic neurons (Seiler et al., 2024; Moons and De Groef, 2022). Dopamine (DA), a crucial neurotransmitter released by these neurons, plays a pivotal role in modulating motor control and cognitive processes (Lelos et al., 2023; Cramb et al., 2023; Chu et al., 2022). In particular, the decrease in this neurotransmitter in the initial disease phase is often imperceptible, contributing to the insidious onset and gradual progression of clinical symptoms (Zhang, 2022).

This pathophysiological profile imposes a dual challenge on the early identification of PD. In the initial stages, patients may experience nonspecific symptoms such as mild tremors and bradykinesia, which are easily ignored (Subasree et al., 2025). The diagnostic delay window between the onset of the symptoms and the formal diagnosis directly affects the optimal therapeutic intervention opportunity. As neurons

continue to be lost, patients develop basic activities of daily living impairments, such as dressing and eating difficulties. This causes a drastic reduction in their quality of life and exponential increases in caregiving costs for families and society (Tolosa et al., 2021). To overcome this dilemma, creating an effective early diagnosis system of PD has become a research focus (Almeida et al., 2019). Evidence shows that early intervention slows the progression of PD, but achieving this requires overcoming technical bottlenecks in traditional diagnosis (Zhou et al., 2023). For example, in pharmacotherapy, levodopa, although the core symptomatic drug, carries a risk of motor complications, leading to a delayed prescription strategy and highlighting the need for a precise early diagnosis (Aghanavesi et al., 2019).

These challenges motivate objective and continuous measurements that complement clinician-rated scales and enable earlier, more precise assessments. In this context, AI is reshaping PD diagnosis through multi-dimensional biomarker mining (Kumar and Ghosh, 2025; Morgan et al., 2023). Wearable devices can continuously quantify movement parameters such as tremor frequency and gait symmetry (Chen et al., 2023), while video analysis systems capture micro-expressions and posture control abnormalities (Sarapata et al., 2023; Talitckii et al., 2022). Deep learning models integrate multi-modal data (e.g., voice and gait) to build predictive systems (Skaramagkas et al., 2023; Vázquez-Correa et al., 2018). These advances enable objective symptom quantification

\* Corresponding author at: College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321004, China.

E-mail addresses: [Xingkai\\_fu@zjnu.edu.cn](mailto:Xingkai_fu@zjnu.edu.cn) (X. Fu), [nisikee@zjnu.edu.cn](mailto:nisikee@zjnu.edu.cn) (S. Ni), [alqaness@zjnu.edu.cn](mailto:alqaness@zjnu.edu.cn) (M.A.A. Al-qaness).

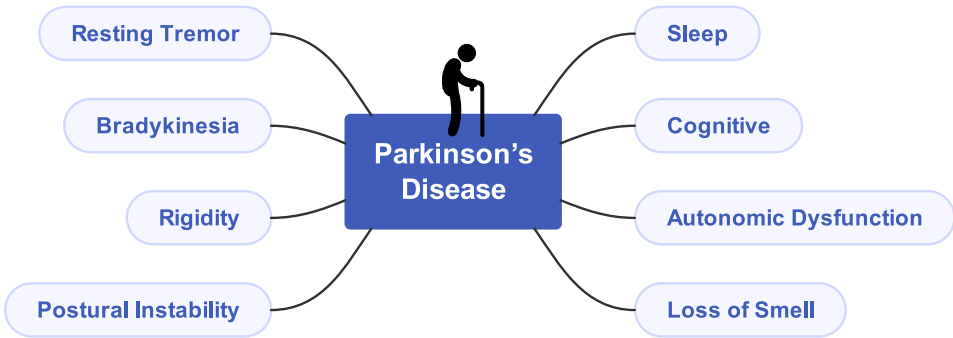


Fig. 1. Main symptoms of PD (motor and non-motor).

**Table 1**  
Advantages/contributions of existing surveys.

Study	Coverage years	Contributions
ul Haq et al. (2022)	2008–2021	<ul style="list-style-type: none"><li>*Followed PRISMA guidelines to conduct a structured review.</li><li>*Summarized characteristics, usage frequency, and strengths/limitations of commonly used PD datasets; compared evaluation metrics.</li><li>*Discussed major challenges (e.g., data scarcity, limited interpretability) and outlined potential solutions including federated learning and GAN-based augmentation.</li></ul>
Skaramagkas et al. (2023)	2016–2023	<ul style="list-style-type: none"><li>*Reviewed performance of advanced deep learning models across major PD diagnostic tasks.</li><li>*Highlighted effective multimodal fusion frameworks and argued for fusion as a key direction to overcome single-modality limitations.</li><li>*Systematized sensor types (IMU, pressure insoles, microphones, cameras), placement, and sampling specifications relevant to deployment.</li></ul>
Islam et al. (2024)	2000–2023	<ul style="list-style-type: none"><li>*Summarized methods and datasets with usage frequency and compared reported metrics.</li><li>*Provided cross-study performance comparisons of ML/DL models under similar datasets and metrics to indicate baseline levels.</li><li>*Discussed potential of multimodal approaches (e.g., voice plus handwriting) for PD diagnosis.</li></ul>
Altham et al. (2024)	2008–2024	<ul style="list-style-type: none"><li>*Analyzed multiple modalities (imaging, EEG, voice, clinical and neuropsychological data) for cognitive impairment detection in PD, noting gains from multimodal fusion.</li><li>*Summarized performance of ML families (RF, SVM, DL) across tasks, with ensemble and DL methods frequently strongest.</li><li>*Reported MoCA outperforming MMSE for cognitive screening and recommended using large public datasets (e.g., PPMI) for model development and validation.</li></ul>
Sigcha et al. (2023)	2012–2022	<ul style="list-style-type: none"><li>*Systematic review of deep learning and wearable-sensor approaches for PD diagnosis covering motor and non-motor symptoms.</li><li>*Outlined commonly used architectures (CNN, RNN, LSTM) and tasks.</li><li>*Applied TRIPOD to assess study quality; Most of studies exhibited high risk of bias, largely due to lack of external validation.</li></ul>
Our (2025)	2018–2025	<ul style="list-style-type: none"><li>*This paper encompasses AI-based PD detection methods, ranging from conventional ML algorithms to state-of-the-art DL models.</li><li>*It delves into the features of public datasets with a focus on data quality and completeness.</li><li>*The auxiliary role of AI in PD diagnosis is elaborated.</li><li>*The application of AI in disease progression monitoring is discussed.</li><li>*Key areas for future research, including multimodal data integration and explainable model development, are proposed.</li></ul>

and expand disease management through remote monitoring, marking a new data-driven era in PD diagnosis and therapeutic intervention (Freire-Álvarez et al., 2025; Bhandari et al., 2023).

1.1. Motivation and contribution

The severity of PD is typically assessed using the Movement Disorder Society-Unified PD Rating Scale (MDS-UPDRS) (Goetz et al., 2008). This unified rating scale provides a standardized, objective framework for assessing patients’ symptoms, aiding physicians in better understanding and monitoring changes in patients’ conditions (Balaji et al., 2020). Although several studies have demonstrated high

test-retest reliability of the UPDRS motor scores, this assessment tool still has limitations in practical clinical trials and patient care (Parisi et al., 2015). For instance, the UPDRS scoring depends on the clinical expertise and experience of healthcare providers, which may be influenced by subjective judgment. Furthermore, the standardized scoring method typically relies on regular clinical visits and patient cooperation, limiting flexibility and real-time application in long-term disease management (Exley et al., 2022).

In this context, early prediction based on sensor technology and AI models, particularly ML and DL, offers new directions for clinical monitoring and management of PD (Kumar and Ghosh, 2025; Pragadeeswaran and Kannimuthu, 2024). Modern wearable devices can

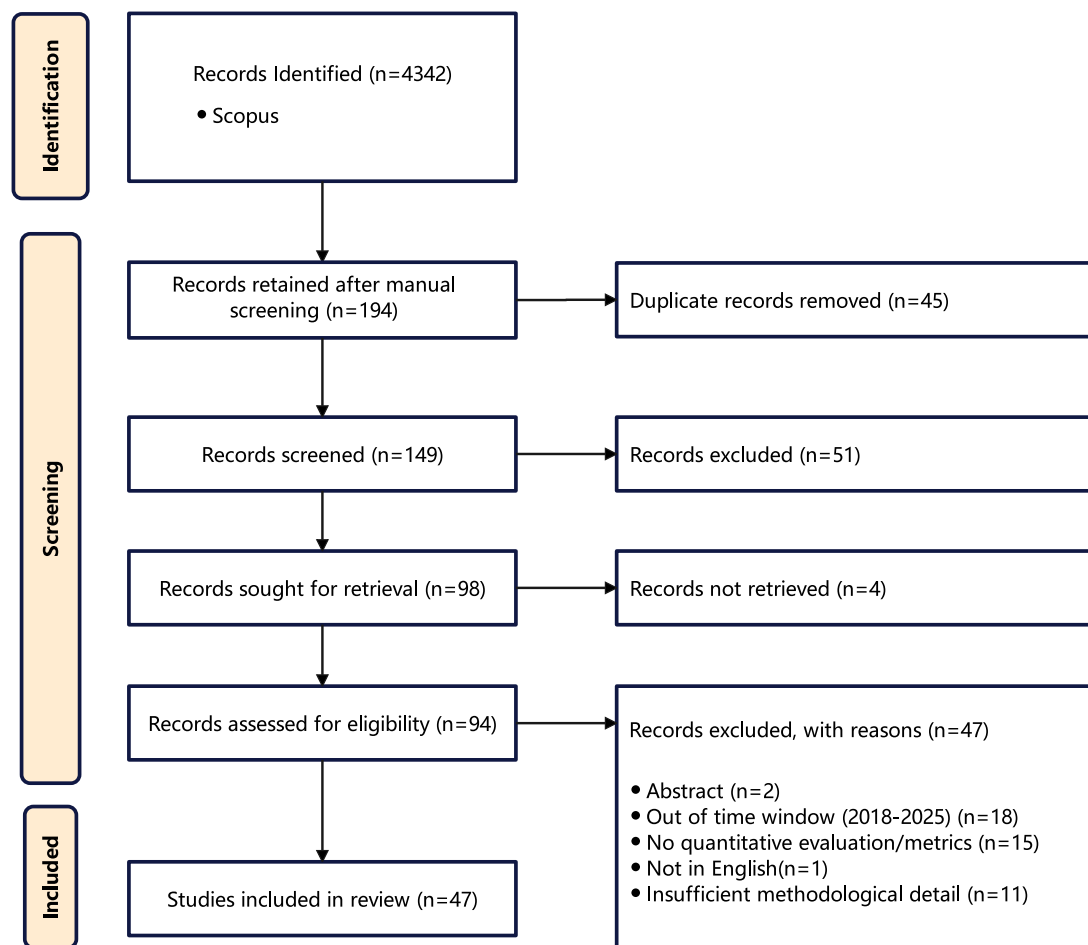


Fig. 2. PRISMA flow diagram of study selection, detailing identification, de-duplication, screening, full-text eligibility assessment, and final inclusion of studies.

monitor patients' motor conditions in real-time, collecting multidimensional data such as gait, tremor, bradykinesia, and more (Huang et al., 2023a). These technologies enable more accurate and continuous tracking of disease progression, addressing the limitations of traditional assessment methods.

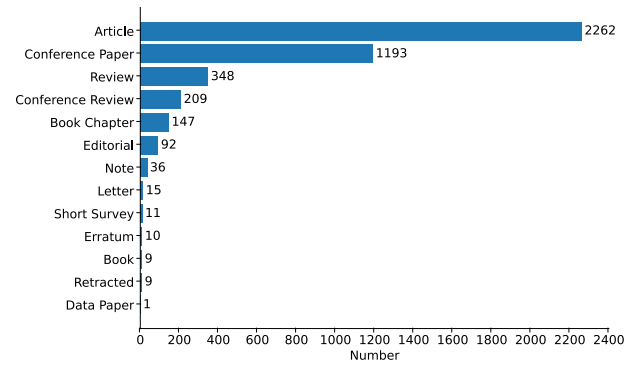
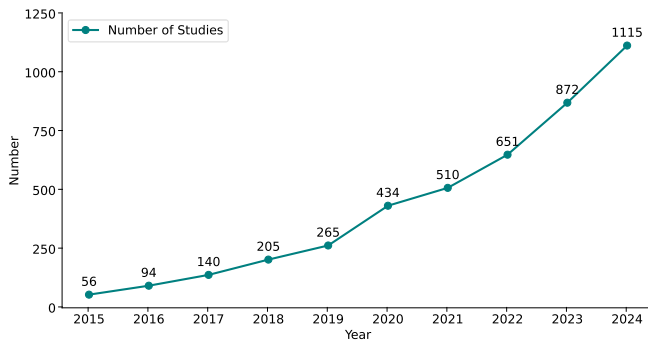
This study offers a thorough review and analysis of the latest advancements in AI models for PD prediction. We discuss different sensor technologies, such as accelerometers and gyroscopes, and their use in monitoring motor symptoms and assessing non-motor symptoms. We also concentrate on predictive models based on ML and DL and their real-world clinical applications. Through this study, we aim to provide a comprehensive reference framework for future researchers, clinicians, and technology developers, thereby promoting early diagnosis, real-time monitoring, and personalized treatment of PD. Ultimately, this paper aims to offer more precise and timely support for PD patients' disease management, enhancing their quality of life and facilitating the implementation of more scientific and personalized clinical treatment plans. With ongoing technological advancements and improvements in data analysis methods, more intelligent and efficient disease monitoring and management will be achievable in the future, protecting the health of PD patients.

This paper aims to supplement existing literature and summarize the contemporary implementations of artificial intelligence and portable sensor technologies in diagnosing, predicting disease trajectories, and longitudinal tracking of PD progression. By examining the latest research findings, we hope to enhance the integration into clinical practice, real-world feasibility, and transformative potential of these technologies as supportive tools in PD management. This study covers relevant research articles published from 2018 to 2025 focusing on DL

methods and portable sensor technologies for PD diagnostic evaluation and continuous physiological surveillance.

Table 1 provides an overview and comparison of our work and previous review works, which cover different research time spans, as well as the limitations and contributions of past research. First, multimodal fusion is repeatedly highlighted as a path to overcome single-modality limitations. At the same time, deployment-relevant sensor specifications (types, placement, sampling) are documented to varying depth (notably in Skaramagkas et al. (2023)). Second, evaluation setups remain heterogeneous — splitting policies, reported metrics, and the prevalence of external validation differ widely — complicating cross-study comparability (surveyed in ul Haq et al. (2022), Islam et al. (2024)). Third, quality assessments indicate substantial risk of bias when external validation is absent (Sigcha et al., 2023). Complementary emphases are visible: (ul Haq et al., 2022) linked dataset/metric auditing with potential remedies (federated learning, GAN augmentation). Islam et al. (2024) provided harmonized baselines under shared datasets/metrics. Altham et al. (2024) evidenced benefits of multimodal approaches for cognitive impairment detection and offers pragmatic guidance. The main contributions of this paper are outlined as follows:

- This study delivers a methodical synthesis and systematic review of AI-driven methodologies for PD detection. It covers various techniques, including traditional ML algorithms and advanced DL models, along with their applications.
- We explored the characteristics of publicly available datasets, which is essential for enhancing the accuracy and reliability of AI models.



**Fig. 3.** Annual distribution of parkinson's disease research literature (2015–2024) using keywords ((parkinson's AND disease OR parkinson) AND (machine learning)).

- The practical applications of AI in clinical settings were explored, highlighting its auxiliary role in disease diagnosis and progression monitoring. This focus on clinical applicability offers essential support for connecting research outcomes with real-world healthcare solutions.
- This paper identifies crucial areas for future research. These directions are designed to overcome current limitations and boost the clinical application of AI in PD treatment.

### 1.2. Paper selection criteria

To map the literature on artificial intelligence-based detection of Parkinson's disease, we conducted a structured search. The databases queried were IEEE Xplore, ScienceDirect, PubMed, Scopus, and Google Scholar (as a supplementary source). The core query used the terms “parkinson's”, “disease”, and “machine learning”. Records published before 2018 and studies of low methodological quality were excluded. Duplicates were removed, and relevance was confirmed through manual screening of titles/abstracts and, when required, full texts. Based on Scopus counts, Fig. 3 visualizes annual publications from 2015 to 2024, revealing a pronounced increase in the last four years, which indicates growing research attention to this topic. Across the observation period, journal articles and conference papers constitute the majority of outputs, whereas review articles remain comparatively scarce. This gap highlights the need for a comprehensive synthesis to clarify current trends, methodological challenges, and future directions; the present paper addresses this need by providing an integrated assessment of recent advances in AI-based detection of Parkinson's disease. In addition, Fig. 2 presents the PRISMA flow diagram for the literature search and screening process.

### 1.3. Paper structure

The remainder of this article is organized as follows. Section 2 examines wearable devices, ML, and DL. Section 3 reviews datasets for PD symptom detection and discusses their limitations and strengths. Section 4 focuses on clinical applications of AI. Section 5 outlines current research challenges. Section 6 presents future research directions. Section 7 concludes the paper. Appendix provides the abbreviation list. The overall framework of this research is depicted in Fig. 4.

## 2. AI methodologies for PD detection

### 2.1. PD detection system

Fig. 5 illustrates a system framework for PD detection. Multimodal inputs (e.g., gait, speech, other non-motor data) are acquired and undergo modality-appropriate preprocessing followed by feature extraction. The learning block comprises model training, internal validation,

and classification. A performance gate checks predefined criteria; if unmet, the pipeline iterates on preprocessing/features/learning; if met, the finalized system outputs PD/HC (or a severity score). Evaluation is summarized below the pipeline: speaker- or site-independent validation (LOSO/external/cross-task) and reporting of performance with appropriate metrics.

To ensure that the terminology and abbreviations in this paper are clear and understandable, we first list some commonly used abbreviations and their corresponding descriptions. These abbreviations frequently appear in the following discussion, so the reader can refer to Table A.7.

### 2.2. Wearable sensor systems

Wearable devices, designed for direct attachment to the human body or integration into personal attire and accessories, represent a significant technological advancement in health monitoring (Sigcha et al., 2023). These portable instruments are embedded with an array of sensors and microprocessors that enable real-time monitoring and recording of various physiological metrics and behavioral patterns. The category of wearable devices is extensive, including smartwatches, wristbands, integrated garments, footwear, and biosensor patches (Mazilu et al., 2015; Demrozi et al., 2019; Ricci et al., 2019).

Recent technological advances in wearable sensor systems have achieved breakthroughs in the application of PD detection (Guo et al., 2019). These ambulatory monitoring technologies provide valuable objective motion data, enabling healthcare professionals to conduct more precise early assessments of patients' conditions (Tam et al., 2023). The motion characteristics captured by wearable sensors have played a crucial role in differentiating early-stage untreated PD patients from healthy individuals (Crowe et al., 2024). The continuous progress in wearable sensor technology has created new opportunities for monitoring PD symptoms in home and community settings, thereby enhancing the prospects for early intervention and continuous health monitoring (Schalkamp et al., 2023).

Sensors integrated into wearable devices can be classified based on their functionality:

- Inertial sensors, such as gyroscopes and accelerometers, are utilized to detect changes in motion and posture.
- Acoustic sensors, including microphones, are capable of identifying disruptions in speech signals.
- Optoelectronic biosensing systems utilize electromagnetic radiation within the visible spectrum to non-invasively quantify cardiopulmonary biomarkers, including pulse wave feature extraction and peripheral blood oxygen saturation monitoring.

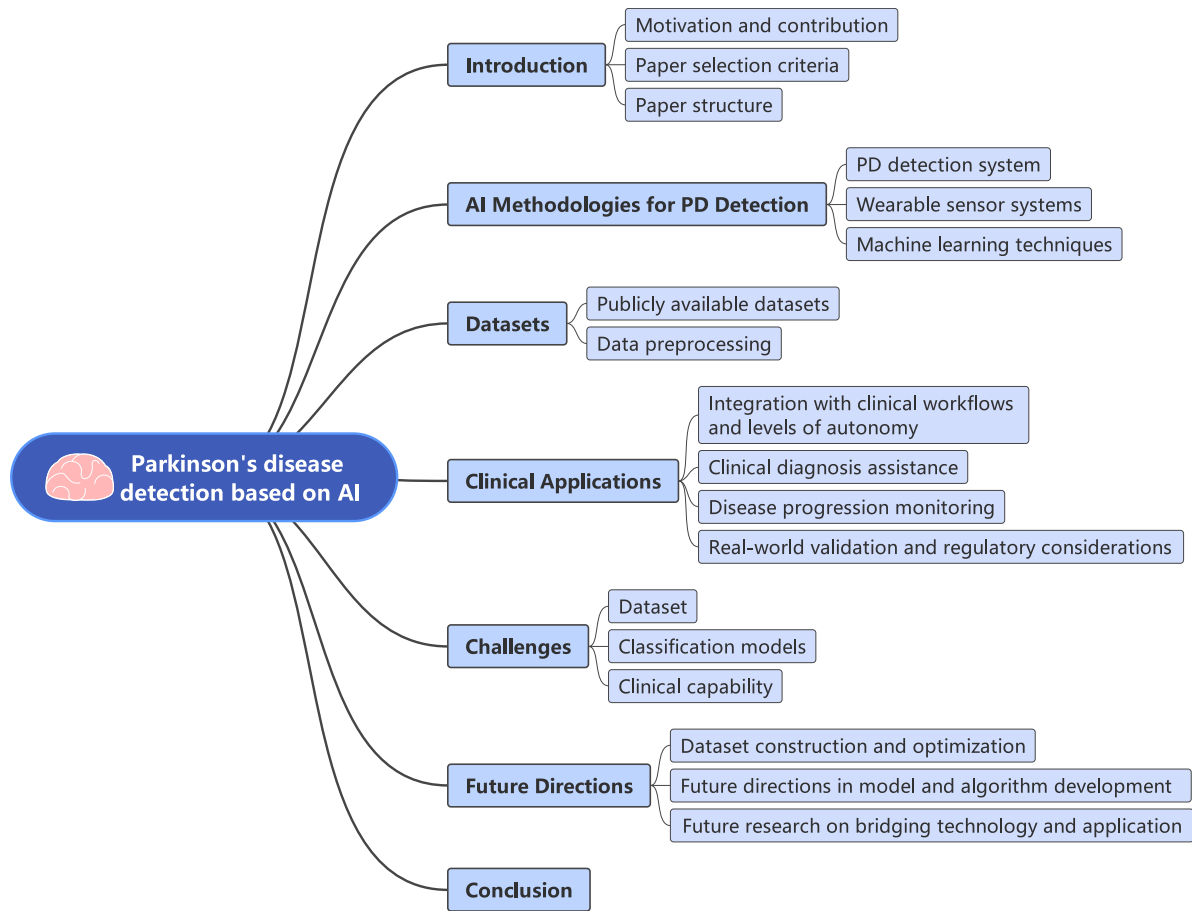


Fig. 4. Study directory.

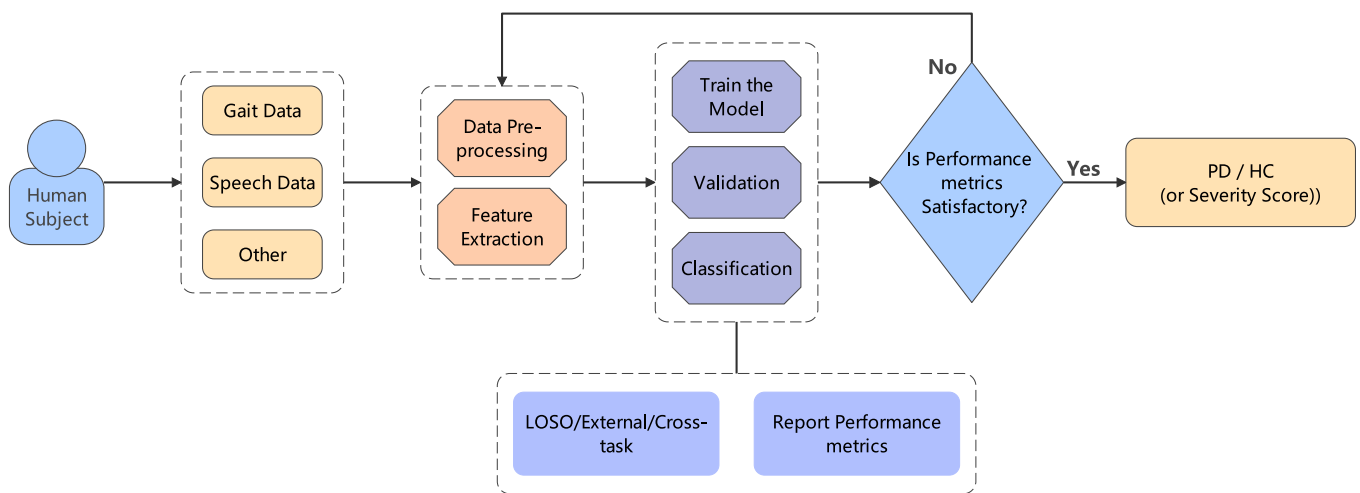


Fig. 5. PD detection system framework description.

- Electrophysiological monitoring technology integrates three diagnostic modules: analysis of myocardial electrical activity patterns (ECG), capture of cortical oscillatory signals (EEG), and acquisition of neuromuscular junction signals (EMG), which respectively correspond to functional assessments of the cardiovascular, central nervous, and motor systems.

### 2.3. Machine learning techniques

#### 2.3.1. Traditional models

- Decision Tree (DT): DT is an interpretable ML method for classification and regression tasks, constructing hierarchical decision rules through feature-based splits. In PD early detection, they analyze biomechanical features (e.g., gait velocity, step length

variability) to identify early signs of motor symptoms via a tree structure: internal nodes perform feature evaluations, branches denote decision outcomes, and leaf nodes classify the presence or absence of PD. While offering high model transparency and adaptability to heterogeneous sensor data, decision trees are prone to overfitting with excessive depth, necessitating optimization strategies like pruning and hyperparameter tuning to enhance generalizability in clinical applications (Ghane et al., 2022; Zhao et al., 2021).

- Random Forest (RF): RF is an ensemble learning method comprising multiple decision trees, where each tree is independently constructed with randomized feature subset selection during node splitting (Xue et al., 2022; Chen et al., 2022b). In PD early detection, RF effectively processes multimodal sensor data (e.g., accelerometers, gyroscopes, and vision sensors) and identifies discriminative biomechanical features such as gait rhythm, amplitude, and symmetry through embedded feature importance ranking. The model aggregates predictions via majority voting across constituent trees, enhancing classification accuracy and robustness against overfitting. For instance, RF demonstrates clinical efficacy in differentiating PD-specific gait abnormalities, including gait dragging and postural instability through multi-sensor fusion (Xu and Pan, 2020; Celik and Başaran, 2023).
- Support Vector Machine (SVM): The SVM is a powerful supervised learning classification algorithm extensively applied across multiple domains, including image recognition, text classification, financial forecasting, and biomedical diagnostics (Avolio and Fuduli, 2020; Zhang and Yang, 2024). The core principle of SVM is to identify an optimal hyperplane that separates data from distinct classes while maximizing the classification margin, known as the maximum-margin classification principle. This methodology emphasizes selecting a hyperplane that achieves correct classification while maximizing the margin between two-class samples, thereby enhancing the model's generalization capability (Chen et al., 2022a; Shrivastava et al., 2024). In early PD prediction, SVM has been widely used to process feature-extracted data. Through the appropriate selection of kernel functions (e.g., linear, polynomial, or Gaussian kernels) and optimized model parameters, SVM achieves high-precision classification of diverse activity types, such as gait and vocal patterns. For instance, in gait analysis, SVM is utilized to differentiate abnormal gait patterns in PD patients from normal gait in healthy populations. By extracting gait features (e.g., stride length, cadence, and gait symmetry), SVM learns decision boundaries to distinguish between gait classes during training and delivers accurate classification of new samples during testing phases (Vidya and Sasikumar, 2021; Alcaraz et al., 2022; Cristianini, 2000).
- K-Nearest Neighbor (KNN): KNN algorithm is a learning approach grounded in instances, which hinges on the principle that if most of a sample's K closest instances (i.e., nearest neighbors) are in a specific class, the sample is also categorized into that class. KNNs can be used for real-time activity recognition in the early detection of PD. By collecting accelerometer and gyroscope data from patients, feature vectors are constructed and then compared with labeled activity samples using distance metrics to identify the KNNs (Chen et al., 2013). Based on the activity labels of these neighbors, such as sitting, walking, or picking up objects, the majority voting method is applied to determine the current activity state of the patient. The advantage of KNNs lies in their simplicity, ease of implementation, and the fact that it does not require complex data preprocessing, making them suitable for scenarios that demand rapid response (Lv et al., 2024; Makarious et al., 2022).

### 2.3.2. DL architectures

- Convolutional Neural Networks (CNNs): CNNs belong to DL models adept at handling data with spatial correlations, such as images and videos. Due to their unique architecture, CNNs can automatically and effectively identify multi-level features from data, thereby demonstrating high accuracy in pattern recognition tasks. The traditional CNNs structure diagram is shown in Fig. 6. For the early prediction of PD patients, CNNs can process time-series features from sensor data (such as accelerometers, gyroscopes, electromyograms, etc.) to achieve efficient symptom recognition and disease prediction (Lin et al., 2022; Chen et al., 2023). Despite its promise in PD prediction, the performance of CNNs remains contingent upon the quality and quantity of training data. To enhance robustness, techniques such as data augmentation and regularization are typically integrated to mitigate overfitting and improve generalization capability. Collectively, CNNs exhibit broad potential for early PD prediction, especially for integrating multimodal sensor data and automatically learning features from high-dimensional datasets (Parisi et al., 2022; Celik and Başaran, 2023).
- Recurrent Neural Networks (RNNs): RNNs are a category of neural architectures optimized for handling sequential data, incorporating memory mechanisms to model temporal dependencies. Distinct from feedforward networks, RNNs possess cyclic connectivity that propagates information across time steps. This design preserves historical states to inform current predictions, rendering RNNs effective for temporal pattern recognition tasks, including speech processing, text analysis, gait cycle quantification, and kinematic trajectory modeling (Li et al., 2023; Singh et al., 2025). As shown in Fig. 7,  $A$  denotes the feedforward network layer,  $X_t$  represents the model input, and  $h_t$  is the output. RNNs demonstrate extensive applications in the early prediction of PD patients, particularly in processing continuous motor data such as gait cycles and hand movement trajectories, which often exhibit temporal characteristics. RNNs can analyze these time-series data, capture dynamic changes within the data, and identify abnormal patterns and regularities, such as the slowing of gait, asymmetry in stride length, or tremors in movement. These features may not be easily recognizable by traditional methods in the early stages, but RNNs can effectively detect these subtle changes (Riasi et al., 2024; Dar et al., 2022; Ahmed et al., 2022).
- Generative Adversarial Networks (GANs): GANs achieve a dual-network architecture composed of a generator (G) and a discriminator (D). The generator is trained to synthesize data approximating the statistical distribution of real samples, while the discriminator learns to discriminate between authentic and generated data through adversarial optimization. Through adversarial training of these two components, the generator gradually produces more realistic data, and the discriminator continuously improves its ability to discriminate (Pan et al., 2025; Ramesh and Bilal, 2022). The GANs structure diagram is shown in Fig. 8. GANs show notable effectiveness in the early detection of PD, particularly in cases where conventional diagnostic methods are restricted due to the subtle early signs. By synthesizing data that closely mimics the distribution of real-world patient data, GANs enhance pattern recognition to facilitate the identification of early pathophysiological features. This synthetic data generation capability can help researchers and clinicians better understand and identify the early features of PD (Shen et al., 2025a; Rey-Paredes et al., 2024).
- Graph Convolutional Networks: GCNs represent a category of DL models tailored for handling graph-based data, like social networks and molecular structures. Unlike traditional CNNs, which focus on regular grid-based data (e.g., images), GCNs directly model complex relationships between nodes, making them particularly suitable for analyzing non-Euclidean data (Mohanraj

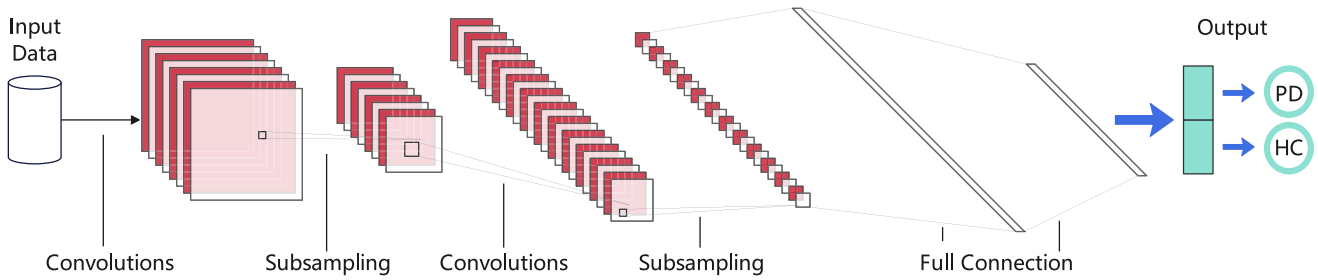


Fig. 6. Convolutional Neural Network.

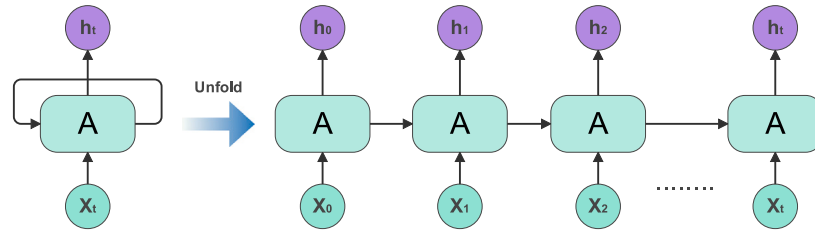


Fig. 7. Simple RNN structure.

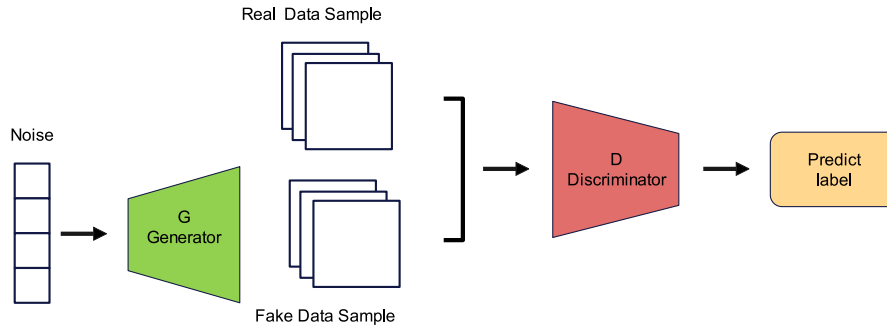


Fig. 8. Framework of Generative Adversarial Networks.

et al., 2024). The GCNs structure diagram is shown in Fig. 9. In the context of early prediction of PD, GCNs demonstrate significant potential. For instance, by transforming brain imaging data (e.g., functional magnetic resonance imaging, fMRI) into graph structures — where nodes represent brain regions and edges denote functional connectivity strengths — GCNs can automatically capture abnormal connectivity patterns associated with PD (Chang et al., 2023). This capability enables GCNs to identify early pathological features that are challenging to detect using conventional methods, such as reduced information transfer efficiency in specific neural networks or functional degradation of critical hub nodes. Furthermore, GCNs can integrate multimodal data (e.g., gene expression, clinical scale scores, and motion sensor data) to construct heterogeneous graphs across dimensions, thereby enabling comprehensive disease risk assessment. Zhao et al. (2022a), Zhang et al. (2023).

- Transformer: Transformers are a self-attention-based neural network architecture, originally designed for Natural Language Processing (NLP) tasks. Their core strength lies in capturing dependencies between different positions in sequential data via the self-attention mechanism, which effectively extracts long-range features (Tougui et al., 2024). In PD detection, Transformer architectures with attention mechanisms show unique advantages. Sun et al. (2023) proposed a Higher-order Polynomial Transformer (HP-Transformer) for fine-grained FoG detection in PD, combining pose and appearance feature sequences through higher-order self-attention mechanisms to achieve precise FoG

event identification with an AUC of 0.92. Chen et al. (2024) introduced CTFF-Net, a Deep Gray Matter (DGM) nuclei segmentation network for PD, based on a CNN-Transformer interleaved encoder that captures global and local features. By incorporating a Cascaded Channel-Spatial Fusion (CCSF) block and a Symmetric Boundary Attention (SymBA) module, it significantly enhances segmentation accuracy, reaching a Dice Similarity Coefficient (DSC) of 0.854. These studies show that the Transformer architecture is effective for processing complex sequence and image data, enabling the capture of global dependencies and local features, and providing new approaches for medical image analysis and disease diagnosis (Özdemir and Özyurt, 2025).

### 2.3.3. Performance evaluation metrics

Different studies have adopted diverse evaluation criteria to evaluate the performance of AI methods in PD prediction. Table 2 summarizes the commonly used performance indicators and their mathematical expressions.

### 2.3.4. Motor symptom

Relevant studies have applied wearable devices and AI approaches for the early detection of motor symptoms in PD. Details regarding samples, sensor technologies, methodologies, and findings are summarized in Table 3 and further elaborated in the following sections.

He et al. (2024a) presented a DL-based approach for analyzing smartphone-based walking records to detect PD in its early stages. The study utilized inertial sensors (gyroscopes and accelerometers) in

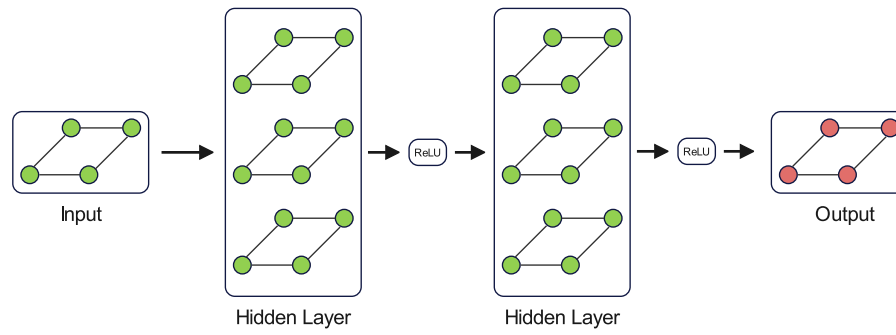


Fig. 9. Framework of Graph Convolutional Networks.

**Table 2**  
Performance metrics for assessing model performance.

Performance metrics	Definition
F1-score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Area Under the Curve	$\int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR})$
Specificity	$\frac{TN}{TN + FP}$
Kappa	$\frac{p_o - p_e}{1 - p_e}$
Precision	$\frac{TP}{TP + FP}$
Sensitivity (Recall)	$\frac{TP}{TP + FN}$
MAE	$\frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
$R^2$	$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
FNR	$\frac{FN}{FN + TP}$

smartphones to gather data from 119 PD patients and 467 healthy participants. Through data preprocessing (including normalization, scaling, and rotation) and the application of the deep NeuroEnhanceNet model, early detection of PD was achieved. The evaluation showed that the method achieved the best results on data from the “Rest” phase, with an average AUC of 0.883 and an FNR of 0.053. This indicates that smartphone-based walking records can serve as an effective digital biomarker for early PD detection.

Lin et al. (2022) presented a PD detection system based on a neural network model. The study enrolled 32 drug-naïve individuals with PD and 16 healthy controls matched for age and gender, measuring their motion data using IMU sensors. The neural network model was able to identify late-stage PD patients with an accuracy of 92.72% during the validation process. It also separated initial-stage PD patients from healthy elderly individuals with a 99.67% accuracy. An additional cohort of independent participants assessed these models, which effectively distinguished PD patients, including those with varying severity, from healthy elderly individuals.

Sarapata et al. (2023) developed a video-based activity recognition framework for the automatic motor assessment of PD patients. The framework leverages DL models and unmarked pose estimation techniques to record patients’ movements through video, enabling scalable and asynchronous evaluation of motor disorders in PD patients. The study utilized 7310 video clips documenting 1170 PD motor assessments, with data sourced from five independent clinical centers. The model achieved balanced accuracy of 96.51% in video-level activity classification and produced high-precision frame-level activity annotations. This framework can dynamically classify the motor assessment tasks implemented by PD patients and demonstrates good generalization ability across different patient populations. Additionally, the framework can automatically label extensive motor assessment data, thereby increasing statistical power in future PD studies.

Chen et al. (2023) proposed a wearable daily detection system optimized by explainable DL for gait detection in PD patients. The study aimed to develop a precise, objective, and non-intrusive detection algorithm, optimize daily gait detection for PD patients using an explainable DL architecture, and identify the most representative spatiotemporal motion features. It used five IMUs on the wrists, ankles, and waist to collect motion data from 100 subjects during a 10-meter walking test. This data, processed by CWT, trained the constructed 6-channel CNN classification model. The findings indicated that the waist-mounted sensor achieved a high accuracy of 98.01% and an area under the receiver operating characteristic curve (AUC) of  $0.9981 \pm 0.0017$ , based on ten-fold cross-validation.

Meng et al. (2023) suggested a gait assessment model based on IMU sensors for assessing the gait of early-stage PD patients. The study demonstrated that turning-related gait parameters exhibited significant discriminative capacity in distinguishing early-stage PD patients from healthy controls, particularly through biomechanical metrics of joint mobility and postural stability, with AUC values exceeding 0.7, demonstrating the diagnostic potential of this model in preclinical PD identification.

Zhang et al. (2020) presented a method for predicting FoG in PD patients by identifying impaired gait patterns. The study collected acceleration signals from accelerometers installed on the waists of 12 patients and extracted gait-based impaired gait features and traditional FoG detection features. Then, two FoG prediction frameworks were established using AdaBoost to verify whether impaired gait features could more accurately predict FoG. The results showed that the model using impaired gait features had higher prediction accuracy than the model using traditional FoG detection features, and a personalized labeling method further improved prediction accuracy. The final FoG prediction model attained an accuracy of 82.7% in patient-dependent testing and 77.9% in patient-independent testing, with an average delay time of 0.93 s.

Dong et al. (2023) proposed a Static–Dynamic Temporal Network (SDTN) model for PD prediction based on vertical ground reaction force (VGFR) signals. They combined time-series data and gait force transfer characteristics. Then, they used a 1D-ConvNet to extract temporal features and force transfer features using a 2D-ConvNet. The study demonstrated that the model recorded an accuracy of 96.7% in PD diagnosis and 92.3% in severity prediction, showing strong classification performance.

Perez-Ibarra et al. (2020) proposed an IMU-based framework for real-time detection of gait events (e.g., heel-strike (HS) and toe-off (TO)) in PD patients and healthy controls. The proposed methodology employed an adaptive unsupervised learning algorithm to extract key gait phases from kinematic signals. Experimental results demonstrated F1-scores  $\geq 0.95$  for all participants (healthy and PD cohorts), with absolute mean differences (AMD) in gait event timing ranging from 36 ms to 82 ms. The study highlighted the algorithm’s adaptability to individual gait patterns through real-time parameter adaptation, enabling enhanced detection accuracy across diverse physiological conditions.

**Table 3**  
Detection of motor symptoms using wearable sensors and AI methods.

Study (Year)	Target symptoms	Sensor type	Algorithm	#PD subjects (#Controls)	Performance metrics
He et al. (2024a)	Motor Symptoms	Accelerometer, Gyroscope	DNN	119 (467)	AUC: 0.883 FNR: 0.053
Lin et al. (2022)	Motor Symptoms	IMU	CNN	32 (16)	ACC: 99.67%
Sarapata et al. (2023)	Motor Symptoms	RGB Camera	ST-GCN	1170 PD motor assessments (7310 video clips)	ACC: 96.51%
Chen et al. (2023)	Bradykinesia, Rigidity, Gait Disorders	IMU	CNN, CWT	50 (50)	ACC: 98.01 $\pm$ 0.85% AUC: 0.9981 $\pm$ 0.0017
Meng et al. (2023)	Postural Instability, Gait Disorders	IMU	SVM	21 (19)	AUC $\geq$ 0.7
Zhang et al. (2020)	Freezing of Gait	Accelerometer	AdaBoost	12 (0)	ACC: 82.7%
Dong et al. (2023)	Bradykinesia, Posture Instability, Gait Disorders	Pressure Sensors	1D/2D-CNN	93 (73)	ACC: 96.7%
Perez-Ibarra et al. (2020)	Gait Event Detection	IMU	HMM	7 (5)	F1 $\geq$ 95%
Zhao et al. (2024)	Bradykinesia, Posture Instability, Gait Disorders	Smartphone Camera	ResNet50, RF	30 (42)	ACC: 91.67%
El Maachi et al. (2020)	Gait Disorders	Foot Pressure Sensors	1D-CNN	93 (73)	ACC: 98.7%
Özdemir and Özyurt (2025)	Bradykinesia, Resting Tremor	Hand drawings	Vision Trans- formers(ViT), ElasticNet, DT, Linear Discriminant, SVM, KNN	1632 (1632)	ACC:99.9% PR:99.69% Recall:100% F1:99.84%
Rangel-Cascajosa et al. (2025)	Bradykinesia, Gait Disorders	Foot Pressure Sensors	LSTM, GRU	92 (72)	ACC:93.75%
Sigcha et al. (2022)	Freezing of Gait	Accelerometer	FOG-Transformer	21	AUC:0.957
Borzi et al. (2023)	Freezing of Gait	Accelerometer	Multi-head CNN	118(21)	SE:87.7% SP:83.3% F1:83% AUC:0.946
Zhao et al. (2018)	Gait Disorders	Foot Pressure Sensors	LSTM, CNN	93(73)	ACC:98.61%
Huan et al. (2025)	Gait Disorders	Foot Pressure Sensors	Conv1D- Transformer- GRU	93(73)	ACC:98.61%
Naimi et al. (2024)	Gait Disorders	Foot Pressure Sensors	Conv1D, Transformer	93(73)	ACC:88%
Veeraragavan et al. (2020)	Gait Disorders	Foot Pressure Sensors	ANN	93(73)	ACC:97.4%
Navita et al. (2025)	Gait Disorders	Foot Pressure Sensors	RFT	93(73)	ACC:97.5 $\pm$ 2.1% SP:95 $\pm$ 2.2% SE:97 $\pm$ 2.5%
Alharthi et al. (2020)	Gait Disorders	Foot Pressure Sensors	Parallel 2D-DCNN	93(73)	ACC:95.5 $\pm$ 0.28%
Guo et al. (2022)	Freezing of Gait	IMU	LSTM-PM	12	ACC:93.6 $\pm$ 1.8% SE:88.5 $\pm$ 10.1% F1:89.4 $\pm$ 7.7%

(continued on next page)

Zhao et al. (2024) presented a novel diagnostic method for PD that utilizes smartphone videos to capture patients' gait, extracts 3D joint data through MediaPipe, and converts it into 2D images for

analysis. The study demonstrated improved efficiency and accuracy of the diagnostic system by simplifying the data collection process and reducing the number of parameters. By combining ResNet50 and the

Table 3 (continued).

Study (Year)	Target symptoms	Sensor type	Algorithm	#PD subjects (#Controls)	Performance metrics
Veer et al. (2022)	Gait Disorders	Foot Pressure Sensors	NB,NN,SVM	25(14)	NB,NN,SVM ACC:100%,92.3%, 88.88%
Torghabeh et al. (2024)	Gait Disorders	Foot Pressure Sensors	Uni-LSTM	93(73)	ACC:99.19%
Ghaderyan and Fathi (2021)	Gait Disorders	Foot Pressure Sensors	NNLS	93(73)	ACC:97.22%
Wang et al. (2024)	Freezing of Gait	IMU	PhysioGPN	93(73)	ACC:85.2 $\pm$ 6.5% AUC:85.8 $\pm$ 4.1% SE:91.6 $\pm$ 7.3% SP:80.1 $\pm$ 4.4% F1:85.5 $\pm$ 4.8%
Vidya and Sasikumar (2021)	Gait Disorders	Foot Pressure Sensors	MCSVM	93(73)	ACC:98.65% SE:97.6% SP:99.1%
Zhao et al. (2022b)	Gait Disorders	Foot Pressure Sensors	EnKNN	91(71)	ACC:95.02 $\pm$ 0.44%
Sun et al. (2023)	Freezing of Gait	RGB Cameras	HP-Transformer	45	ACC:86.1% SE:86.6% SP:84.7% AUC:0.92
Abdulhay et al. (2018)	Gait Disorders, Tremor	Foot Pressure Sensors	SVM	93(73), 16	ACC:92.7%

RF algorithm, the system can effectively identify gait abnormalities and accurately differentiate PD patients from healthy controls. The results showed that when classifying with limb joint data, accuracy can reach 91.67%. Compared with traditional video processing methods, the number of parameters is reduced by 22% and the testing time is reduced by 62.92%. This method provided a simple and contactless solution for the early detection of PD and has significant clinical application potential.

El Maachi et al. (2020) proposed a PD detection method based on a deep 1D-Convnet. The model takes 18 vertical VGRF signals as input, with each signal coming from 8 sensors placed under each foot. Through parallel processing of multi-channel signals, the proposed architecture extracts discriminative features from gait-related kinematic data and executes classification through a fully connected neural network. Experimental results demonstrated a detection accuracy of 98.7% for Parkinson's disease identification and 85.3% precision in symptom severity staging, quantitatively validating the methodological novelty and diagnostic efficacy in pathological gait analysis. In addition, the study also verified its superior performance through cross-validation and comparison with other existing methods, indicating a high potential for clinical application.

Özdemir and Özyurt (2025) proposed an ElasticNet-optimized Vision Transformer (ViT) framework for the early detection of PD through automated analysis of hand-drawn spirals and waveforms. The study utilized a publicly available Kaggle dataset comprising digitized hand-drawn images (scanned or digitally rendered) from 1632 PD patients and 1632 healthy controls. Following preprocessing steps — including image resizing, intensity normalization, and ElasticNet-based feature selection — the ViT architecture extracted hierarchical spatial features (global shape patterns and local stroke irregularities) from static drawings. These features were subsequently classified using ML algorithms (DT, Linear Discriminant, SVM, and KNNs). The results demonstrated that the ViT\_Base\_16 + SVM configuration achieved optimal performance, with 99.9% accuracy, 99.69% precision, 100% recall, and an F1-score of 99.84%, while reducing computational training time by 427 $\times$  compared to other ViT variants (e.g., ViT\_Large\_16/32). This underscores the potential of ViT-driven hand-drawing analysis combined with ElasticNet regularization as a rapid, cost-effective, and highly

accurate screening tool for early PD diagnosis in clinical settings, particularly in resource-constrained environments.

Rangel-Casajosa et al. (2025) proposed an early PD detection method using insole pressure sensors and gated recurrent neural networks (RNNs). Gait data were captured by insole-embedded sensors, and RNN variants based on LSTM and GRU were compared for classifying PD patients versus healthy controls. A single-layer GRU achieved the best trade-off between accuracy (up to 93.75%) and computational efficiency. The lightweight model is suitable for deployment on resource-constrained wearables, enabling early screening and continuous monitoring.

Sigcha et al. (2022) introduced a freezing of gait (FoG) detection approach using a single waist-worn triaxial accelerometer and an FoG-Transformer network. The study used data collected in patients' homes to improve detection accuracy for long-term monitoring and real-time applications. By combining CNN and Transformer components, FOG-Transformer exploited temporal dependencies across adjacent windows, improving AUC by 4.1% over baselines and achieving strong sensitivity and specificity. Despite slightly increased computational cost, the performance gain and architectural simplification support deployment on embedded and mobile devices, including real-time cueing systems.

Borzi et al. (2023) developed a real-time FoG detection algorithm using a single inertial sensor (waist accelerometer) and a multi-head CNN. Across three datasets (118 PD patients and 21 healthy older adults), the architecture analyzed inertial data at multiple spatial resolutions to capture local and global FoG characteristics. On REMPARK, sensitivity reached 0.877 and specificity 0.883 with an AUC of 0.946; on an independent dataset (6MWT), AUC was 0.953, evidencing generalization. Low computational cost and a short per-window test time (43 ms) support real-time deployment, enabling immediate feedback and assistive interventions.

Zhao et al. (2018) proposed a dual-branch deep model combining LSTM and CNN for automatic PD detection and severity assessment from gait. The hybrid architecture captured spatiotemporal patterns in gait data and outperformed traditional methods. Trained and tested on three public VGRF datasets, the study primarily addressed four-level severity classification and laid groundwork for integrating multi-source data to enable finer-grained severity prediction.

Huan et al. (2025) introduced a parallel hybrid architecture combining 1D convolutions (Conv1D), an efficient Transformer, and bidirectional GRUs for PD detection and severity assessment from gait. Conv1D layers extracted local spatial features, the Transformer captured contextual dependencies, and BiGRU modeled temporal patterns, jointly representing complex gait characteristics. The method achieved 95.7% accuracy for PD detection and 87.3% for severity assessment.

Naimi et al. (2024) proposed a deep hybrid model combining Conv1D and a Transformer to detect PD and assess severity from gait. Conv1D extracted local spatial features, while the Transformer modeled long-range spatiotemporal dependencies. Reported accuracies were 88% for PD detection and 87.89% for severity assessment.

Veeraragavan et al. (2020) described a non-invasive approach for early PD diagnosis and severity assessment using VGRF signals. An artificial neural network extracted gait features, achieving 97.4% accuracy for PD detection and 87.1% for severity assessment. With a simple architecture and SMOTE-based handling of class imbalance, the method demonstrated effectiveness for remote monitoring and wearable applications.

Navita et al. (2025) developed a two-stage model for PD detection and severity estimation from gait signals collected via VGRF sensors, addressing gradual symptom progression and early detection challenges. Stage 1 used a hyperparameter-optimized random forest tree (RFT) to classify PD vs. non-PD with  $97.5\% \pm 2.1\%$  accuracy. Stage 2 applied a hyperparameter-optimized ensemble regressor (ER) to estimate severity, achieving  $96.4\% \pm 2.3\%$  average accuracy. SMOTE balanced the dataset, and recursive feature elimination selected key gait features (e.g., double-support time). The framework offers strong early detection and severity estimation performance, though explainability and severity granularity warrant further improvement.

Alharthi et al. (2020) proposed a deep convolutional neural network (DCNN) for PD detection and severity grading from spatiotemporal gait signals. End-to-end learning with automated feature extraction enabled classification of PD versus healthy participants and severity estimation. Layer-wise relevance propagation elucidated salient gait features driving predictions. The model achieved  $95.5\% \pm 0.28\%$  classification accuracy and a 98% F1-score for severity grading, and demonstrated robustness to noise across large longitudinal GRF datasets.

Guo et al. (2022) presented a high-accuracy wearable FoG detection method using pseudo-multimodal features. An LSTM-based proxy measurement model (LSTM-PM) estimated EEG features from easily acquired accelerometer signals to produce pseudo-EEG (pmEEG), which was fused with raw accelerometry (pmEEG-ACC). An SVM classifier then detected FoG. Without collecting accurate EEG signals, the approach retained multimodal advantages and improved wearability for long-term home monitoring.

Veer et al. (2022) examined sex-specific differences in gait rhythmicity under single- and dual-task conditions and proposed an early PD detection approach combining statistical and kinematic features. Analysis of gait parameters (e.g., stance time, swing time, stride time, and variability) revealed stronger dual-task effects in female PD. Naïve Bayes, SVM, and neural network classifiers distinguished PD from healthy controls, with Naïve Bayes achieving 100% accuracy in single-task settings, supporting sex-specific gait markers for early diagnosis.

Torghabeh et al. (2024) proposed a lightweight automated model based on cumulative gait signals and time–frequency–fuzzy features. Using cumulative VGRF (CVGRF) as the core signal, four features — instantaneous frequency (IF), spectral entropy (SE), fuzzy recurrence image entropy (FRIE), and fuzzy recurrence entropy (FRE) — were extracted. Bayesian optimization tuned key hyperparameters (initial learning rate, hidden units) for Uni-LSTM and Bi-LSTM architectures. The system achieved 99.19% accuracy for PD detection (Uni-LSTM) and 92.28% for 5-class UPDRS severity grading (Bi-LSTM), with only 51 KB memory, supporting microcontroller deployment for real-time monitoring. The “single-signal + few features + no preprocessing”

design reduces device and computational costs while maintaining high accuracy.

Ghaderyan and Fathi (2021) developed a low-cost, noninvasive PD detection and severity grading method using VGRF signals, time-varying singular value decomposition (TSVD), and inter-limb time-varying singular values (ITSV). In-shoe VGRF sensors recorded bilateral gait; TSVD (via Hankel or non-overlapping trajectory matrices) extracted intrinsic time-varying characteristics. The difference between left and right time-varying singular values yielded ITSV, which, with a sparse nonnegative least squares classifier, enabled PD classification and severity grading.

Wang et al. (2024) proposed PhysioGPN, a physiology-informed deep learning framework with knowledge distillation (KD) for lightweight FoG prediction from accelerometer (ACC) and gyroscope (GYRO) signals. Four physiologically motivated design choices — large kernels for gradual motion changes, multi-dimensional multi-scale convolutions for coordination, a dual-tower structure for gait self-similarity and asymmetry, and multi-domain attention for cross-domain fusion — enabled prediction 2–2.5 s before FoG onset. KD transferred knowledge from a teacher trained on all sensors (waist, left wrist, both legs) to students using fewer sensors (e.g., only legs or only waist), reducing sensor burden while limiting performance loss.

Vidya and Sasikumar (2021) presented a PD diagnosis and severity grading framework using VGRF gait signals and a multi-class SVM (MCSVM). Twenty-two spatiotemporal gait features (e.g., step time, stance time, stride length, cadence) were extracted; Pearson correlation selected 15 biomarkers, and multivariate regression normalized data to mitigate physiological confounders (e.g., height, weight). A one-vs-one strategy decomposed multi-class tasks into binary subproblems, comparing linear, quadratic, Gaussian, and cubic SVM kernels.

Zhao et al. (2022b) proposed an ensemble K-nearest neighbors (EnKNN) diagnostic method and identified gait markers of PD progression from VGRF features. Using PhysioBank gait data, standardized VGRF (VGRF/BW), coefficient of variation (CV), and asymmetry index (AI) were extracted. Bootstrap sampling generated base KNN classifiers; G-mean and F-measure jointly determined classifier weights, and weighted voting yielded the ensemble. The method addressed class imbalance without over-/undersampling, avoiding information loss and overfitting.

Sun et al. (2023) proposed a fine-grained FoG detection method from visual inputs using a High-order Polynomial Transformer (HP-Transformer). Clinical FoG assessment videos were converted into skeleton pose sequences (joint coordinates) and appearance feature sequences (RGB texture). First-, second-, and third-order self-attention mechanisms formed multi-stream Transformers capturing fine-grained FoG patterns across dimensions. A cross-order self-attention fusion aggregated multi-stream features for end-to-end FoG detection.

Abdulhay et al. (2018) proposed a machine-learning framework that diagnoses Parkinson's disease (PD) from VGRF gait analysis and assesses disease severity from tremor signals. The system integrates time- and frequency-domain gait features with frequency-domain tremor features to achieve high diagnostic accuracy and robust severity estimation. Using PhysioNet public datasets, the gait cohort comprised 93 PD patients (279 records) and 73 healthy controls (HC) recorded via 16 force sensors (8 per foot). In contrast, the tremor cohort included index-finger tremor velocity signals from 16 low-medication PD patients. Preprocessing employed Chebyshev Type II high-pass and Butterworth filters to suppress noise and mitigate body-weight effects. Gait features included stance, swing, and stride time, as well as the foot-strike profile (time domain), and fast Fourier transform (FFT) and power spectral density (PSD) descriptors (frequency domain). Tremor features comprised FFT- and PSD-based measures and median frequency. Support vector machines (linear and radial basis function kernels) and decision trees were trained to evaluate PD diagnostic performance, and tremor-derived features were analyzed to quantify their association with disease severity.

Studies leveraging IMU, insole pressure/VGRF, and smartphone video generally frame gait abnormalities as PD vs. HC or multi-level severity classification (Lin et al., 2022; Chen et al., 2023; Meng et al., 2023; El Maachi et al., 2020; Vidya and Sasikumar, 2021; Zhao et al., 2024). On controlled or single-center datasets, both classical pipelines (handcrafted features + SVM/KNN) and deep Conv/Transformer hybrids report high accuracy/AUC; however, performance can be inflated by subject-dependent splits, protocol homogeneity, and device consistency. Robust designs pair subject-independent (LOSO) splits with external-site tests, quantify calibration, and explore cross-placement robustness (waist vs. ankle vs. insole). For clinical utility, mapping scores to UPDRS thresholds/MCID and reporting confidence intervals are as important as raw accuracy. Lightweight, explainable models remain attractive for routine screening, while deeper hybrids add value for severity grading when leakage is controlled. Beyond gait impairment, freezing of gait (FoG) presents a more specialized detection challenge due to its rarer, context-dependent, and short-lived events, which require temporal context modeling and event-level evaluation (Sigcha et al., 2022; Borzi et al., 2023; Zhang et al., 2020; Wang et al., 2024; Sun et al., 2023). CNN-Transformer architectures that aggregate adjacent windows improve detection AUC over baselines (Sigcha et al., 2022), and multi-resolution CNNs enable real-time deployment with low latency (Borzi et al., 2023). Prediction, rather than detection, demands reporting lead time and false alarms per hour; impaired-gait feature engineering and personalized labeling can raise predictive accuracy (Zhang et al., 2020). Knowledge-distilled frameworks (Wang et al., 2024) help reconcile sensor burden and accuracy. Stronger FoG studies enforce patient-independent splits, include home/clinic generalization, and disclose time-to-detection, latency, and energy. Overall, FoG pipelines should optimize the trade-off between early warning (longer horizon) and false alarms, target cross-device robustness, and document deployment metrics for real-world cueing systems.

Taken together, motor symptom pipelines exhibit a predictable tension between model capacity and interpretability. Feature-based or statistical models that operate on clinically meaningful spatiotemporal parameters (e.g., step, stance, and stride time; asymmetry indices) with SVM, KNN, or Naïve Bayes provide transparent biomarkers and strong performance under controlled conditions (Vidya and Sasikumar, 2021; Zhao et al., 2022b). However, these approaches often degrade when data comes from different devices and sensor placements or from multiple sites. By contrast, higher-capacity convolutional and Transformer hybrids and physiology-informed networks generally yield superior robustness and AUC on heterogeneous or real-world data (Zhao et al., 2018; Alharthi et al., 2020; Borzi et al., 2023; Sigcha et al., 2022), albeit with reduced transparency. Two pragmatic strategies have emerged to reconcile this trade off: post hoc attribution to expose salient gait windows or features (e.g., relevance or saliency analyses in interpretable CNNs) (Chen et al., 2023); and knowledge distillation that transfers a high capacity teacher (often trained with multiple sensors and axes) to compact students suitable for wearable deployment while retaining most of the teacher's accuracy (Wang et al., 2024). For clinical use, high capacity models should be accompanied by calibrated probabilities and operating points referenced to UPDRS or MCID, and FoG applications should additionally report event level latency and false alarm rates.

### 2.3.5. Non-motor symptom

Non-motor symptom studies remain relatively understudied and challenging due to heterogeneous, fluctuating phenotypes and low-amplitude signals that often require longer or longitudinal capture; ecologically valid acquisition (e.g., passive phone calls) introduces device, channel, and language variability and privacy constraints that depress generalization and limit data sharing (Chang et al., 2023; Laganas et al., 2021; La Quatra et al., 2024). Cohorts, especially EEG and facial-affect, are frequently small and fragmented, and supervision commonly relies on coarse PD vs. HC labels rather than symptom-specific

severities aligned with clinical scales, weakening clinical interpretability (Huang et al., 2023b; Dao et al., 2025). Confounders such as age, sex, accent, microphone, and non-PD voice pathologies are hard to control in the wild (Pah et al., 2023; La Quatra et al., 2025). Limited use of XAI and calibration further slows clinical adoption (Shen et al., 2025b).

Some studies have used wearable devices and AI methods to detect non-motor symptoms of PD early. Table 4 lists the samples, sensor technologies, methods, and results. The following sections provide further details.

Solana-Lavalle et al. (2020) proposed a PD early detection method based on a small number of voice features. The study used the latest largest public dataset comprised of voice signals from 188 PD patients and 64 healthy controls. By employing the Wrappers feature selection method, the number of features was reduced from 754 to between 8 and 20. Four classifiers (KNN, multilayer perceptrons, SVM, and RF) were employed to voice-based PD detection. The method achieved 94.7%, 98.4%, 92.68%, and 97.22% in terms of accuracy, sensitivity, specificity, and precision, respectively.

Quan et al. (2022) proposed an end-to-end DL method for detecting PD from speech signals. The method uses a time-distributed 2D-Convnet to extract temporal dynamic features and then employs a 1D-Convnet to capture the dependencies between these time series. The study was validated using two databases, with results indicating that the model achieved an accuracy of 81.6% for the sustained vowel /a/ speech task and 75.3% for the short sentence reading task. In Spanish reading tasks involving both simple and complex sentences, an accuracy rate of 92% was achieved.

Karan et al. (2020) presented a PD prediction method based on intrinsic mode function (IMF) features. The method uses empirical mode decomposition (EMD) to decompose speech signals, extracting intrinsic mode function cepstral coefficients (IMFCC) to characterize the dynamic features of the speech signals. The study used two datasets from different experimental environments for PD classification prediction. The evaluation outcomes showed that the IMFCC features performed excellently on both datasets, significantly outperforming traditional acoustic features and MFCC features, with an increase in classification accuracy of 10%–20%. This method can effectively capture the non-linear characteristics of speech signals and handle dynamic changes, offering an innovative and efficient solution for early PD detection.

Karaman et al. (2021) presented an automatic detection method for PD based on speech signals and transfer learning. The study utilized a deep CNN to train and evaluate data collected from the mPower speech database, aiming to identify PD through biomarkers in speech signals automatically. The study employed three architectures, SqueezeNet1.1, DenseNet161, and ResNet101, and fine-tuned them using transfer learning. The evaluation showed that the DenseNet-161 architecture achieved 89.75% accuracy, 91.50% sensitivity, and 88.40% precision on the test set. This model's success indicates that combining the developed model with smart electronic devices can provide an alternative pre-diagnostic method for PD detection and assist doctors in clinical assessments.

Parisi et al. (2018) presented a hybrid ML approach based on feature selection to enhance PD early diagnosis. The study combined a multilayer perceptron (MLP) for feature selection and a Lagrange SVM (LSVM) for the classification. Using the MLP, the initial 27 features were reduced to the 20 most relevant, thereby enhancing classification accuracy. Experimental results demonstrated that the hybrid MLP-LSVM achieved 100% classification accuracy and, compared with traditional algorithms, improved classification performance while reducing training time and the number of iterations. Compared with existing commercial software and models reported in the literature, this method achieved significant improvements in accuracy, sensitivity, specificity, and AUC, demonstrating its potential for clinical application.

**Table 4**

Detection of non-motor symptoms using wearable sensors and AI methods.

Study (Year)	Target symptoms	Sensor type	Algorithm	#PD subjects (#Controls)	Performance metrics
Solana-Lavalle et al. (2020)	Dysphonia	Microphone	KNN, MLP, SVM, RF	188(64)	ACC: 94.7%
Quan et al. (2022)	Dysphonia	Microphone	2D-CNN, 1D-CNN	1.30(15) 2.50(50)	1.ACC:81.6% 2.ACC:92.0%
Karan et al. (2020)	Dysphonia	Microphone	SVM, RF	1.25(20) 2.25(20)	1.ACC:100% 2.ACC:96%
Karaman et al. (2021)	Dysphonia	Microphone	CNN	mPower Dataset	ACC:89.75% SE:91.50% PR:88.40%
Parisi et al. (2018)	Dysphonia, Dysarthria	Microphone	MLP, LSVM	48(20)	ACC:100% AUC:1.00 SE:100% SP:100%
Huang et al. (2023b)	Emotional Expression Disorder	Camera	StarGAN, CNN, Swin Transformer	(PDFE, CK+, RaFD, TFED, Oulu-CASIA) Dataset	ACC:100%
Chang et al. (2023)	Cognitive Impairment	EEG	ASGCNN	24(24)	ACC:87.67% Recall:90.36% Pr:88.43% F1:88.41% kappa:75.24%
Dao et al. (2025)	Dysarthria	Microphone	wav2vec2.0, Whisper, SeamlessM4T, SVM, CNN	148(99)	AUC:91.35%
Laganas et al. (2021)	Dysphonia	Microphone	SVM, LR, RF	145(416)	AUC: English/Greek/German 0.84,0.93,0.83
Shen et al. (2025b)	Dysarthria	Microphone	MLP-CNN-RNN-MKL	40(41)	ACC:91.11% Recall:92.50% PR:89.84% F1:91.13% AUC: 0.9125
Iyer et al. (2023)	Dysarthria	Microphone	Inception V3 CNN	40(41)	AUC: 0.97
Pah et al. (2023)	Dysarthria	Microphone	SVM	50(50)	ACC:77.46% Recall:94.00% F1:74.60%
La Quatra et al. (2024)	Dysarthria	Microphone	Transformer	PC-GITA	ACC:88.33% F1:88.13% AUC 88.23% Pr:85.00% SP:91.6%
Rey-Paredes et al. (2024)	Dysarthria	Microphone	ResNet, LSTM-FCN, InceptionTime, CDIL-CNN	50(50)	ACC:73.00%
Narendra et al. (2021)	Dysarthria	Microphone	SVM, CNN-MLP	PC-GITA	ACC:68.56%
He et al. (2024b)	Dysphonia	Microphone	SVM, GB, LR	MDVR-KCL, mPower	ACC:98.40% F1:97.90% AUC:0.996
Pah et al. (2024)	Dysarthria	Microphone	AVQI	PC-GITA	ACC:75.33% AUC:0.74
La Quatra et al. (2025)	Dysarthria	Microphone	Bilingual Dual-Head DM	PC-GITA, EWA-DB	ACC:90.00% F1:90.00%

Huang et al. (2023b) suggested a PD automatic diagnostic method based on mixed emotional facial expressions, combining Generative Adversarial Networks (StarGAN), CNN, and Swin Transformer to identify facial expression abnormalities in PD patients. The method initially used StarGAN to generate normal facial expression images from PD patients to enhance the dataset's diversity. Subsequently, FaceQnet was employed for image quality assessment to filter out high-quality synthetic images, which were then combined with the original PD

facial expression data and healthy control data to train a deep feature extractor and an expression classifier. Ultimately, the method used fused features of six basic emotions for PD diagnosis, achieving 100% diagnostic accuracy on the PDFE dataset and 70.08% accuracy on the facial expression recognition task. The study demonstrated that this method could effectively detect non-motor symptoms in PD patients and showed its potential for clinical early screening.

Chang et al. (2023) proposed an attention-based sparse graph convolutional neural network (ASGCNN) for PD recognition based on EEG data. The study utilized the public PD auditory oddball dataset, which included 24 PD patients (in both “on” and “off” medication states) and 24 matched healthy controls. The experimental results demonstrated good performance of the method in PD recognition, with an accuracy of 87.67%, recall of 90.36%, precision of 88.43%, F1-score of 88.41%, and Kappa value of 75.24%.

Dao et al. (2025) proposed an early Parkinson’s disease (PD) detection method based on three speech foundation models — wav2vec 2.0, Whisper, and SeamlessM4T — using alterations in speech (e.g., reduced prosody, articulatory imprecision) as key indicators. Leveraging the ICEBERG public datasets, speech was recorded with professional equipment, denoised, and segmented into 5 s clips. Two base strategies were examined: (i) feature extraction from the frozen models followed by SVM classification, and (ii) fine-tuning after freezing the optimal layers. An ensemble that averaged the posterior probabilities of the three models further improved performance relative to a 5-layer 1D-CNN baseline, without complex class rebalancing. The final ensemble of three fine-tuned models achieved an AUC of 91.35% and showed significant correlation with MDS-UPDRS clinical scores and DaTSCAN neuroimaging, supporting a low-cost, reliable approach to early PD screening.

Laganas et al. (2021) presented a PD detection method from passively captured continuous telephone speech on smartphones, enabling ecologically valid and privacy-preserving daily screening. Using speech impairments in PD (e.g., reduced intensity, abnormal vocal fold vibration) as core markers, the iPrognosis app collected multilingual (English, Greek, German, Portuguese) call audio. Local MFCCs, BBE, and demographic features were extracted, and both single-instance (Models A/B) and multi-instance (Model C) learners were designed. Language-aware training and leave-one-subject-out cross-validation addressed class imbalance and inter-call variability. The approach requires no active participation and preserves privacy, providing a low-cost, scalable solution for remote early PD screening and longitudinal monitoring.

Shen et al. (2025b) proposed an explainable AI (XAI) framework for speech-based early PD detection that is non-invasive and low-cost. Core acoustic features (MFCCs, jitter, shimmer, HNR) were extracted via Parselmouth, and a hybrid MLP+CNN+RNN+MKL model was built: CNN captured local spectrotemporal patterns, RNN modeled temporal dynamics, MKL integrated multimodal features, and MLP handled nonlinearity. SHAP was employed to quantify feature contributions and mitigate the “black-box” issue. To support disease monitoring, a 0–1 probability scoring system was designed to reflect PD likelihood by score intervals. Despite a small proof-of-concept dataset, results demonstrated that purely speech-based biomarkers can effectively diagnose early PD with real-time, low-cost advantages. Future work may exploit synthetic or multimodal data to enhance generalization, enabling remote screening and individualized monitoring.

Iyer et al. (2023) validated PD detection from low-resolution telephone recordings via transfer learning with CNNs to establish clinical utility for remotely collected speech. Twenty-three acoustic and four spectral feature sets were extracted for traditional classifiers (logistic regression and random forest). In addition, a pretrained Inception V3 CNN analyzed spectrograms via transfer learning, confirming the clinical reliability of low-resolution phone audio and demonstrating that Inception V3 effectively exploits time–frequency energy distributions, offering a low-cost, practical tool for remote screening.

Pah et al. (2023) developed an SVM-based speech-feature model to identify PD, focusing on hypokinetic dysarthria. Eighteen features related to glottal control, pulmonary control, and vocal-tract control were extracted. Three SVM experimental settings were investigated to distinguish PD from other non-PD voice pathologies. Results indicate that an SVM trained only on PD and healthy controls can reliably discriminate PD vs. HC; incorporating additional pathological voice

data further improves performance, whereas multi-class SVMs are less effective at separating non-PD categories.

La Quatra et al. (2024) investigated PD detection from speech captured under real-world operating conditions using foundation models and speech enhancement. Models fine-tuned on the standard PC-GITA dataset outperformed prior approaches in ideal conditions but degraded substantially on the in-the-wild e-PC-GITA dataset. Applying off-the-shelf enhancement (voice activity detection, dereverberation, denoising) markedly improved performance on e-PC-GITA. An ensemble of the two best foundation models achieved the best results on the enhanced e-PC-GITA data.

Rey-Paredes et al. (2024) introduced an approach combining GAN-driven data augmentation with temporal classifiers, focusing for the first time on direct classification from raw speech waveforms to address the underuse of raw temporal data in PD detection. BigVSAN (GAN) operated on mel spectrograms of real speech: the generator used periodic activations and anti-aliasing modules to produce raw waveforms highly similar to real speech, with randomized factors balancing fidelity and diversity. The combined real-plus-synthetic data were then trained four temporal models — ResNet, LSTM-FCN, InceptionTime, and CDIL-CNN — under speaker-independent 5-fold cross-validation to ensure generalization.

Narendra et al. (2021) explored using glottal source information for PD detection and compared a traditional pipeline with an end-to-end system. The pipeline extracted baseline and glottal features and used SVMs to assess their complementarity. The end-to-end system employed a CNN+MLP network fed by raw speech or glottal waveforms, segmented into 250 ms frames to avoid hand-crafted feature engineering. Both systems used 10-fold cross-validation to support generalization.

He et al. (2024b) proposed a smartphone speech-based digital biomarker approach for PD diagnosis combining preprocessing, feature extraction and selection, and traditional ML classification. Raw recordings from two smartphone datasets were denoised using MATLAB Voicebox; MDVR-KCL data were segmented into 10s clips, and all data were resampled to 44.1 kHz. The openSMILE toolkit extracted 446-dimensional acoustic features. Eight feature selection methods identified optimal subsets to reduce redundancy and overfitting. SVM (RBF kernel), gradient boosting, and logistic regression were tuned via 5-fold cross-validation with grid search to classify PD vs. healthy controls (HC), and cross-task evaluation (train on one task, test on another) assessed generalization.

Pah et al. (2024) proposed a preprocessing method that concatenates short sustained vowels /a/ to 3 s and systematically evaluated the Acoustic Voice Quality Index (AVQI v.03.01) for PD identification. The study provided the first evidence for AVQI’s utility in female PD detection and showed that concatenated short vowels can support portable PD voice diagnostics. Limitations include reliance on a single Spanish dataset and limited sensitivity to male PD voice abnormalities; larger multilingual datasets and additional voice metrics are needed to improve generalization.

La Quatra et al. (2025) developed a bilingual, dual-head deep model that fuses self-supervised learning (SSL) representations with wavelet features, task-specific heads, and cross-lingual optimization to unify PD detection across Slovak (EWA-DB) and Spanish (PC-GITA). The architecture uses a shared backbone with dual heads: WavLM Base embeddings concatenated with wavelet transforms (capturing time-domain detail) after layer normalization. Two task heads handle DDK and continuous speech, activating only the relevant head per input. Adaptive layers, convolutional bottlenecks, and contrastive learning enhance cross-lingual generalization. The study demonstrates the first bilingual unified model for PD speech detection, facilitating remote screening across languages; current coverage is limited to two languages and tasks, motivating future extensions to more languages and tasks (e.g., sustained vowels).

On controlled corpora (often comprising sustained vowels), feature-engineered pipelines combined with classical classifiers — such as

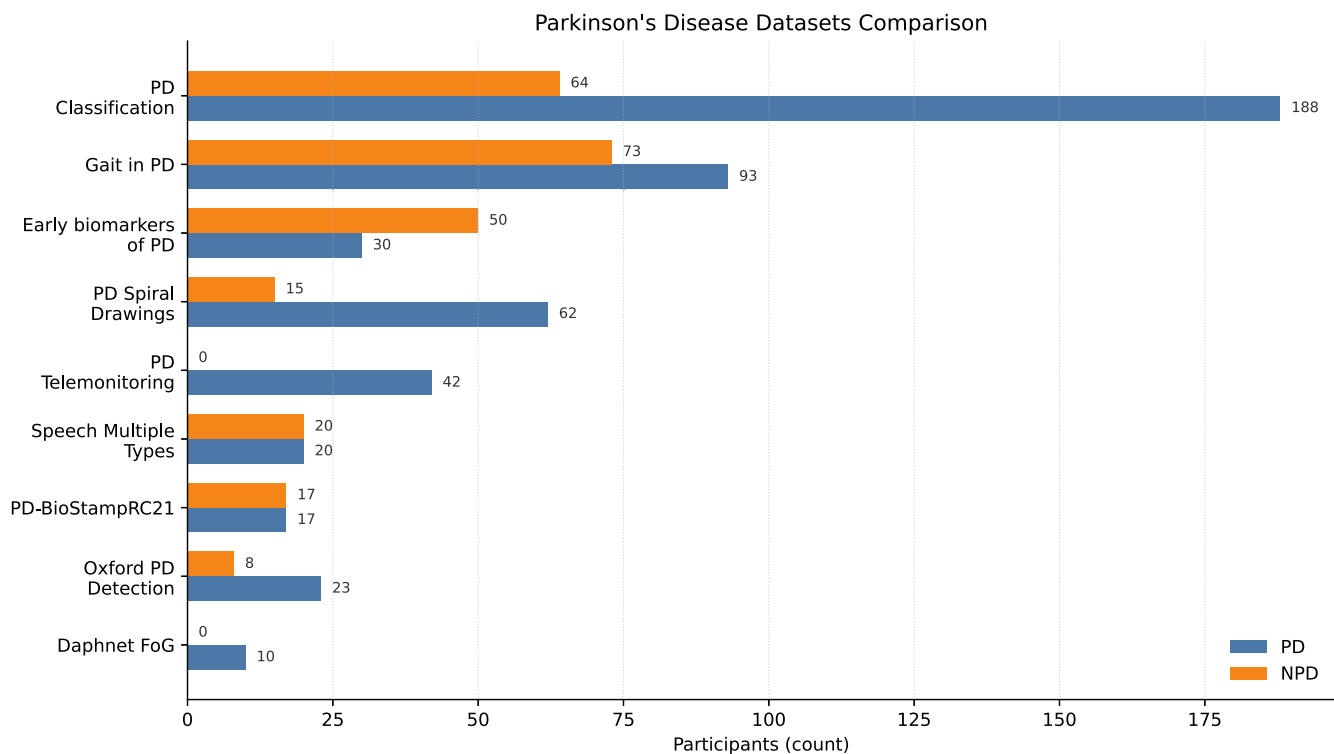


Fig. 10. Datasets: instance and PD patient comparison.

those using Wrapper, IMFCC, or openSMILE features with SVM, random forest, or MLP-LSTM — report strong performance but exhibit limited generalization across languages and recording devices (Solana-Lavalle et al., 2020; Parisi et al., 2018; Karan et al., 2020). In contrast, for tasks involving read or spontaneous speech, as well as multilingual telephone conversations that offer higher ecological validity, end-to-end CNN-based or hybrid transfer learning approaches (e.g., InceptionV3) demonstrate greater robustness (Quan et al., 2022; Iyer et al., 2023; Laganas et al., 2021). Self-supervised learning (SSL) and foundation models — including wav2vec2.0, Whisper, and SeamlessM4T — as well as bilingual dual-head architectures, have shown improved cross-lingual consistency. Nonetheless, their performance degrades under real-world conditions (e.g., in the e-PC-GITA corpus) without further enhancement. Techniques such as speech enhancement (via voice activity detection, dereverberation, and denoising), along with speaker-independent and external-site validation, help narrow this gap (Dao et al., 2025; La Quatra et al., 2024, 2025). Beyond acoustic speech, emerging modalities are broadening the capabilities of non-motor phenotyping. For instance, EEG-based models (e.g., ASGCNN) and facial affect analysis pipelines (e.g., StarGAN combined with Swin Transformer) offer complementary insights. Additionally, GAN-driven augmentation has proven beneficial for modeling raw waveform time-series data (Chang et al., 2023; Huang et al., 2023b; Rey-Paredes et al., 2024).

To enhance translational credibility, methods incorporating explainability (e.g., SHAP) and clinical alignment — such as consensus with UPDRS scores or neuroimaging findings — are increasingly adopted (Shen et al., 2025b; Dao et al., 2025). It is important to note that high benchmark scores do not necessarily imply deployment readiness. Future work should prioritize: standardizing tasks and subject-independent or external validation splits across diverse languages and devices; reporting performance metrics (e.g., F1, sensitivity, specificity, BACC) with 95% confidence intervals and calibration plots; and developing lightweight, multimodal late-fusion frameworks that improve real-world robustness without compromising interpretability or clinical utility (He et al., 2024b; Pah et al., 2024; Laganas et al., 2021; La Quatra et al., 2024, 2025).

Overall, across non-motor modalities, a performance–explainability tension is evident. Feature-based speech pipelines built on MFCC/IMFCC or openSMILE features with classical learners provide transparent acoustic markers, yet they generalize poorly across languages and recording devices (Solana-Lavalle et al., 2020; Parisi et al., 2018; Karan et al., 2020). In contrast, end to end and transfer/SSL/foundation models — ranging from time-distributed ConvNets and Inception V3 to wav2vec2.0, Whisper, WavLM, and bilingual dual-head designs — yield stronger accuracy under realistic conditions but offer limited interpretability (Quan et al., 2022; Iyer et al., 2023; La Quatra et al., 2024; Dao et al., 2025; La Quatra et al., 2025). Two complementary strategies have been adopted: post hoc attribution to quantify feature contributions in hybrid architectures (Shen et al., 2025b), and alignment of predictions with clinical scales or neuroimaging to establish construct validity (Dao et al., 2025). In practice, high-capacity systems should be accompanied by calibrated probability outputs, prespecified decision thresholds, and concise clinician-facing summaries of salient acoustic or temporal features; analogous reporting is advisable for EEG and face modalities, with attention to session/site effects and quality control (Chang et al., 2023; Huang et al., 2023b).

Table 5 synthesizes PD detection by modality, detailing typical preprocessing, representations, key limitations/advantages, and a uniform validation/reporting standard. Speech uses VAD/denoise/dereverb with short-segment analysis and handcrafted or SSL embeddings; its main risks are device/language/channel variability and in-the-wild degradation, mitigated by enhancement and SSL. Gait applies filtering, stride/step detection and windowing with spatiotemporal parameters; it is clinically intuitive but confounded by speed/medication and sensor placement, so subject-independent and external testing with placement notes is advised. EEG relies on bandpass/notch filtering, artifact removal (e.g., ICA), and epoching with PSD/coherence or conv/graph features; results are competitive but limited by small cohorts and inter-subject/session variability, motivating session-/site-independent evaluation. Face requires QC, detection/landmarking and alignment with AU or deep embeddings; privacy and capture conditions (lighting/camera) dominate. Across modalities, recommended practice is to

**Table 5**

Modality to approach summary for PD detection. Abbreviations: L = leave-one-subject/speaker-out (subject/speaker-independent); E = external/site-independent; TS = time series; AU = Action Units.

Modality	Typical preprocessing	Representation	Limitations	Advantages	Recommended validation/reporting
Speech	VAD, denoise, dereverb, normalize; segment (5–10 s)	Handcrafted (MFCC, jitter, shimmer, HNR) or SSL embeddings	*Language/device/channel variability; *In-the-wild degradation; privacy constraints; *Results sensitive to task/split	*Mature toolchain; enhancement and SSL reduce domain shift; *Supports passive smartphone screening	Speaker-/site-independent (L/E); report AUC, F1, SEN, SP with 95% CI; calibration; language/device notes
Gait	Filtering; stride/step detection; windowing; normalization	Spatiotemporal parameters; generic TS features	*Speed/medication confounds; sensor placement variability; *Fewer public datasets	*Intuitive clinical linkage; *Wearable-friendly; *Robust temporal structure	Subject-independent (L) and external (E) where possible; AUC/F1/SE/SP with 95% CI; calibration; placement notes
EEG	Bandpass/notch; artifact removal (e.g., ICA); epoching	PSD/coherence; learned conv/graph features	*Small cohorts; *Inter-subject/session variability; *Setup burden	*Objective neural signal; *complements speech/gait; *supports graph/attention	Session-/site-independent evaluation; AUC/F1/SE/SP with 95% CI; calibration; session/site notes
Face	Quality control; detection/landmarking; alignment; intensity normalization	AU or deep face embeddings	*Privacy/consent; *Lighting/camera bias; *Curated sets may inflate accuracy	*Non-invasive; *affective/motor correlates; *benefits from augmentation/QA	Person-/site-independent splits; AUC/F1/SE/SP with 95% CI; calibration; device/lighting notes

validate L/E and report AUC, F1, sensitivity, and specificity with 95% CIs, and to assess calibration, while disclosing language/device/session (or placement/lighting) metadata.

### 3. Datasets

#### 3.1. Publicly available datasets

Studies on PD detection are highly dependent on diverse datasets that offer high-quality samples for AI-based automated diagnosis and progression prediction. These datasets, used for both motor and non-motor symptom detection, are summarized in Table 6, including their data types, sensor sources, sample sizes, and applicable tasks, which provides a data foundation for subsequent analysis. Several important PD-related datasets have been employed in research. Additionally, Fig. 10 shows the counts of PD and NPD participants in selected PD datasets.

In the existing literature, common PD-related datasets include "the Early Biomarkers of PD based on Natural Connected Speech Signals, PD Spiral Drawings, Parkinson's Telemonitoring (PDT), Parkinson Speech Dataset with Multiple Types of Sound Recordings, Gait Dataset, Parkinson's Progression Markers Initiative, and mPower Dataset".

The sources of these datasets are varied. Some are accessible via the UC Irvine ML Repository and the Kaggle ML Repository, while others, created by clinical hospitals or medical research institutions, are not publicly available. For instance, the PPMI dataset, collected by the Parkinson's Progress Markers Initiative, offers multimodal data, including imaging, genomic information, and biomarkers. The mPower dataset gathers PD patients' voice, gait, and touchscreen interaction data through smartphone apps for remote monitoring and disease progression prediction.

- Oxford PD Detection Dataset (Little et al., 2008): This dataset encompasses 195 acoustic recordings from 31 participants, comprising 23 PD patients and 8 healthy controls. 22 clinically validated bioacoustic parameters characterize each recording. The dataset is structured for binary classification, where the status output variable assumes binary values (0 for healthy controls and 1 for PD instances), with complete data integrity.

- Early biomarkers of PD based on natural connected speech (Hlavnicka et al., 2017): This dataset was constructed from naturally connected speech recordings of 130 individuals, including 30 early-stage untreated PD patients, 50 REM sleep behavior disorder (RBD) patients (at high risk of developing PD), and 50 healthy controls. It contains 65 speech-related parameters extracted from reading and monologue tasks. As a multivariate resource, it is applicable to classification and regression analysis. The dataset is publicly accessible in XLS and CSV formats through the UCI Machine Learning Repository. It provides critical utility for investigating early-stage PD biomarkers through speech features.
- PD Spiral Drawings Using Digitized Graphics Tablet (Isenkul et al., 2014): This dataset, derived from digitized graphics tablet recordings, encompasses handwriting data from 62 individuals with PD and 15 healthy controls. It includes three types of handwriting assessments: the Static Spiral Test (SST), the Dynamic Spiral Test (DST), and the Stability Test on Certain Point (STCP). The dataset also includes spiral drawing images from PD patients. Facilitating both classification and regression analyses, it contains a status output variable (0 for healthy controls and 1 for PD cases). This dataset serves as a critical resource for investigating motor dysfunction and tremor characteristics in PD.
- Parkinson's Telemonitoring (Tsanas et al., 2009): This dataset originates from biomedical recordings of 42 individuals in the early stages of PD. It encompasses 5875 voice recording samples and multiple attributes, including 16 biomedical voice measures. As a regression dataset, its primary goal is to predict motor and total UPDRS scores using voice measures. It provides a helpful foundation for researching the progression of PD and assessing telemonitoring methods for remote symptom monitoring.
- Parkinson's Speech with Multiple Types of Sound Recordings (Sakar et al., 2013): The dataset, archived in the UCI ML Repository, provides a comprehensive multimodal vocal corpus for PD biomarker research. This resource comprises 1040 annotated speech samples from 40 participants (20 PD patients, 20 healthy controls) across diverse phonatory tasks, including sustained vowels (e.g., /a/, /o/), numerical sequences, lexical repetitions, and

**Table 6**  
Comparison of commonly used datasets in PD detection studies.

Dataset	Data Type	Instances	Limitation	Advantages
Oxford Parkinson's Disease Detection (Little et al., 2008)	Speech data	195 (PD:23 NPD:8)	*Small sample size. *Only for binary classification.	*No missing values. *Simple data preprocessing.
Early biomarkers of Parkinson's disease based on natural connected speech (Hlavnicka et al., 2017)	Speech data	130 (PD:30 NPD:50)	*This dataset has missing values. *Limited sample size.	*This dataset includes a high-risk RBD group for early prediction.
Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet (Isenkul et al., 2014)	Handwriting data	77 (PD: 62, NPD: 15)	*Small sample size.	*No missing values. *Suitable for both classification and regression tasks.
Parkinson's Telemonitoring (Tsanas et al., 2009)	Speech data	5875 (PD: 42)	*No healthy control group.	*No missing values. *Suitable for developing and evaluating telemonitoring methods for remote symptom monitoring in PD.
Parkinson's Speech with Multiple Types of Sound Recordings (Sakar et al., 2013)	Speech data	1040 (PD:20, NPD:20)	*The dataset is specific to voice recordings. *Small sample size.	*No missing values. *suitable for ML models.
Daphnet Freezing of Gait (Roggen et al., 2010)	Accelerometry data	237 (PD:10)	*Small sample size. *No healthy control group included.	*Focused on FoG detection, useful for gait analysis in PD patients.
Parkinson's Disease Classification (Sakar et al., 2018)	Speech data	756 (PD:188, NPD:64)	*Class imbalance. *Acoustic limitations of the data.	*Focuses on voice features. *Provides binary labels for classification tasks.
Gait in Parkinson's Disease (Frenkel-Toledo et al., 2005a,b; Hausdorff et al., 2007; Yogev et al., 2005)	VGRF data	305 (PD:93, NPD:73)	*Class imbalance.	*Diverse experimental conditions. *Standardized task design and acquisition protocols. *High clinical applicability. *No missing values.
PD-BioStampRC21: Dataset (Adams et al., 2020)	Accelerometry data	34 (PD:17, NPD:17)	*Small sample size.	*Rich movement data for multi-sensor analysis. *Synchronized clinical assessments (UPDRS) enhance interpretability of sensor data.
Parkinson's Progression Markers Initiative (Marek et al., 2011) (PPMI)	SPECT imaging	–	–	*No missing values. *This dataset is widely utilized by researchers for method evaluation.
M-Power dataset (Bot et al., 2016)	–	–	–	*Large-scale and Diverse Data Collection.

short sentence recitations. Each recording is characterized by 26 clinically validated acoustic features, such as jitter (local, RAP), shimmer (APQ3, APQ5), harmonic-to-noise ratios (HNR), and nonlinear pitch dynamics (median/mean pitch, standard deviation). Additionally, UPDRS scores are provided for regression-based severity assessment. The dataset is partitioned into a training set (26 task types per subject) and an independent test set

(168 recordings from 28 PD patients, focusing on sustained vowels), enabling robust validation of diagnostic models.

- Daphnet Freezing of Gait Dataset (Roggen et al., 2010): This dataset was created from biomedical accelerometer readings from 10 patients with PD experiencing FoG during walking tasks. The dataset consists of multivariate time-series data, including acceleration measurements from the hip and leg. The data is annotated with freeze events, making it suitable for classification tasks

aimed at recognizing FoG. This dataset is designed to benchmark automatic methods for gait freeze recognition.

- **PD Classification Dataset (Sakar et al., 2018):** This dataset includes data from 188 PD patients (107 males, 81 females aged 33–87) and 64 healthy controls (23 males, 41 females aged 41–82). Speech recordings of the sustained vowel /a/ were collected at 44.1 kHz, and 754 features were extracted using various speech signal processing algorithms for PD assessment and classification. This dataset contains complete data without any missing values, making it appropriate for binary classification analysis. The target variable “status” is defined as 0 for healthy individuals and 1 for those with PD. This dataset offers a foundation for constructing AI models aimed at PD prediction using recorded data.
- **PD-BioStampRC21 Dataset (Adams et al., 2020):** This dataset is available on IEEE DataPort and includes biomedical accelerometer data from 34 participants (17 PD patients and 17 controls). The dataset comprises sensor data from five wearable sensors placed on the trunk, left and right anterior thighs, and left and right anterior forearms. It contains multivariate time-series data, including acceleration measurements and annotations for UPDRS assessments. The dataset is designed to support research on automatic methods for gait freeze recognition and other motor symptom analyses in PD.
- **Gait in Parkinson’s Disease (Frenkel-Toledo et al., 2005a,b; Hausdorff et al., 2007; Yogeve et al., 2005):** This PhysioNet gait database is designed for Parkinson’s disease research and provides high-resolution VGRF recordings collected during real-world walking. It includes 93 patients with idiopathic PD and 73 age-matched healthy controls who walked for approximately two minutes at a self-selected pace on level ground. Eight force sensors were placed under each foot (16 channels in total) and sampled at 100 Hz; two additional channels contain the per-foot sums. Demographic information and clinical scales (Hoehn–Yahr stage and UPDRS) are also provided. The raw force traces enable the derivation of key spatiotemporal measures (e.g., stride time, swing and stance durations, center-of-pressure trajectories) and stride-to-stride variability. A dual-task subset (walking while performing serial-7 subtraction) allows quantitative assessment of cognitive–motor interference. The standardized acquisition protocol, adequate cohort size, multi-condition coverage, and rich annotations make this corpus suitable both as a benchmark for traditional feature-based classification and for end-to-end learning with clinically aligned analyses.

### 3.2. Data preprocessing

In AI-based PD prediction, data preprocessing is crucial for enhancing data quality and model performance. Different data types (e.g., voice, gait) require distinct processing methods. Proper techniques can reduce noise, improve data quality, and boost model accuracy. Preprocessing for PD classification comprises data cleaning, feature extraction, normalization, and data augmentation. The following are some of the pretreatment methods studied.

Chen et al. (2023) enhanced the detection algorithm using an interpretable DL architecture. In data preprocessing, they applied CWT to the raw data to obtain detailed time–frequency information. CWT can dynamically analyze non-stationary signals and capture more time–frequency details. Specifically, they performed CWT on each time series (triaxial acceleration and angular velocity) from each sensor using a Morlet wavelet, providing richer inputs for the CNN model.

Zhao et al. (2021) employed preprocessing methods on the PD dataset, including the removal of initial and terminal effects, data segmentation, and normalization. In the removal of initial and terminal effects, the first 100 points and the last 100 points of the gait data were eliminated to reduce artifacts caused by gait initiation and termination.

Subsequently, the raw signals were divided into multiple continuous segments using time windows, with the “Ju” section segmented into 20 intervals and the “Ga” and “Si” sections each segmented into 60 intervals. Additionally, to mitigate the influence of individual body weight on VGRF, VGRF signals were normalized by the subject’s body weight, yielding the “VGRF/BW” signal data.

Chang et al. (2023) applied signal removal, independent component analysis (ICA), signal re-referencing, and feature extraction to preprocess the PD dataset. During signal removal, electrodes near the ventral temporal lobe with unreliable signals were excluded. ICA was then used to remove blink artifacts and ensure signal purity. Subsequently, the data were re-referenced to an average reference to standardize the signal reference point. Finally, log-spectral power features from each electrode in five frequency bands were extracted and mapped onto a brain topographic map to visualize the spatial distribution of frequency-specific features across brain regions.

Solana-Lavalle et al. (2020) used standardization and feature subset selection as preprocessing methods on the PD dataset. For standardization, they standardized all features to have zero mean and unit standard deviation, eliminating dimensional differences between features. For feature subset selection, they used a Wrappers feature selection algorithm that combines forward and backward stepwise selection. This selected 8–20 of the most relevant and non-redundant features from 754, improving model performance and reducing computational complexity.

## 4. Clinical applications

Clinical diagnosis plays a pivotal role in the early detection and therapeutic intervention of PD (Zhang, 2022). Nevertheless, the diagnostic process faces significant challenges due to the disease’s insidious onset and progressive symptom exacerbation. Current clinical protocols predominantly rely on subjective clinical assessments and standardized evaluation instruments, such as the UPDRS. Notwithstanding these established protocols, empirical studies have identified intrinsic limitations in conventional methodologies, including elevated misdiagnosis rates and substantial inter-rater variability in subjective evaluations (Pan et al., 2025). This diagnostic uncertainty stems from the inherent complexity of differentiating PD from atypical Parkinsonian syndromes during early disease stages (Kondo et al., 2024).

### 4.1. Integration with clinical workflows and levels of autonomy

To maximize clinical utility, AI tools should map cleanly onto existing neurological workflows (history taking, focused examination, imaging/lab as indicated, and longitudinal follow-up), while making their level of autonomy explicit:

- **Assistive decision support (ADS):** AI produces a calibrated risk score and an interpretable summary for clinician review (e.g., salient acoustic features for dysarthria, stride-time variability for gait). Decisions remain with the neurologist. Evidence required: speaker/subject-independent external validation; reporting of F1/AUC with 95% CIs and calibration.
- **Semi-automated triage/monitoring:** AI screens remote inputs (phone speech, wearables) to flag visits or medication adjustments. Evidence required: prospective longitudinal validation, stability under device/language shift, and decision-analytic reporting (net benefit).
- **Automated action:** Rare in current PD care; would require prospective clinical trials demonstrating safety/effectiveness and regulatory clearance as Software as a Medical Device (SaMD). We view this as a future step.

Practical workflow insertions include:

- Intake screening: a 30–60 s speech task or phone-call sample is analyzed to provide a calibrated PD likelihood with confidence intervals and an EHR-ready note;
- Clinic assessment: wearable gait sensors quantify bradykinesia and instability alongside UPDRS;
- Home monitoring: passive calls or smartwatch accelerometry trend severity between visits. Reports should align scores with clinical thresholds (e.g., UPDRS cut-points), provide uncertainty estimates, and include brief explanations to support shared decision-making.

## 4.2. Clinical diagnosis assistance

### 4.2.1. Automated imaging analysis

Computer vision technology plays a pivotal role in early PD detection by analyzing neuroimaging data, such as MRI and computed tomography (CT) scans. Zhang (2022) demonstrated that AI can automatically extract lesion regions from these images and analyze them using DL models, such as convolutional neural networks. This approach enables the identification of PD-related structural changes in brain regions, including damage to the substantia nigra. For instance, DL models can detect volumetric alterations in the substantia nigra, providing early indicators of PD onset.

*Workflow fit:* imaging analysis functions as ADS—radiology or neurology reviews an AI heatmap, calibrated score, and differential list before final reporting.

### 4.2.2. Speech analysis and diagnosis

Speech features in voice signals, such as pitch, speech rate, and articulation clarity, are important early indicators of PD. Quan et al. (2022) emphasized that over 90% of PD patients suffer from dysarthria, which is characterized by reduced vocal loudness, monotonous speech, and impaired articulation. AI can analyze patients' speech samples to identify potential PD symptoms. Goyal et al. (2021) proposed a hybrid approach combining resonance-based and time–frequency-based features to improve PD diagnosis through voice signals. AI models use ML algorithms (e.g., SVM, RF) in conjunction with patients' voice characteristics to classify speech features of healthy individuals and PD patients, thereby assisting physicians in diagnosing PD. Laganas et al. (2021) demonstrated the potential of using telephone speech data for PD detection, achieving significant results in distinguishing PD patients from healthy controls across different languages. *Workflow fit:* in clinic, a short, sustained vowel and a reading passage can be recorded on a tablet; the tool returns a calibrated PD likelihood and an explanation (e.g., SHAP attributions over acoustic cues) for ADS. For remote monitoring, passive phone calls are processed server-side with language/device normalization (Laganas et al., 2021; La Quatra et al., 2024; Iyer et al., 2023). Foundation-model ensembles can report scores correlated with UPDRS and DaTSCAN to support risk stratification (Dao et al., 2025).

### 4.2.3. Multimodal data fusion

In the domain of neurodegenerative disease diagnosis, particularly for PD, the integration of multimodal data has become a crucial strategy to enhance diagnostic accuracy (Lee et al., 2022). By amalgamating data from diverse sources such as genetic information, transcriptomics, clinical assessments, and imaging, and employing advanced DL techniques, a more comprehensive evaluation of the disease state can be achieved. For instance, (Makarious et al., 2022) demonstrated the efficacy of integrating genetic, transcriptomic, and clinico-demographic data from the PPMI dataset using ML to predict PD risk, achieving an AUC of 89.72% in their initial model and 85.03% in an external validation cohort. Similarly, integrating gait and speech data has been shown to improve the accuracy of PD detection significantly. Lv et al. (2024) developed a multimodal DL framework that combines visual and audio features, achieving a PD detection accuracy of 92.68%,

surpassing conventional models. These studies underscore the potential of multimodal data fusion, particularly via DL methods, to capture the complex, subtle changes indicative of PD, thereby providing more accurate diagnostic results for clinical practice. Integrating such diverse data modalities not only enhances diagnostic accuracy but also offers deeper insights into the disease mechanisms, potentially aiding in the development of targeted therapeutic interventions.

*Workflow fit:* multimodal tools operate as ADS/triage—outputs are fused risk scores with explanation and cross-modal consistency checks, flagged for clinician review.

## 4.3. Disease progression monitoring

### 4.3.1. Real-time symptom monitoring

The integration of wearable sensors, such as smartwatches, accelerometers, and gyroscopes, with AI, enables continuous, real-time monitoring of patients' motor symptoms, such as gait, tremors, and bradykinesia, and provides a quantitative assessment of changes in motor function through ML models. For instance, (Crowe et al., 2024) developed a wearable-enabled algorithm for estimating Parkinson's symptoms in a continuous home-monitoring setting using inertial sensors, achieving balanced accuracies of 63%–67% in unscripted home monitoring, highlighting the potential of such systems for objective symptom assessment. Additionally, (Gonçalves et al., 2021) introduced a gait monitoring system for Parkinson's disease patients, which demonstrated high sensitivity (99.53%) and accuracy (97.42%) in detecting gait events, and the ability to estimate spatiotemporal gait parameters with a good error margin. These studies underscore the importance of real-time monitoring for identifying subtle changes in motor symptoms, which can help clinicians make informed treatment decisions and adjust therapies accordingly (Raza et al., 2020). The advancements in wearable technology and AI-driven analysis have the potential to transform the management of motor symptoms in Parkinson's disease, offering a more objective and comprehensive approach to symptom assessment and disease progression monitoring. *Workflow fit:* home monitoring operates at the semi-automated level—alerts are generated from calibrated thresholds and reviewed by clinicians; deployment should report latency, false-alarm rates, and energy use.

## 4.4. Real-world validation and regulatory considerations

Real-world evidence now spans several application types and clarifies their role in care pathways. For speech, multilingual smartphone calls have been evaluated with leave-one-subject-out designs using language- and device-aware training, supporting assistive screening in practice (Laganas et al., 2021). Foundation-model ensembles operating on short segments show external correlations with UPDRS and DaTSCAN, indicating utility for risk stratification (Dao et al., 2025). On the e-PC-GITA corpus collected under real-world operating conditions, combining enhancement with foundation models yields site-independent gains and improved robustness across devices and languages (La Quatra et al., 2024). Transfer learning on low-resolution phone audio demonstrates subject-independent performance under variable phone quality (Iyer et al., 2023). For gait, prospective home-monitoring studies report continuous estimation with attention to sensor placement and are positioned as semi-automated monitoring (Crowe et al., 2024), whereas clinic-based systems reliably detect gait events and derive stance/stride parameters for clinician decision support (Gonçalves et al., 2021).

Regulatory and implementation considerations. Most systems currently function as clinician-in-the-loop decision support or triage tools and would be regulated as Software as a Medical Device (SaMD). Demonstration of robustness should include external/site-independent validation, calibrated probabilities with 95% confidence intervals, and prespecified subgroup analyses; for monitoring applications, latency

and false-alarm rates should also be reported. Privacy and security require data minimization, on-device preprocessing when feasible, secure transmission/storage, and explicit consent management. Seamless EHR integration and concise clinician-facing explanations (e.g., rationale summaries) are key enablers of adoption.

## 5. Challenges

### 5.1. Dataset

ML and DL can efficiently handle complex data in PD detection. They automatically extract features, enabling rapid and accurate diagnosis and enhancing diagnostic efficiency (Ghane et al., 2022). However, their data dependency poses several challenges. Many projects have small sample sizes (e.g., < 50 participants, as low as 12) and data from a single source, which limit the model's generalization and adaptability to diverse data (Ahmadi et al., 2024).

In addition, data annotation is challenging. Acquiring precise clinical diagnostic labels requires professional medical knowledge and is time-consuming, especially for early-stage PD and cognitive impairment cases. Differences in data collection methods, equipment, and standards across studies result in inconsistent data quality. This impacts model training and diagnostic accuracy, causing unstable performance across different datasets (Sigcha et al., 2023; ul Haq et al., 2022).

Furthermore, there is a significant data imbalance issue in PD prediction datasets. The number of samples between healthy individuals and patients is often uneven. This imbalance can cause models to be biased towards predicting the class with more samples, affecting prediction performance, especially for the minority class, such as early-stage PD patients (Deharab and Ghaderyan, 2022).

### 5.2. Classification models

Many models see performance declines in independent datasets or real-world settings, particularly with cross-device or cross-center data. DL models, often regarded as “black boxes”, have trouble meeting the clinical demand for transparent decision-making (Amato et al., 2023). Data fusion for different sensors (e.g., inertial, voice, EEG) remains immature, with feature redundancy and ineffective utilization of complementary information (Ali et al., 2019).

Moreover, model selection and adaptation present significant challenges. Selecting the optimal model for specific data types and research objectives from a wide array is no easy task, and models show substantial performance variations across datasets.

### 5.3. Clinical capability

In addition to challenges in data and algorithms, most studies lack sufficient validation in real-world medical settings, making it hard to assess their effectiveness and practicality in actual healthcare scenarios (Zhang, 2022). Communication and collaboration barriers among professionals in computer science, neurology, psychology, etc., hinder research progress and technology transfer. In addition, the field lacks unified standards and evaluation frameworks. Different studies use heterogeneous evaluation metrics, preventing direct comparisons and comprehensive analyses.

In terms of technical implementation and ethics, medical AI faces multi-faceted challenges. During clinical translation, there is insufficient real-world validation. Most studies remain in the laboratory stage, not fully considering individual differences between patients and tracking long-term treatment effects (Mohammed et al., 2017). Data privacy protection needs to balance information sharing with security requirements. Insufficient cross-disciplinary collaboration leads to model designs detached from clinical reality, such as ignoring doctors' intuitive judgments and patient behavioral data. Ethical issues include algorithmic bias (e.g., diagnostic disparities for ethnic minorities) and liability definition (e.g., legal attribution for AI misdiagnosis).

## 6. Future directions

In light of the aforementioned challenges, future PD prediction research should focus on the following directions: data expansion, multimodal fusion, and enhanced interpretability.

### 6.1. Dataset construction and optimization

Previous studies have used small-scale datasets. Future research should focus on building better datasets to boost model performance and generalization. This involves expanding data collection and integrating multimodal data (e.g., sensors, audio, and images) to build comprehensive, representative datasets. Unified data standards and formats should be established to promote sharing and exchange, facilitating comparisons across studies. Data quality control must be strengthened to minimize noise, ensure accuracy, and include diverse samples covering various population characteristics and disease stages.

Innovative data augmentation methods (e.g., transfer learning, or GANs) should be explored to address small-sample or data-imbalance issues. Moreover, testing models on larger datasets can yield more reliable performance and generalization results. Cross-validation can tackle data imbalance, and optimizing feature sets by removing irrelevant features can further enhance model performance.

As a practical recommendation for comparability, future studies should report, on a held-out test set, AUC, F1 score, sensitivity, and specificity with 95% confidence intervals and calibration, together with a clearly stated validation protocol indicating speaker/subject and site independence.

### 6.2. Future directions in model and algorithm development

Future research should break new ground in multimodal fusion, model explainability, and personalized modeling to boost diagnostic and assessment accuracy. This involves exploring more efficient multimodal fusion frameworks (e.g., cross-modal attention mechanisms and graph neural networks) and combining explainability techniques (e.g., SHAP and LIME) to strengthen clinical credibility. For example, Makarios et al. (2022) proposed a multimodal ML method that integrates clinical, genetic, and transcriptomic data to develop a model for predicting PD risk.

To meet wearable devices' real-time monitoring needs, optimizing model compression and edge computing capabilities is essential. Targeting individual differences and developing personalized models based on patient history and dynamic physiological data for adaptive parameter adjustment is also crucial.

### 6.3. Future research on bridging technology and application

Future clinical translation of AI requires bridging the gap between technical validation and real-world medical scenarios, creating an intelligent diagnostic–therapeutic system that balances efficacy and ethics. Technically, multi-center clinical trials should be conducted to test the robustness of AI models in complex clinical settings (e.g., comorbidity interference and individualized treatment responses), and dynamic tracking systems should be developed to evaluate long-term efficacy.

To address interdisciplinary collaboration barriers, a collaborative innovation framework between computer science and clinical medicine should be established, incorporating physicians' expertise and patient behavior data into model design. Ethically and implementation-wise, algorithmic fairness verification mechanisms should be developed to eliminate diagnostic biases (e.g., data compensation strategies for minority populations), and AI-related medical accountability rules (e.g., responsibility definitions in misdiagnosis cases) should be clarified.

Table A.7

Acronyms used in the paper.

Acronym	Description
ML	Machine Learning
DL	Deep Learning
PD	Parkinson's Disease
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
RF	Random Forest
ACC	Accuracy
SP	Specificity
SE	Sensitivity
F1	F1-Score
LOSO	Leave One Subject Out
AUC	Area Under the Curve
FNR	False Negative Rate
ECG	Electrocardiography
EEG	Electroencephalogram
PPMI	Parkinson's Progression Markers Initiative
MRI	Magnetic Resonance Imaging
IMU	Inertial Measurement Unit
CWT	Continuous Wavelet Transform
FoG	Freezing of Gait
SSL	Self-supervised Learning
CI	confidence interval
MLP	Multilayer Perceptron

7. Conclusion

This study presented an in-depth exploration of the application of ML and DL in the diagnosis of PD, with a particular emphasis on early diagnosis and the analysis of both motor and non-motor symptoms. Through the analysis of diverse data types, such as gait, speech, and imaging data, the significant advantages of AI algorithms in enhancing diagnostic accuracy and differentiating PD patients from healthy individuals were evident. Specifically, when dealing with data related to motor symptoms (e.g., gait analysis, tremor, bradykinesia) and non-motor symptoms (e.g., Dysphonia, Dysarthria), AI can automatically extract key features. This assists clinicians in conducting a more comprehensive assessment of the patient's disease status. However, despite the substantial potential demonstrated by AI in automated diagnosis, several challenges persist. Issues such as data standardization, dataset imbalance, and model interpretability continue to act as major obstacles in current research, restricting the extensive clinical application of AI technologies.

CRediT authorship contribution statement

**Xingkai Fu:** Writing – original draft, Visualization, Resources, Methodology, Formal analysis, Data curation. **Sike Ni:** Writing – original draft, Validation, Formal analysis, Conceptualization. **Mohammed A.A. Al-qaness:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Conceptualization.

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.

Appendix. Abbreviations

See Table A.7.

Data availability

No data was used for the research described in the article.

References

Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaipappan, E., Venkatraman, V., 2018. Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Future Gener. Comput. Syst.* 83, 366–373.

Adams, J.L., Dinesh, K., Snyder, C.W., Xiong, M., Tarolli, C.G., Sharma, S., Dorsey, E.R., Sharma, G., 2020. PD-BioStampRC21: Parkinson's disease accelerometry dataset from five wearable sensor study. <http://dx.doi.org/10.21227/g2g8-1503>.

Aghanavesi, S., Bergquist, F., Nyholm, D., Senek, M., Memedi, M., 2019. Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge. *IEEE J. Biomed. Health Inform.* 24 (1), 111–119.

Ahmadi, H., Huo, L., Arji, G., Sheikhtaheri, A., Zhou, S.-M., 2024. Early diagnosis of Parkinson's disease using a hybrid method of least squares support vector regression and fuzzy clustering. *Biocybern. Biomed. Eng.* 44 (3), 569–585.

Ahmed, S., Komeili, M., Park, J., 2022. Predictive modelling of Parkinson's disease progression based on RNA-sequence with densely connected deep recurrent neural networks. *Sci. Rep.* 12 (1), 21469.

Alcaraz, J., Labbé, M., Landete, M., 2022. Support Vector Machine with feature selection: A multiobjective approach. *Expert Syst. Appl.* 204, 117485.

Alharthi, A.S., Casson, A.J., Ozanyan, K.B., 2020. Gait spatiotemporal signal analysis for Parkinson's disease detection and severity rating. *IEEE Sensors J.* 21 (2), 1838–1848.

Ali, L., Zhu, C., Zhou, M., Liu, Y., 2019. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. *Expert Syst. Appl.* 137, 22–28.

Almeida, J.S., Rebouças Filho, P.P., Carneiro, T., Wei, W., Damaševičius, R., Maskeliūnas, R., de Albuquerque, V.H.C., 2019. Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognit. Lett.* 125, 55–62.

Altham, C., Zhang, H., Pereira, E., 2024. Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's disease: A systematic review. *Plos One* 19 (5), e0303644.

Amato, F., Saggio, G., Cesarini, V., Olmo, G., Costantini, G., 2023. Machine learning- and statistical-based voice analysis of Parkinson's disease patients: A survey. *Expert Syst. Appl.* 219, 119651.

Avolio, M., Fuduli, A., 2020. A semiproximal support vector machine approach for binary multiple instance learning. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (8), 3566–3577.

Balaji, E., Brindha, D., Balakrishnan, R., 2020. Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Appl. Soft Comput.* 94, 106494.

Bhandari, N., Walambe, R., Kotecha, K., Kaliya, M., 2023. Integrative gene expression analysis for the diagnosis of Parkinson's disease using machine learning and explainable AI. *Comput. Biol. Med.* 163, 107140.

Borzi, L., Sigcha, L., Rodríguez-Martín, D., Olmo, G., 2023. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. *Artif. Intell. Med.* 135, 102459.

Bot, B.M., Suver, C., Neto, E.C., Kellen, M., Klein, A., Bare, C., Doerr, M., Pratap, A., Wilbanks, J., Dorsey, E., et al., 2016. The mpower study, Parkinson disease mobile data collected using ResearchKit. *Sci. Data* 3 (1), 1–9.

Cao, Z., John, A.R., Chen, H.-T., Martens, K.E., Georgiades, M., Gilat, M., Nguyen, H.T., Lewis, S.J., Lin, C.-T., 2021. Identification of EEG dynamics during freezing of gait and voluntary stopping in patients with Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1774–1783.

Celik, G., Başaran, E., 2023. Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals. *Appl. Acoust.* 211, 109476.

Chang, H., Liu, B., Zong, Y., Lu, C., Wang, X., 2023. EEG-based Parkinson's disease recognition via attention-based sparse graph convolutional neural network. *IEEE J. Biomed. Health Inform.*

Chen, H., Fu, J., Liu, X., Zheng, Z., Luo, X., Zhou, K., Xu, Z., Geng, D., 2024. A Parkinson's disease-related nuclei segmentation network based on CNN-transformer interleaved encoder with feature fusion. *Comput. Med. Imaging Graph.* 118, 102465.

Chen, H.-L., Huang, C.-C., Yu, X.-G., Xu, X., Sun, X., Wang, G., Wang, S.-J., 2013. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst. Appl.* 40 (1), 263–271.

Chen, Y., Mao, Q., Wang, B., Duan, P., Zhang, B., Hong, Z., 2022a. Privacy-preserving multi-class support vector machine model on medical diagnosis. *IEEE J. Biomed. Health Inform.* 26 (7), 3342–3353.

Chen, M., Sun, Z., Su, F., Chen, Y., Bu, D., Lyu, Y., 2022b. An auxiliary diagnostic system for Parkinson's disease based on wearable sensors and genetic algorithm optimized random forest. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 2254–2263.

- Chen, M., Sun, Z., Xin, T., Chen, Y., Su, F., 2023. An interpretable deep learning optimized wearable daily detection system for Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* 31, 3937–3946.
- Chu, C., Zhang, Z., Song, Z., Xu, Z., Wang, J., Wang, F., Liu, W., Lu, L., Liu, C., Zhu, X., et al., 2022. An enhanced EEG microstate recognition framework based on deep neural networks: an application to Parkinson's disease. *IEEE J. Biomed. Health Inform.* 27 (3), 1307–1318.
- Cramb, K.M., Beccano-Kelly, D., Cragg, S.J., Wade-Martins, R., 2023. Impaired dopamine release in Parkinson's disease. *Brain* 146 (8), 3117–3132.
- Cristianini, N., 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.
- Crowe, C., Sica, M., Kenny, L., O'Flynn, B., Mueller, D.S., Timmons, S., Barton, J., Tedesco, S., 2024. Wearable-enabled algorithms for the estimation of Parkinson's symptoms evaluated in a continuous home monitoring setting using inertial sensors. *IEEE Trans. Neural Syst. Rehabil. Eng.*
- Dao, Q., Jeancolas, L., Mangone, G., Sambin, S., Chalançon, A., Gomes, M., Lehericy, S., Corvol, J.-C., Vidailhet, M., Arnulf, I., et al., 2025. Detection of early Parkinson's disease by leveraging speech foundation models. *IEEE J. Biomed. Health Inform.*
- Dar, M.N., Akram, M.U., Yuvraj, R., Khawaja, S.G., Murugappan, M., 2022. EEG-based emotion charting for Parkinson's disease patients using Convolutional Recurrent Neural Networks and cross dataset learning. *Comput. Biol. Med.* 144, 105327.
- Deharab, E.D., Ghaderyan, P., 2022. Graphical representation and variability quantification of handwriting signals: New tools for Parkinson's disease detection. *Biocybern. Biomed. Eng.* 42 (1), 158–172.
- Demrozi, F., Bacchin, R., Tamburin, S., Cristani, M., Pravadelli, G., 2019. Toward a wearable system for predicting freezing of gait in people affected by Parkinson's disease. *IEEE J. Biomed. Health Inform.* 24 (9), 2444–2451.
- Diaz, M., Moetesum, M., Siddiqi, I., Vessio, G., 2021. Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs. *Expert Syst. Appl.* 168, 114405.
- Dong, C., Chen, Y., Huan, Z., Li, Z., Zhou, B., Liu, Y., 2023. Static-Dynamic temporal networks for Parkinson's disease detection and severity prediction. *IEEE Trans. Neural Syst. Rehabil. Eng.* 31, 2205–2213.
- El Maachi, I., Bilodeau, G.-A., Bouachir, W., 2020. Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst. Appl.* 143, 113075.
- Exley, T., Moudy, S., Patterson, R.M., Kim, J., Albert, M.V., 2022. Predicting updrs motor symptoms in individuals with Parkinson's disease from force plates using machine learning. *IEEE J. Biomed. Health Inform.* 26 (7), 3486–3494.
- Freire-Álvarez, E., Ramírez, I.L., García-Ramos, R., Carrillo, F., Santos-García, D., Gómez-Esteban, J.C., Martínez-Castrillo, J.C., Martínez-Torres, I., Madrid-Navarro, C.J., Pérez-Navarro, M.J., et al., 2025. Artificial intelligence for identification of candidates for device-aided therapy in Parkinson's disease: DELIST-PD study. *Comput. Biol. Med.* 185, 109504.
- Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005a. Effect of gait speed on gait rhythmicity in Parkinson's disease: variability of stride time and swing time respond differently. *J. Neuroeng. Rehabil.* 2 (1), 23.
- Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005b. Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Mov. Disord.: Off. J. Mov. Disord. Soc.* 20 (9), 1109–1114.
- Ghaderyan, P., Fathi, G., 2021. Inter-limb time-varying singular value: A new gait feature for Parkinson's disease detection and stage classification. *Measurement* 177, 109249.
- Ghane, M., Ang, M.C., Nilashi, M., Sorooshian, S., 2022. Enhanced decision tree induction using evolutionary techniques for Parkinson's disease classification. *Biocybern. Biomed. Eng.* 42 (3), 902–920.
- Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., et al., 2008. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.: Off. J. Mov. Disord. Soc.* 23 (15), 2129–2170.
- Gonçalves, H.R., Rodrigues, A., Santos, C.P., 2021. Gait monitoring system for patients with Parkinson's disease. *Expert Syst. Appl.* 185, 115653.
- Goyal, J., Khandnor, P., Aseri, T.C., 2021. A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals. *Expert Syst. Appl.* 182, 115283.
- Guo, Y., Huang, D., Zhang, W., Wang, L., Li, Y., Olmo, G., Wang, Q., Meng, F., Chan, P., 2022. High-accuracy wearable detection of freezing of gait in Parkinson's disease based on pseudo-multimodal features. *Comput. Biol. Med.* 146, 105629.
- Guo, Y., Wang, L., Li, Y., Guo, L., Meng, F., 2019. The detection of freezing of gait in Parkinson's disease using asymmetric basis function TV-ARMA time-frequency spectral estimation method. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (10), 2077–2086.
- ul Haq, A., Li, J.P., Agbley, B.L.Y., Mawuli, C.B., Ali, Z., Nazir, S., Din, S.U., 2022. A survey of deep learning techniques based Parkinson's disease recognition methods employing clinical data. *Expert Syst. Appl.* 208, 118045.
- Hausdorff, J.M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., Giladi, N., 2007. Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *Eur. J. Neurosci.* 26 (8), 2369–2375.
- He, T., Chen, J., Xu, X., Fortino, G., Wang, W., 2024a. Early detection of Parkinson's disease using deep NeuroEnhanceNet with smartphone walking recordings. *IEEE Trans. Neural Syst. Rehabil. Eng.*
- He, T., Chen, J., Xu, X., Wang, W., 2024b. Exploiting smartphone voice recording as a digital biomarker for Parkinson's disease diagnosis. *IEEE Trans. Instrum. Meas.* 73, 1–12.
- Hlavnicka, J., Tykalov, T., Onka, K., Rika, E., Rusz, J., J., 2017. Early biomarkers of Parkinson's disease based on natural connected speech. <http://dx.doi.org/10.24432/C5W02Q>, UCI Machine Learning Repository.
- Huan, X., Zhou, H., Jung, B., Ma, L., 2025. Enhancing gait analysis for Parkinson's disease detection and severity staging with a parallel Conv1D-efficient transformer and bidirectional GRU hybrid architecture. *IEEE Access*.
- Huang, T., Li, M., Huang, J., 2023a. Recent trends in wearable device used to detect freezing of gait and falls in people with Parkinson's disease: A systematic review. *Front. Aging Neurosci.* 15, 1119556.
- Huang, W., Xu, W., Wan, R., Zhang, P., Zha, Y., Pang, M., 2023b. Auto diagnosis of Parkinson's disease via a deep learning model based on mixed emotional facial expressions. *IEEE J. Biomed. Health Inform.*
- Icenkul, M., Sakar, B., Kursun, O., et al., 2014. Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease. In: *The 2nd International Conference on E-Health and Telemedicine. ICEHTM-2014*, vol. 5, pp. 171–175.
- Islam, M.A., Majumder, M.Z.H., Hussein, M.A., Hossain, K.M., Miah, M.S., 2024. A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. *Heliyon*.
- Iyer, A., Kemp, A., Rahmatallah, Y., Pillai, L., Glover, A., Prior, F., Larson-Prior, L., Virmani, T., 2023. A machine learning method to process voice samples for identification of Parkinson's disease. *Sci. Rep.* 13 (1), 20615.
- Karaman, O., Çakın, H., Alhudhaif, A., Polat, K., 2021. Robust automated Parkinson disease detection based on voice signals with transfer learning. *Expert Syst. Appl.* 178, 115013.
- Karan, B., Sahu, S.S., Mahto, K., 2020. Parkinson disease prediction using intrinsic mode function based features from speech signal. *Biocybern. Biomed. Eng.* 40 (1), 249–264.
- Kondo, Y., Bando, K., Suzuki, I., Miyazaki, Y., Nishida, D., Hara, T., Kadone, H., Suzuki, K., 2024. Video-based detection of freezing of gait in daily clinical practice in patients with parkinsonism. *IEEE Trans. Neural Syst. Rehabil. Eng.*
- Kumar, K., Ghosh, R., 2025. A multi-modal Parkinson's disease diagnosis system from EEG signals and online handwritten tasks using grey wolf optimization based deep learning model. *Biomed. Signal Process. Control.* 100, 106946.
- La Quatra, M., Orozco-Arroyave, J.R., Siniscalchi, M.S., 2025. Bilingual dual-head deep model for Parkinson's disease detection from speech. In: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE*, pp. 1–5.
- La Quatra, M., Turco, M.F., Svendsen, T., Salvi, G., Orozco-Arroyave, J.R., Siniscalchi, M.S., 2024. Exploiting foundation models and speech enhancement for Parkinson's disease detection from speech in real-world operative conditions. *arXiv preprint arXiv:2406.16128*.
- Laganas, C., Iakovakis, D., Hadjimitsi, S., Charisis, V., Dias, S.B., Bostantzopoulou, S., Katsarou, Z., Klingelhoefer, L., Reichmann, H., Trivedi, D., et al., 2021. Parkinson's disease detection based on running speech data from phone calls. *IEEE Trans. Biomed. Eng.* 69 (5), 1573–1584.
- Lee, J.-Y., Martin-Bastida, A., Murueta-Goyena, A., Gabilondo, I., Cuenca, N., Piccini, P., Jeon, B., 2022. Multimodal brain and retinal imaging of dopaminergic degeneration in Parkinson disease. *Nat. Rev. Neurol.* 18 (4), 203–220.
- Lelos, M.J., Murphy, E.M., Lindgren, H.S., Dunnett, S.B., Lane, E.L., 2023. Impaired cognitive and motor function are coincident with L-DOPA-induced dyskinesia in a model of Parkinson's disease. *Sci. Rep.* 13 (1), 17697.
- Li, J., Cao, X., Hou, B., Li, X., Ding, Q.-A., 2023. A deep learning hybrid model for Parkinson's disease diagnosis based on electroencephalogram signals. In: *2023 17th International Conference on Complex Medical Engineering. CME, IEEE*, pp. 55–58.
- Lin, C.-H., Wang, F.-C., Kuo, T.-Y., Huang, P.-W., Chen, S.-F., Fu, L.-C., 2022. Early detection of Parkinson's disease by neural network models. *IEEE Access* 10, 19033–19044.
- Little, M., McSharry, P., Hunter, E., Spielman, J., Ramig, L., 2008. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nat. Preced.* 1–1.
- Lv, C., Fan, L., Li, H., Ma, J., Jiang, W., Ma, X., 2024. Leveraging multimodal deep learning framework and a comprehensive audio-visual dataset to advance Parkinson's detection. *Biomed. Signal Process. Control.* 95, 106480.
- Makarios, M.B., Leonard, H.L., Vitale, D., Iwaki, H., Sargent, L., Dadu, A., Violich, I., Hutchins, E., Saffo, D., Bandres-Ciga, S., et al., 2022. Multi-modality machine learning predicting Parkinson's disease. *Npj Parkinson's Dis.* 8 (1), 35.
- Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kiebert, K., Flagg, E., Chowdhury, S., et al., 2011. The Parkinson progression marker initiative (PPMI). *Prog. Neurobiol.* 95 (4), 629–635.
- Mazilu, S., Calatrani, A., Gazit, E., Mirelman, A., Hausdorff, J.M., Tröster, G., 2015. Prediction of freezing of gait in Parkinson's from physiological wearables: an exploratory study. *IEEE J. Biomed. Health Inform.* 19 (6), 1843–1854.
- Meng, L., Pang, J., Yang, Y., Chen, L., Xu, R., Ming, D., 2023. Inertial-based gait metrics during turning improve the detection of early-stage Parkinson's disease patients. *IEEE Trans. Neural Syst. Rehabil. Eng.* 31, 1472–1482.

- Mohammed, A., Zamani, M., Bayford, R., Demosthenous, A., 2017. Toward on-demand deep brain stimulation using online Parkinson's disease prediction driven by dynamic detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 (12), 2441–2452.
- Mohanraj, P., Raman, V., Ramanathan, S., 2024. Deep learning for Parkinson's disease diagnosis: A Graph Neural Network (GNN) based classification approach with graph wavelet transform (GWT) using protein–peptide datasets. *Diagnostics* 14 (19), 2181.
- Moons, L., De Groef, L., 2022. Multimodal retinal imaging to detect and understand Alzheimer's and Parkinson's disease. *Curr. Opin. Neurobiol.* 72, 1–7.
- Morgan, C., Tonkin, E.L., Masullo, A., Jovan, F., Sikdar, A., Khaire, P., Mirmehdi, M., McConville, R., Tourte, G.J., Whone, A., et al., 2023. A multimodal dataset of real world mobility activities in Parkinson's disease. *Sci. Data* 10 (1), 918.
- Naimi, S., Bouachir, W., Bilodeau, G.-A., 2024. 1D-convolutional transformer for Parkinson disease diagnosis from gait. *Neural Comput. Appl.* 36 (4), 1947–1957.
- Narendra, N., Schuller, B., Alku, P., 2021. The detection of Parkinson's disease from speech using voice source information. *IEEE/ACM Trans. Audio Speech Lang. Process.* 29, 1925–1936.
- Navita, Mittal, P., Sharma, Y.K., Rai, A.K., Simaiya, S., Lihore, U.K., Kumar, V., 2025. Gait-based Parkinson's disease diagnosis and severity classification using force sensors and machine learning. *Sci. Rep.* 15 (1), 328.
- Özdemir, E.Y., Özyurt, F., 2025. Elasticnet-Based Vision Transformers for early detection of Parkinson's disease. *Biomed. Signal Process. Control.* 101, 107198.
- Pah, N.D., Indrawati, V., Kumar, D.K., 2023. Voice-based SVM model reliability for identifying Parkinson's disease. *IEEE Access* 11, 144296–144305.
- Pah, N.D., Indrawati, V., Prayitno, A., Kumar, D.K., 2024. The effectiveness of acoustic voice quality index to identify people with Parkinson's disease. *IEEE Access*.
- Pan, C., Tian, Y., Ma, L., Zhou, T., Ouyang, S., Li, J., 2025. CISL-PD: A deep learning framework of clinical intervention strategies for Parkinson's disease based on directional counterfactual Dual GANs. *Expert Syst. Appl.* 261, 125506.
- Parisi, F., Ferrari, G., Giuberti, M., Contin, L., Cimolin, V., Azzaro, C., Albani, G., Mauro, A., 2015. Body-sensor-network-based kinematic characterization and comparative outlook of UPDRS scoring in leg agility, sit-to-stand, and gait tasks in Parkinson's disease. *IEEE J. Biomed. Health Inform.* 19 (6), 1777–1793.
- Parisi, L., Neagu, D., Ma, R., Campean, F., 2022. Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinson's disease and COVID-19. *Expert Syst. Appl.* 187, 115892.
- Parisi, L., RaviChandran, N., Manaog, M.L., 2018. Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Syst. Appl.* 110, 182–190.
- Perez-Ibarra, J.C., Siqueira, A.A., Krebs, H.I., 2020. Identification of gait events in healthy subjects and with Parkinson's disease using inertial sensors: An adaptive unsupervised learning approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (12), 2933–2943.
- Pragadeeswaran, S., Kannimuthu, S., 2024. An adaptive intelligent polar bear (aipb) optimization-quantized contempo neural network (qcn) model for Parkinson's disease diagnosis using speech dataset. *Biomed. Signal Process. Control.* 87, 105467.
- Quan, C., Ren, K., Luo, Z., Chen, Z., Ling, Y., 2022. End-to-end deep learning approach for Parkinson's disease detection from speech signals. *Biocybern. Biomed. Eng.* 42 (2), 556–574.
- Ramesh, V., Bilal, E., 2022. Detecting motor symptom fluctuations in Parkinson's disease with generative adversarial networks. *NPJ Digit. Med.* 5 (1), 138.
- Rangel-Cascajosa, C., Luna-Perejón, F., Vicente-Díaz, S., Domínguez-Morales, M., 2025. Gait-based Parkinson's disease detection using recurrent neural networks for wearable systems. *Big Data Cogn. Comput.* 9 (7), 183.
- Raza, M., Awais, M., Singh, N., Imran, M., Hussain, S., 2020. Intelligent IoT framework for indoor healthcare monitoring of Parkinson's disease patient. *IEEE J. Sel. Areas Commun.* 39 (2), 593–602.
- Rey-Paredes, M., Pérez, C.J., Mateos-Caballero, A., 2024. Time series classification of raw voice waveforms for Parkinson's disease detection using generative adversarial network-driven data augmentation. *IEEE Open J. Comput. Soc.*
- Riasi, A., Delrobaei, M., Salari, M., 2024. A decision support system based on recurrent neural networks to predict medication dosage for patients with Parkinson's disease. *Sci. Rep.* 14 (1), 8424.
- Ricci, M., Di Lazzaro, G., Pisani, A., Mercuri, N.B., Giannini, F., Saggio, G., 2019. Assessment of motor impairments in early untreated Parkinson's disease patients: the wearable electronics impact. *IEEE J. Biomed. Health Inform.* 24 (1), 120–130.
- Roggen, D., Plotnik, M., Hausdorff, J., 2010. Daphnet freezing of gait. <http://dx.doi.org/10.24432/C56K78>, UCI Machine Learning Repository.
- Sakar, B.E., Isenkul, M.E., Sakar, C.O., Sertbas, A., Gorgen, F., Delil, S., Apaydin, H., Kursun, O., 2013. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J. Biomed. Health Inform.* 17 (4), 828–834.
- Sakar, C., Serbes, G., Gunduz, A., Nizam, H., Sakar, B., 2018. Parkinson's disease classification. <http://dx.doi.org/10.24432/C5MS4X>, UCI Machine Learning Repository.
- Sarapata, G., Dushin, Y., Morinan, G., Ong, J., Budhdeo, S., Kainz, B., O'Keeffe, J., 2023. Video-based activity recognition for automated motor assessment of Parkinson's disease. *IEEE J. Biomed. Health Inform.*
- Schalkamp, A.-K., Peall, K.J., Harrison, N.A., Sandor, C., 2023. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nature Med.* 29 (8), 2048–2056.
- Seiler, J.L., Zhuang, X., Nelson, A.B., Lerner, T.N., 2024. Dopamine across timescales and cell types: Relevance for phenotypes in Parkinson's disease progression. *Exp. Neurol.* 374, 114693.
- Shen, Y., Chen, L., Liu, J., Chen, H., Wang, C., Ding, H., Zhang, Q., 2025a. PADS-Net: GAN-based radiomics using multi-task network of denoising and segmentation for ultrasonic diagnosis of Parkinson disease. *Comput. Med. Imaging Graph.* 120, 102490.
- Shen, M., Mortezaagha, P., Rahgozar, A., 2025b. Explainable artificial intelligence to diagnose early Parkinson's disease via voice analysis. *Sci. Rep.* 15 (1), 11687.
- Shrivastava, S., Shukla, S., Khare, N., 2024. Support vector machine with eagle loss function. *Expert Syst. Appl.* 238, 122168.
- Sigcha, L., Borzi, L., Amato, F., Rechichi, I., Ramos-Romero, C., Cárdenas, A., Gascó, L., Olmo, G., 2023. Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: a systematic review. *Expert Syst. Appl.* 229, 120541.
- Sigcha, L., Borzi, L., Pavón, I., Costa, N., Costa, S., Arezes, P., López, J.M., De Arcas, G., 2022. Improvement of Performance in Freezing of Gait detection in Parkinson's disease using transformer networks and a single waist-worn triaxial accelerometer. *Eng. Appl. Artif. Intell.* 116, 105482.
- Silva, A.B.R.L., de Oliveira, R.W.G., Diógenes, G.P., de Castro Aguiar, M.F., Sallem, C.C., Lima, M.P.P., de Albuquerque Filho, L.B., de Medeiros, S.D.P., de Mendonça, L.L.P., de Santiago Filho, P.C., et al., 2023. Premotor, nonmotor and motor symptoms of Parkinson's disease: a new clinical state of the art. *Ageing Res. Rev.* 84, 101834.
- Singh, K., Khare, M., Khare, A., Kohli, N., 2025. Review on computational methods for the detection and classification of Parkinson's disease. *Comput. Biol. Med.* 187, 109767.
- Skaramagkas, V., Pentari, A., Kefalopoulou, Z., Tsiknakis, M., 2023. Multi-modal deep learning diagnosis of Parkinson's disease—A systematic review. *IEEE Trans. Neural Syst. Rehabil. Eng.* 31, 2399–2423.
- Solana-Lavalle, G., Galán-Hernández, J.-C., Rosas-Romero, R., 2020. Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybern. Biomed. Eng.* 40 (1), 505–516.
- Subasree, S., Priya, S., Brinda, S., Sakthivel, N., 2025. Early Parkinson's disease diagnosis using transition propagation graph neural network with dynamic hunting leadership optimization. *Biomed. Signal Process. Control.* 101, 107196.
- Sun, R., Hu, K., Martens, K.A.E., Hagenbuchner, M., Tsoi, A.C., Bennamoun, M., Lewis, S.J., Wang, Z., 2023. Higher order polynomial transformer for fine-grained freezing of gait detection. *IEEE Trans. Neural Netw. Learn. Syst.*
- Talitic, A., Kovalenko, E., Shcherbak, A., Anikina, A., Bril, E., Zimniakova, O., Semenov, M., Dylov, D.V., Somov, A., 2022. Comparative study of wearable sensors, video, and handwriting to detect Parkinson's disease. *IEEE Trans. Instrum. Meas.* 71, 1–10.
- Tam, W., Alajlani, M., Abd-Alrazaq, A., 2023. An exploration of wearable device features used in UK hospital Parkinson disease care: Scoping review. *J. Med. Internet Res.* 25, e42950.
- Tolosa, E., Garrido, A., Scholz, S.W., Poewe, W., 2021. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol.* 20 (5), 385–397.
- Torghabeh, F.A., Modaresnia, Y., Hosseini, S.A., 2024. An efficient tool for Parkinson's disease detection and severity grading based on time-frequency and fuzzy features of cumulative gait signals through improved LSTM networks. *Med. Nov. Technol. Devices* 22, 100297.
- Tougui, I., Zakroum, M., Karrakchou, O., Ghogho, M., 2024. Transformer-based transfer learning on self-reported voice recordings for Parkinson's disease diagnosis. *Sci. Rep.* 14 (1), 1–21.
- Tsanas, A., Little, M., McSharry, P., Ramig, L., 2009. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *Nat. Preced.* 1–1.
- 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 J. Biomed. Health Inform.* 23 (4), 1618–1630.
- Veer, K., Pahuja, S., et al., 2022. Gender based assessment of gait rhythms during dual-task in Parkinson's disease and its early detection. *Biomed. Signal Process. Control.* 72, 103346.
- Veeraragavan, S., Gopalai, A.A., Gouwanda, D., Ahmad, S.A., 2020. Parkinson's disease diagnosis and severity assessment using ground reaction forces and neural networks. *Front. Physiol.* 11, 587057.
- Vidya, B., Sasikumar, P., 2021. Gait based Parkinson's disease diagnosis and severity rating using multi-class support vector machine. *Appl. Soft Comput.* 113, 107939.
- Wang, W., Lin, J., Le, X., Li, Y., Liu, T., Pan, L., Li, M., Yao, D., Ren, P., 2024. Addressing multiple challenges in early gait freezing prediction for Parkinson's disease: A practical deep learning approach. *IEEE J. Biomed. Health Inform.*
- Xu, S., Pan, Z., 2020. A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset. *Int. J. Med. Inform.* 144, 104283.
- Xue, Z., Zhang, T., Lin, L., 2022. Progress prediction of Parkinson's disease based on graph wavelet transform and attention weighted random forest. *Expert Syst. Appl.* 203, 117483.
- Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E.S., Hausdorff, J.M., 2005. Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? *Eur. J. Neurosci.* 22 (5), 1248–1256.

- Zhang, J., 2022. Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson's disease. *Npj Parkinson's Dis.* 8 (1), 13.
- Zhang, Y., Yan, W., Yao, Y., Bint Ahmed, J., Tan, Y., Gu, D., 2020. Prediction of freezing of gait in patients with Parkinson's disease by identifying impaired gait patterns. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (3), 591–600.
- Zhang, J., Yang, H., 2024. Bounded quantile loss for robust support vector machines-based classification and regression. *Expert Syst. Appl.* 242, 122759.
- Zhang, X., Zhou, Y., Lu, Z., Zhai, D., Luo, H., Li, T., Li, Y., 2023. Multi-level graph neural network with sparsity pooling for recognizing Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.*
- Zhao, H., Cao, J., Wang, R., Lei, Y., Liao, W.-H., Cao, H., 2021. Accurate identification of Parkinson's disease by distinctive features and ensemble decision trees. *Biomed. Signal Process. Control.* 69, 102860.
- Zhao, M., Lei, H., Huang, Z., Zhang, Y., Li, Z., Liu, C.-M., Lei, B., 2022a. Attention-based graph neural network for the classification of Parkinson's disease. In: 2022 26th International Conference on Pattern Recognition. ICPR, IEEE, pp. 4608–4614.
- Zhao, Y., Liu, Y., Li, J., Wang, X., Yang, R., Lian, C., Shan, P., Wang, Y., Zhan, Z., Fu, C., 2024. Global joint information extraction convolution neural network for Parkinson's disease diagnosis. *Expert Syst. Appl.* 243, 122837.
- Zhao, A., Qi, L., Li, J., Dong, J., Yu, H., 2018. A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data. *Neurocomputing* 315, 1–8.
- Zhao, H., Wang, R., Lei, Y., Liao, W.-H., Cao, H., Cao, J., 2022b. Severity level diagnosis of Parkinson's disease by ensemble K-nearest neighbor under imbalanced data. *Expert Syst. Appl.* 189, 116113.
- Zhou, Z.D., Yi, L.X., Wang, D.Q., Lim, T.M., Tan, E.K., 2023. Role of dopamine in the pathophysiology of Parkinson's disease. *Transl. Neurodegener.* 12 (1), 44.