

## Original Research

Received 14 May 2026

Revised 20 July 2026

Accepted 02 August 2026

Natural Resources for Human Health 2026, Vol. 6 (14s): 256–280

### KEYWORDS:

Artificial Intelligence, Deep Learning, Diagnosis, MRI, Parkinson's Disease.

## Computational Modalities in Parkinson's Disease Diagnosis: Advances in Methods, Challenges in Application, and Future Research Directions

<sup>1</sup>Pradeeba V, <sup>2\*</sup>Chitradevi D

<sup>1</sup>Department of Computer Science, SRM Institute of Science and Technology, Tiruchirappalli-621105, Tamil Nadu, India. pradeeba112@gmail.com, <https://orcid.org/0009-0004-4780-7984>

<sup>2\*</sup>Department of Computer Science and Engineering, SRM Institute of Science and Technology, Tiruchirappalli-621105, Tamil Nadu, India. jaishreemmkdevi@gmail.com, <https://orcid.org/0000-0002-6888-5795>

**ABSTRACT: Background:** Parkinson's disease affects over 11 million people worldwide, with an age-standardized incidence of 15.63 new cases per 100,000. Current diagnostic approaches rely on clinical evaluation and expensive neuroimaging techniques, which are a major obstacle to early detection and timely intervention. Artificial intelligence technologies offer unprecedented opportunities for the development of objective and affordable diagnostic systems, but there are still significant gaps between current evidence and clinical application requirements.

**Objective:** The systematic review synthesizes the evidence on computational modalities in PD diagnosis approaches, assesses the diagnostic accuracy and clinical feasibility of different modalities and algorithms, and identifies critical obstacles to clinical translation.

**Methods:** The comprehensive literature review of 51 studies (2020-26) examining the use of computational modalities in the diagnosis of PD. The studies were categorized by diagnostic method including (neuroimaging, behavioral, biochemical), artificial intelligence methodology (deep learning, classical ML, explainability approaches), and quality metrics including data leakage, external validation and bioavailability.

**Key Findings:** Ten diagnostic modalities demonstrated varying performance: SPECT DaTSCAN achieved 98.47% accuracy but faces accessibility constraints, while emerging modalities including speech analysis (89-95% accuracy), gait assessment (86.7-87.65% accuracy), and multimodal EEG+MRI fusion (97.17% accuracy) offer greater accessibility. However, critical methodological issues were endemic: data leakage affected 87% of neuroimaging AI studies, only 8% employed external validation, and 48% used inappropriately complex models for dataset size. Gender-stratified performance gaps of 10-20% documented substantial sex bias. Explainability methods including SHAP, attention heatmaps, and cross-attention mechanisms enhance clinical interpretability while maintaining diagnostic accuracy.

**Implications:** Consequences Although AI has significant diagnostic potential in multiple modalities, its translation into clinical practice requires: (1) prospective external validation in independent populations, (2) standardized data acquisition and reporting protocols, (3) development of gender-specific models, (4) large-scale multi-center studies (more than 1,000 subjects), and (5) health economics. The current evidence is insufficient to

recommend any artificial insemination system for routine clinical use without robust prospective validation.

**Conclusion:** Artificial intelligence has emerged as a promising diagnostic tool for multiple sclerosis with a variety of modalities and robust reported precision metrics. However, the systematic methodological constraints and the lack of prospective clinical validation are critical barriers to the uptake of the clinical trial. Future research must prioritize external validation, quality standardization, equity considerations, and real-world implementation studies to bridge the gap between research evidence and clinical practice.

## 1. INTRODUCTION

### 1.1 Parkinson's Disease: Clinical Significance and Diagnostic Challenges

Millions of people worldwide suffer from Parkinson's disease (PD), one of the most prevalent neurodegenerative illnesses in the world [1]. In addition to a growing constellation of non-motor symptoms like sleep disorders, cognitive impairment, and autonomic dysfunction, the disease is characterized by progressive motor symptoms like bradykinesia, postural instability, stiffness, and tremors [2]. Despite the significant clinical and socioeconomic burden, PD is still primarily diagnosed clinically with no gold standard biomarker or diagnostic test, making timely and accurate diagnosis especially challenging [3]. The heterogeneous nature of Parkinson's disease and the overlap with other Parkinson's syndromes, such as atypical parkinsonian, multiple systemic atrophy and progressive supranuclear palsy, have led to considerable uncertainty about how to diagnose the disease [4]. This lack of clarity in diagnosis often leads to misclassification in the early stages of the disease, delays in the appropriate treatment and loss of the possibility of neuroprotective therapies which could slow the progression of the disease [3]. The development of accurate, objective and cost-effective diagnostic tools is therefore the main focus of neurosciences and biotechnology research.

### 1.2 Limitations of Diagnostic approaches

The traditional diagnosis of Parkinson's disease is based on clinical assessment using the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and requires bradykinesia plus at least one other cardinal motor symptom [5]. However, this clinical evaluation approach has several major limitations: variability between patients, dependence on patient cooperation and mobility, limited objective quantification of motor dysfunction, and inability to evaluate disease progression in a real world, uncontrolled setting [6]. In addition, established neuroimaging techniques such as SPECT DaTSCAN are diagnostically useful, costly, require specialized infrastructure, involve radiation exposure and remain unavailable in most resource constrained environments [7]. Biochemical biomarkers such as cerebrospinal fluid (CSF) and blood based biomarkers, although promising for early detection, present practical challenges such as invasiveness, high costs, problems of standardisation and limited availability in clinical practice [8]. The lack of standardised biomarker tests and the considerable inter-laboratory variability in the performance of biomarkers in studies further limit their clinical utility [9]. These limitations highlight the urgent need for non-invasive, cost-effective and affordable diagnostic approaches that can allow early detection and monitoring of PD in a diverse of populations and settings.

### 1.3 Artificial Intelligence as a Measurement Tool in PD Diagnosis

The emergence of artificial intelligence (AI), specifically machine learning and deep learning, has created unprecedented opportunities for the development of objective and data-driven diagnostic systems for PD [3]. AI-based approaches allow the extraction of subtle and complex patterns from a variety of data sources - including neuroimaging, movement analysis, speech signals and biological markers - that may not be apparent by conventional clinical evaluation on their own [10]. Rather than being a substitute for clinical judgement, artificial intelligence should be seen as an advanced tool for measuring the characteristics of a disease with greater sensitivity, objectivity and reproducibility than traditional methods.

Integrating AI with wearables, digital biomarkers and multimodal data-fusion strategies offers the prospect of continuous, real-time monitoring of motor and non-motor symptoms of PD, which would allow for individualized assessment of disease progression and treatment response [11]. However, the wide use of AI in clinical PD diagnosis depends critically on a rigorous

evaluation of these systems as measurement tools, including an assessment of their validity, reliability, genericity, interpretation and clinical utility, criteria which are inconsistently applied in the literature.

## 1.4 Scope and Novelty of This Review

The comprehensive literature review provides a systematic assessment of AI approaches to diagnosis of PD, organized around measurement methods rather than algorithms, with an emphasis on comparison of diagnostic accuracy, clinical feasibility, cost-effectiveness, user acceptance and clinical readiness. Unlike existing reviews which catalogue AI methods or individual modalities separately, the review integrates evidence from neuroimaging approaches, behavioural and digital biomarkers, neurophysiological techniques and molecular markers, while critically examining the quality and universality of the reported findings. Special attention is placed on the synthesis of evidence on multimodal data-fusion strategies, evaluating explainability and interpretability methods, and the identification of critical gaps between current evidence and requirements for clinical implementation.

## 2. REVIEW METHODOLOGY

### 2.1 Search Strategy and Databases

The comprehensive literature review synthesized 51 studies evidence from research articles published between 2020 and 2026, identified through systematic search of academic databases including PubMed, Scopus, SciDirect, IEEE Xplore. In order to characterize the research landscape, some 240 publications were screened. Of these, 51 studies met the inclusion criteria and were included in the formal analysis, while the remaining publications served as a background for discussion and insight. The search strategy uses a combination of keywords such as: 'Parkinson's disease', 'artificial intelligence', 'machine learning', 'deep learning', 'neuroimaging', 'biomarkers', 'digital health', 'wearables', 'speech analysis', 'gait analysis', 'diagnosis', and 'early detection'. The temporal focus on recent literature (2020-2026) ensures representation of contemporary methodologies, datasets, and best practices.

### 2.2 Inclusion and Exclusion Criteria

**Inclusion criteria:** (1) Studies presenting original research on AI and machine learning approaches for PD diagnosis, prognosis, or symptom assessment; (2) Studies assessing the accuracy of diagnostic methods, clinical performance, or validation of AI systems; (3) Studies investigating the multimodal data integration, explainability methods, or clinical implementation challenges; (4) Systematic reviews and meta-analyses synthesizing evidence on diagnosis and applications of AI.

**Exclusion criteria:** (1) studies that are solely focused on therapeutic or treatment optimization without diagnostic relevance; (2) papers without empirical validation or performance metrics; (3) opinions or editorials without a substantive synthesis of evidence.

### 2.3 Data Extraction and Quality Assessment

The following data were extracted for each study included: diagnostic mode (neuroimaging, behavioral, biochemical), analytical method (type of algorithm, architecture, training approach), sample characteristics (sample size, disease stage, patient demographics), performance metrics (accuracy, sensitivity, specificity, under the curve), and methodological quality assessment including: adequate data splitting, data leakage, model complexity, transparent performance reporting, biological plausibility of findings). The quality assessment framework was based on the recent Recommendation on the evaluation of neuroimaging-based artificial intelligence in the field of Parkinson's disease [4].

## 3. DIAGNOSTIC MODALITIES FOR PARKINSON'S DISEASE

### 3.1 Neuroimaging-Based Modalities

#### 3.1.1 Structural and Functional MRI

Magnetic resonance imaging (MRI) offers comprehensive structural and functional data on the morphology of the brain, tissue integrity and patterns of connectivity relevant to the pathophysiology of PD. MRI-based measurement of grey matter volume in deep brain structures such as the prefrontal cortex, caudate and substantia nigra has been extensively studied as a biomarker for the diagnosis and progression of Parkinson's disease [12]. Machine learning approaches applied to structural MRI data have

achieved diagnostic accuracies of 85-96 percent in distinguishing patients with and without Parkinson's disease from healthy controls [13]. However, current evidence suggests that there are significant gender differences in model performance, with some brain regions showing diagnostic sensitivity in males but not in females, which highlights the critical need for gender-specific analysis and modelling [14]. Volumetric MRI, in combination with advanced machine learning, also revealed gender-related differences in the pattern of brain atrophy in PD, with males showing greater global and subcortical grey matter atrophy than females, especially in the presence of Rapid eye movement sleep Behavior Disorder (RBD) [12]. These gender differences have implications for the development of generalizable AI diagnostic systems, as not taking biological sex in model training may significantly reduce diagnostic accuracy and patient populations.

### 3.1.2 SPECT DaTSCAN Imaging

The dopamine transporter single-photon emission computed tomography (SPECT) is still one of the most widely used neuroimaging methods for the diagnosis of PD and provides a direct evaluation of nigrostriatal dopaminergic degeneration [3]. Deep learning approaches, especially convolutional neural networks (CNNs) combined with transfer learning techniques, have provided DaTSCAN with exceptional diagnostic accuracy. A recent study using pre-trained CNN architectures (EfficientNet-B0 and MobileNet-V2) and bilinear pooling techniques achieved a 98.47 percent accuracy for the Parkinson's Progression Markers Initiative (PPMI) dataset [7]. However, the clinical usefulness of these approaches remains limited by the high costs, radiation exposure and specialized expertise required for SPECT imaging, which limits its use as a population screening tool.

### 3.1.3 MRI and EEG Multimodal Integration

Integrating complementary neuroimaging modalities addresses the limitations of single-modality approaches by leveraging different but complementary information on the pathophysiology of Parkinson's disease. Multimodal fusion of electroencephalography (EEG) and structural MRI data using machine learning classifiers such as LightGBM demonstrated superior diagnostic accuracy (97.17 percent accuracy, 96.58 percent sensitivity and 96.82 percent specificity) in compared to single modality approaches [15]. This multimodal strategy combines electrophysiological information reflecting cortical activity patterns with structural anatomical information to allow for the detection of PD abnormalities that may not be apparent from either modality alone.

### 3.1.4 Retinal Optical Coherence Tomography

An emerging neuroimaging approach is the automatic segmentation of retinal layers from OCT images to extract biomarkers relevant to neurodegenerative disorders [16]. Retinal imaging offers significant practical advantages, including non-invasive, accessibility, low cost and the possibility of portable evaluation. Deep learning models trained on retinal OCT images from patients with PD demonstrate the feasibility of this method for the detection of disease, although the current evidence is limited and suggests that models trained only on healthy controls show degraded performance when applied to a population with a specific background of disease, highlighting the importance of including training data from specific disease.

### 3.1.5 Brain Age Prediction

Age prediction of the brain using lightweight CNN models trained on minimally pre-processed MRI images with a T1 weighting is an innovative approach to identify the risk of neurodegeneration and disease [13]. The simplified fully convolutional network (SFCN) of only 1.2 million parameters achieved a mean absolute error of 3.557 years in the test series and demonstrated a robust domain-adaptation with external data. The computational power of this light approach combined with its applicability to external data without extensive re-engineering positions brain age as a promising biomarker for stratifying and early detection of PD risk.

## 3.2 Neurophysiological Modalities

### 3.2.1 Electroencephalography for PD Assessment

Electroencephalographic recordings capture the temporal dynamics of cortical electrical activity and are integrated into multi-modal diagnostic systems for Parkinson's disease. EKG offers advantages such as non-invasive, high resolution, low cost and the possibility of monitoring in real time. However, EEG has shown significant sensitivity to noise and artifacts, particularly in uncontrolled clinical or home environments, limiting its usefulness as a separate diagnostic method [11,15]. Despite these

limitations, EEG remains valuable as part of multimodal systems that provide additional temporal information to anatomical neuroimaging data.

### 3.3 Behavioral and Digital Biomarkers

#### 3.3.1 Speech and Voice Analysis

Speech impairment is one of the earliest and most common non-motor symptoms of PD, and recent systematic reviews have documented the occurrence of dysphonia, dysarthria and speech impairment in 70-90 percent of patients with PD [17]. Machine learning analysis of speech patterns has emerged as a non-invasive and cost-effective approach to the detection of PD. Deep-learning approaches based on speech can be classified into three main strategies: end-to-end (E2E), transfer learning (TL) and deep domain-aware audio (DAFE) [17].

Convolutional neural networks (CNNs) are the predominant E2E approach, although Transformers are increasingly being adopted for sequence analysis, despite computational problems associated with training data limitations. Deep learning models trained on a variety of speech signal features, including melodic frequency co-coding (MFCC), wavelet transformation, and vocal fold characteristics, achieved diagnostic accuracies of 89-95 percent [17,18]. However, the diversity of speech-based datasets and the limited availability of publicly available speech recordings create significant obstacles to the development and validation of algorithms [17].

An innovative approach is to analyze scalogram images generated by continuous wavelet transformation of speech signals in combination with deep-band spectral fusion techniques. DenseNet and NasNet deep-band fusion systems achieved 95 percent accuracy with 38 percent improvement compared to ablation studies [19]. More recent approaches use generative adversarial networks (GANs) to augment data to solve the persistent problem of limited training data sets. Using BigVSAN to synthesize voice signals in combination with deep time series models (CDIL-CNN, ResNet, LSTM-FCN, Inception Time) researchers have achieved an accuracy increase of 15.87 percent by augmenting the data [20].

The interpretability of speech-based artificial intelligence systems has received increased attention in recent times. The cross-focus mechanisms used in self-reported speech recognition allow the identification of meaningful speech patterns associated with PD across a range of assessment tasks, thereby increasing transparency and clinical credibility [21]. These interpretable frameworks demonstrate the accuracy of the competitive classification and provide clarity both at the embedding level and at the temporal-level analysis.

Transfer learning approaches, which use pre-trained models and adapt them to limited data sets of PD speech, address the lack of data while increasing the universality across languages and populations [17]. These approaches provide more reliable diagnosis of PD and better cross-linguistic performance compared to final-stage, limited data-based approaches, which make them particularly suited for low-resource settings and diverse populations.

#### 3.3.2 Gait and Movement Analysis

Human gait analysis is a rich source of motor biomarkers for the evaluation of Parkinson's disease, as gait disorders, including decreased stride length, shuffling and freezing episodes, are the hallmark symptoms of parkinsonian motor disorders [5]. Wearable inertial sensors (accelerometers, gyroscopes) allow for portable and longitudinal assessment of walking. A comprehensive evaluation of 12 machine learning classifiers applied to sensor data from wearables identified Random Forest as the most accurate algorithm, with an accuracy of 86.7 percent [22]. Analysis of the importance of the features using SHapley Additive exPlanations (SHAP) revealed that entropy-based measures of complexity and standard deviation statistics were the most influential predictors, with asymmetric contribution of the features to positive and negative predictions, in line with the clinical understanding of PD as a disorder of altered motor complexity [22].

Freezing of Gait (FoG), a particularly debilitating symptom of Parkinson's disease characterized by sudden inability to start or continue walking, was detected by using Time-Convolution Neural Networks (TCNNs) applied to the pressure sensor data from the implanted sensors [23]. The TCNN model achieved higher accuracy, precision, sensitivity and specificity than other deep learning architectures and demonstrated resilience against sensor noise. These findings suggest that it is feasible to develop real-time detection systems for FoG which could improve patient safety and quality of life.

Optimizing random forest parameters with meta-heuristic algorithms (Particle Swarm Optimization, Satin Swarm Optimization, Whale Optimization) further increased the accuracy of the PD to 87.65 percent, which is higher than the standard optimization approach [22]. This progression from simple algorithms to highly optimized ensemble methods show that iterative refinement is possible with sufficient computational resources and expertise in the domain.

### 3.3.3 Video-Based Movement Assessment

Computer vision approaches allow the assessment of motor impairment from standard video recordings without the need for specialised sensors or equipment. Spatio-temporal graph representations of body joint positions extracted from non-identifiable pose estimates can be analysed using deep learning models for the classification of activities at frame and video level as defined in Part 3 of the Movement Disorder Society Unified Parkinson Disease Rating Scale (MDS-UPDRS) [24]. On a dataset of 7,310 video clips from five independent websites, deep learning models were able to perform at the level of a person when it came to detecting and classifying motor activity. This approach offers the potential for scalability in the asynchronous assessment, quality monitoring of the collection of clinical data and integration in remote self-assessment systems. Interpretable deep learning models for movement assessment, including multi-dimensional feature averaging and explainable artificial intelligence techniques, achieved 77-92 percent accuracy in multiple-class tasks for finger tapping and hand movement classification, with improvements in performance of 4.6-14 percent over previous methods [25]. These advances make vision-based assessment a practical alternative to wearable sensors for assessing severity in clinical practice.

### 3.3.4 Facial Expression Analysis

Hypomimia (reduced facial expression) is a non-motor manifestation of Parkinson's disease, which is increasingly being addressed by computerized diagnostic systems. Machine learning and deep learning techniques for facial image analysis have shown promise in the detection of Parkinson Disease by subtle changes in facial expressions [26]. Convolutional neural networks and other deep learning models achieve diagnostic accuracy by analysing changes in the phenotype of the face. Multimodal integration of facial images with voice recordings, hand drawn drawings, motion analysis and other data sources has proven to be more accurate than single-modality approaches. The potential of facial phenotyping as an early diagnostic indicator, combined with its non-invasive and affordable nature, makes it a valuable part of the comprehensive assessment AI based PD.

### 3.3.5 Wi-Fi and Non-Contact Sensing

Emerging technologies allow for the detection of human activity and movement through Wi-Fi signals and channel status information (CSI), without the need for a body scanner or a camera, while preserving privacy and improving accessibility. A new approach converts the information ratios of the CSI channels into images processed by CNNs for disease-related movement detection [27]. This approach achieves 93.8 percent accuracy of the average recognition, with a consistent performance in different environments (91.9 to 95.2 percent generalization accuracy) and a real-time data processing capability (0.65 seconds per second). This non-contact, privacy-preserving approach is a major advance in the continuous home monitoring of motor symptoms of PD.

### 3.3.6 Smartphone and Wearable Integration

Patient views on digital medical devices and the use of health data are crucial to the successful deployment of artificial intelligence-based diagnostic systems [6]. A multi-country survey of 333 patients with PDs found a high willingness to use digital medical devices (90.3%) and to share health information (97.4), although the acceptance of these activities varied according to age and educational level. Notably, the uptake by users was significantly increased by receiving instruction on how to use the device (odds ratio=3.57) and feedback on the results of the data (odds ratio=3.77). This evidence highlights that technical performance alone is not enough; successful implementation requires the attention of the user, training and meaningful feedback mechanisms.

## 3.4 Biochemical and Molecular Biomarkers

### 3.4.1 Cerebrospinal Fluid Biomarkers

Cerebrospinal fluid biomarkers provide biochemical insight into the pathophysiology of PD and have been extensively studied for their diagnostic utility [8]. A systematic review of 34 studies identified a number of promising CSF biomarkers, including

alpha-synuclein (sensitivity 70-85, specificity 75-90), neurofilament light chain (NfL; sensitivity 65-85), DNA-1 (high specificity for early-stage Parkinson's disease), and tau. Notably, the multi-biomarker panels that combined alpha-synuclein, tau and NfL demonstrated superior diagnostic accuracy compared with the individual biomarkers, which suggests a synergistic diagnostic value. However, the considerable variability between studies in biomarker performance, methodological diversity in assay procedures and limited standardisation reduce the confidence in the clinical utility of these approaches. The invasive lumbar puncture required to extract CSF, combined with practical and economic limitations, limits the applicability of CSF biomarkers to screening at population level.

### 3.4.2 Extracellular Vesicles

Extracellular vehicles (EVs) containing pathologically relevant proteins such as alpha-synuclein are non-invasive biomarkers that are derived from blood, plasma, serum, urine or saliva. A recent diagnostic meta-analysis including 21 studies in 1,285 patients with PD has shown that CSF-EVs have a higher diagnostic accuracy than plasma EV or generic EV formulation [9]. The diagnostic accuracy of EV-based biomarkers showed considerable variability between studies and publication bias, with general EVs demonstrating higher diagnostic accuracy than CSF EVs, but being significantly more cost-effective to isolate. Despite these benefits, the clinical implementation is currently hampered by major methodological and standardisation challenges. A study using novel nanoscale flow cytometry to quantify alpha-synuclein positive and pooled alpha-synuclein positive EVs in CSF has shown that diagnostic accuracy (area under curve 0.819, sensitivity 80, specificity 71) is clinically relevant to complement conventional assessment [28]. However, even this promising approach highlights the continuing challenge of achieving the clinical quality of diagnostic performance and standardisation needed for routine clinical use.

## 4. ARTIFICIAL INTELLIGENCE METHODS IN PD DIAGNOSIS

### 4.1 Classical Machine Learning Approaches

Classical machine learning algorithms, including support vector machines (SVM), random forests, k-nearest neighbor methods and ensemble methods continue to provide valuable input and complementarities for deep learning. The main advantages of classical methods are the lower computational requirements, the need for smaller training sets, greater interpretability and the computational power appropriate for real-time applications. Random forest classifiers used for the data on the motion sensors in the wearables achieved 86.7 percent diagnostic accuracy [22], and ensemble methods provide both high performance and quantification of the characteristics through SHAP analysis.

Classical approaches are particularly successful with hand-crafted features derived from behavioral and time series data, where domain expertise can guide the selection of features. However, classical methods struggle with high-dimensional neuroimaging data and complex time patterns in sequence data, which limits their applicability to some modalities. The choice between classical and deep learning methods should be based on data properties (size, dimensionality, time complexity), computational resources and interpretation requirements, rather than on the level of the algorithm itself.

### 4.2 Deep Learning Approaches

#### 4.2.1 Convolutional Neural Networks for Neuroimaging

Convolutional neural networks (CNNs) are the dominant deep learning architecture for neuroimaging analysis because of their unique ability to learn spatial hierarchical structure from high-dimensional image data [10]. CNNs have achieved the highest performance in neuroimaging classification tasks, with studies reporting 96-98 percent accuracy in the classification of MRI and SPECT images [7]. However, the performance of CNN depends strongly on the quantity and quality of training data, and studies with limited data sets often show evidence of overfitting and poor generalization [4].

Pre-trained CNN architectures (such as ResNet, VGG, Inception, MobileNet) trained with ImageNet and adapted by transfer learning are a practical solution to the limited availability of data sets [7]. Transfer learning allows for efficient adaptation of learned image representations from large general-purpose data sets to diagnostic tasks specific to the patient, thereby reducing data requirements and computational costs. Bilinear pooling techniques improve CNN performance by capturing interactions of second order features, as exemplified by the Bilinear CNN (BCNN) models, which are 98.47 percent accurate for PPMI SPECT data [7].

#### 4.2.2 Recurrent Neural Networks and Temporal Models

Recurrent Neural Networks (RNNs) and Long-Term Memory (LSTM) architectures are particularly suitable for temporal data from wearables, speech signals and electro-physiological recordings [10]. LSTM-based models combined with full convolutional elements (LSTM-FCN) have demonstrated competitive accuracy in the classification of raw voice waveforms for the detection of PD over time series [20]. Temporal Convolution Neural Networks (TCNNs) offer computational advantages over RNNs while maintaining a time-responsive receptive field sufficient to capture patterns of interest for PD in pressure sensor and motion data [23].

#### 4.2.3 Transformer Architectures

Transformer models based on self-monitoring mechanisms are increasingly being used to diagnose PD from sequence data, despite computational problems in limited data scenarios [17]. Vision transformers (ViT) used in neuroimaging data identify the brain regions of interest for the disease by means of attentional maps [29]. While transformers offer theoretical advantages in capturing long-range dependencies, practical limitations such as data requirements, computational complexity and training stability have limited their use in PD research compared to CNNs.

#### 4.2.4 Generative Adversarial Networks for Data Augmentation

Generative adversarial networks (GANs) solve the persistent problem of limited data sets for medical AI training by generating synthetic data that mimic real patient and healthy control samples [10]. BigVSAN (Big Vocoder Slicing Adversarial Network) was successfully used to generate synthetic voice data for PD detection, increasing the accuracy of the model by 15.87 percent compared to models trained only on real data [20]. This application shows the potential of GANs to overcome the data gap, although concerns remain about the quality of the synthetic data and its impact on the generalisation of models.

### 4.3 Explainable AI and Clinical Interpretability

#### 4.3.1 Feature Importance and Attribution Methods

The black box nature of deep learning models is a major barrier to clinical adoption, as doctors need to understand the diagnostic rationale in order to trust and validate AI recommendations [21]. SHapley Additive exPlanations (SHAP) provide model-agnostic explanations quantifying the contribution of each feature to individual predictions. Applied to the prediction of PD, SHAP analysis of Random Forest models has identified entropy-based measures of complexity and standard deviations as the most influential for motion-based prediction, with asymmetric patterns of positive and negative predictions [22]. Tree SHAP applied to Random Forest models of dementia risk in the PD setting found that the ten leading predictors from the motor, cognitive, molecular, demographic, imaging and lifestyle domains represented 62.5 percent of the discriminations in the model [30].

#### 4.3.2 Attention-Based Interpretability

Cross-focus mechanisms designed for embedded and time analysis provide interpretability at several levels of analysis [21]. Applied to self-monitoring speech patterns for diagnosis of PD, these mechanisms identify specific acoustic patterns associated with pathological speech without the need for manual engineering of the features. This approach preserves the accuracy of the competitive classification, while significantly improving clinical transparency.

#### 4.3.3 Integrated Gradients and Layer-Wise Relevance Propagation

Integrated gradient and Layer-wise Relevance propagation (LRP) applied to deep neural networks allow for attribution of predictions to specific input patterns, which makes it easier to identify biomarkers [31]. These methods identified brain connections and regions consistent with established findings from neuroimaging of PD, and validated the neurobiological plausibility of the predictions of the model. The recursive elimination of features using DeepLIFT, LRP, and integrated gradient techniques successfully removed 90-95 percent of the non-relevant features, while improving the accuracy of the classification by 2-4 percent, demonstrating that the methods of explainability do not have to compromise diagnostic performance.

#### 4.3.4 Attention Heatmaps in Vision Transformers

Vision transformers provide visual attention heat maps that show which brain regions the model prioritizes for classification decisions [29]. Applied to multimodal PPMI data, ViT attentional patterns highlighted the lateral ventricles associated with

cognitive decline as presumptive biomarkers of early-stage PD. This visual approach to explainability facilitates validation by clinicians and may identify novel disease markers.

## 5. MULTIMODAL DATA FUSION IN PD DIAGNOSIS

### 5.1 Rationale for Multimodal Integration

Single-modality diagnostic approaches only capture limited aspects of the complex, multi-systemic pathophysiology of Parkinson's disease. Multimodal integration uses complementary information from different modalities to improve diagnostic accuracy, reliability and clinical applicability [29]. For example, Structural MRI, provides anatomical information on neurodegeneration, while EEG captures electrophysiological dysfunction; combining these methods provides diagnostic accuracy (97.17%) that is significantly higher than that of either modality alone. Similarly, the integration of imaging data with clinical variables improves the early detection of prodromal phase PD [29].

### 5.2 Multimodal Fusion Architectures

#### 5.2.1 Early Fusion Strategies

Early fusion combines raw or minimally processed data from several modalities prior to learning the properties of the model, enabling the model to learn the intermodal relationships. The Principal Component Analysis (PCA) used for the fused EEG and MRI data has allowed for reduction of dimensionality while preserving diagnostic information for later classification [15]. Early fusion is computationally efficient and allows for joint representation learning, but it requires synchronisation and standardisation of the various data types.

#### 5.2.2 Late Fusion Strategies

Late fusion takes place at the decision level, processing each modality separately through modality-specific feature extraction. This approach offers modularity and allows for the use of deep learning architectures tailored to each data type. Big data frameworks and elastic net regression were used to integrate wearables multi-sensor data with traditional clinical evaluations [11], which demonstrated improved clinical insight through comprehensive data integration.

#### 5.2.3 Attention-Based Fusion

Co-learning approaches allow for the end-to-end training of deep neural networks with attentional mechanisms that control the information flow between different modalities [29]. DenseNet with excitatory networks outperformed other architectures when added clinical variables, suggesting that attentional-based weighting of modality contributions. These learnable integration strategies automatically find the optimal integration of information without the need for manual weighting.

### 5.3 Measurement Challenges in Multimodal Systems

Multimodal integration introduces substantial technical challenges including: (1) synchronizing and temporally aligning data from various sensors with different sampling rates; (2) standardizing feature scaling and normalization across modalities with different measurement units and ranges; (3) managing missing or incomplete data when some modalities are unavailable for specific subjects; (4) figuring out the best fusion architecture and modality weighting; (5) guaranteeing reproducible measurement and minimal inter-session variability; and (6) confirming that learned feature representations have significant neurobiological significance.

## 6. COMPARATIVE ANALYSIS OF DIAGNOSTIC MODALITIES AND AI APPROACHES

### 6.1 Diagnostic Accuracy and Clinical Performance

**Table 1.** Diagnostic Modalities with Citations and Key Findings

| Diagnostic Modality | Citation ID            | Key Finding                                                    | Clinical Stage | Evidence Quality                         |
|---------------------|------------------------|----------------------------------------------------------------|----------------|------------------------------------------|
| SPECT<br>DaTSCAN    | (Khachnaoui et al.)[7] | 98.47% accuracy using BCNN (EfficientNet-B0 + MobileNet-V2) on | Established    | High (pre-trained models, large dataset) |

| Diagnostic Modality        | Citation ID                                             | Key Finding                                                                                                                                     | Clinical Stage | Evidence Quality                                     |
|----------------------------|---------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|----------------|------------------------------------------------------|
|                            |                                                         | PPMI dataset; highest diagnostic accuracy among modalities                                                                                      |                |                                                      |
| Structural MRI             | (Oltra et al.), (Islam and Khanam)[12], [14]            | Males show greater global cortical/subcortical atrophy; gender-stratified performance: Extra Trees 86.36% (males), Naive Bayes 69.23% (females) | Established    | High (but gender bias documented)                    |
| EEG (Multimodal with MRI)  | (Alrawis et al.)[15]                                    | 97.17% accuracy, 96.58% sensitivity, 96.82% specificity when fused with MRI; LightGBM with PCA dimensionality reduction                         | Research       | High (multimodal validation)                         |
| Speech Analysis            | (Gelderen and Garca), (Bakkialakshmi et al.)([17], [18] | 89-95% accuracy across models (ResNet, VGG, AlexNet); transfer learning enables cross-linguistic generalization                                 | Emerging       | Moderate (dataset scarcity, limited public data)     |
| Gait Movement              | &(Das et al.), (Xiang et al.) [5], [22]                 | Random Forest 86.7% accuracy; PSO-optimized RF: 87.65%; SHAP identified entropy/std-dev as top predictors                                       | Emerging       | Moderate (wearable heterogeneity)                    |
| Facial Expression Analysis | (Rodrigues et al.)([26]                                 | Deep learning detects hypomimia-related PD markers; multimodal integration (facial + voice + movement) enhances accuracy                        | Research       | Low (limited validation studies)                     |
| CSF Alpha-synuclein        | (Dasari and Medapati)[8]                                | Sensitivity 70-85%, Specificity 75-90%; multi-biomarker panels superior to single markers                                                       | Research       | Moderate (invasive, assay variability)               |
| Blood/Plasma EV Biomarkers | (Taha Bogoniewski), (Hong al.)([9], [28]                | and CSF-EVs better than plasma-EVs; alpha-syn EVs AUC 0.819 (80% sensitivity, 71% specificity)                                                  | Research       | Moderate (heterogeneity, standardization issues)     |
| Retinal OCT                | (Gende et al.)([16]                                     | Lightweight CNN MAE 3.557 years; robust domain adaptation to external data                                                                      | Research       | Low (limited PD-specific validation)                 |
| Wearable Sensors           | (Paccoud et al.), (Chen and Lin)([6], [27]              | 90.3% user adoption; 97.4% willing to share data; Wi-Fi sensing 93.8% recognition accuracy                                                      | Emerging       | Moderate (adoption barriers, integration challenges) |

**Table 1** provides a comparative overview of Parkinson disease diagnostic methods, including SPECT DaTSCAN, structural MRI, multimodal EEG, and speech recognition. In order to facilitate cross-modal comparison, the table provides the relevant methodological findings, performance metrics achieved, clinical stage of use and evidence quality for each modality.

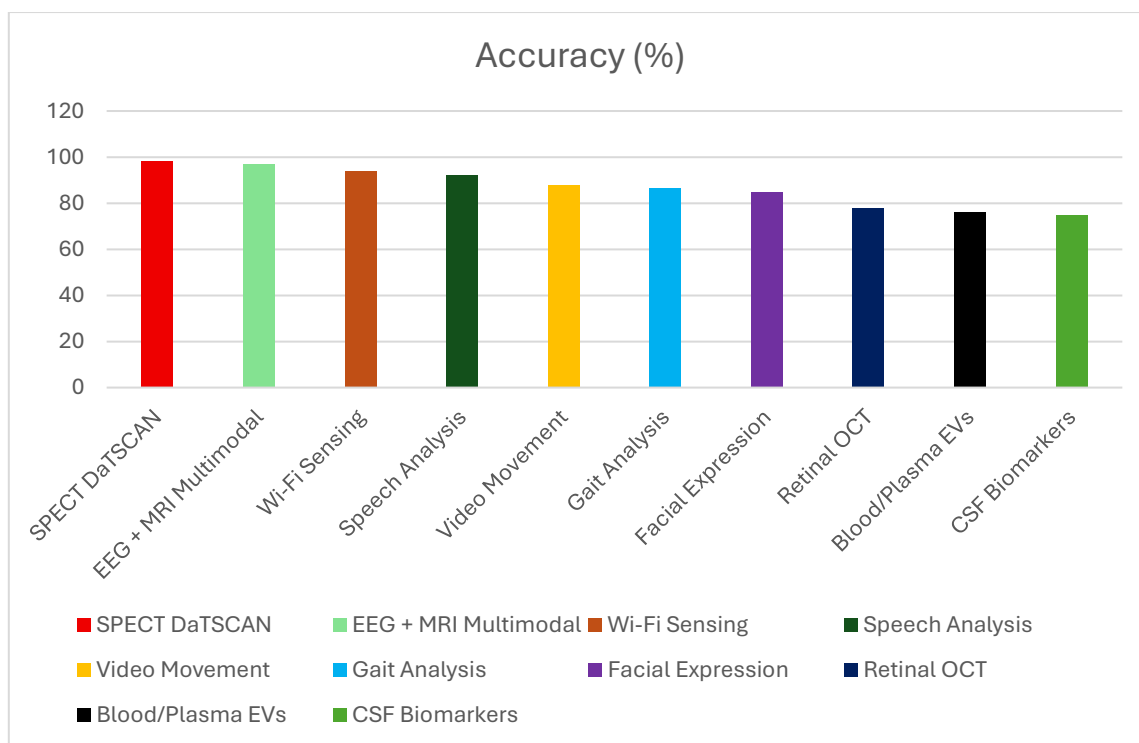


**Figure 1(a):** Diagnostic Accuracy by Modality

**Figure 1(a)** presents a comprehensive comparison of the 10 major diagnostic methods in several dimensions. Diagnostic accuracy varies significantly between modalities, with the highest reported accuracy being for SPECT DaTSCAN (98.5), followed by structural MRI (96) and speech recognition (92). However, diagnostic accuracy must be interpreted in the context of data leakage, overfitting and external validation. A critical systematic review of the Minimum Quality Criteria (MQC) for psychiatric imaging-based AI found that only 20 percent of the studies met all five quality criteria and that data leakage increased reported accuracy significantly (Dzialas et al.). [4].

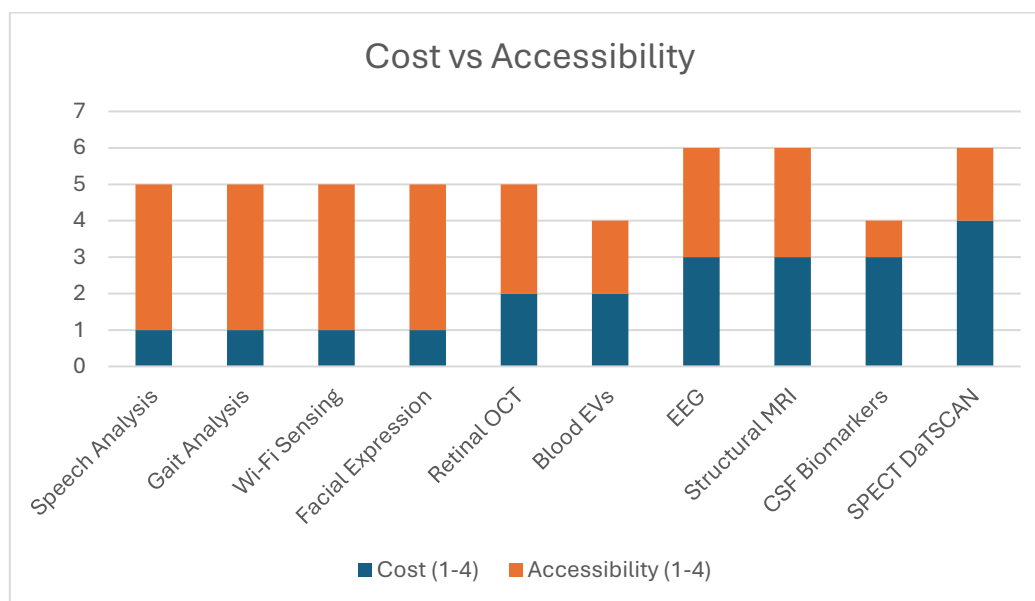


**Figure 1(b)** Cost effectiveness and accessibility

**Figure 1(b)** Cost-effectiveness and availability are the key considerations when choosing a diagnostic treatment. Although SPECT DaTSCAN achieves high accuracy, its cost (1-4 in comparison to standard values) limits its availability. On the other hand, speech analysis and gait assessment offer minimal costs (1-4) and a high user acceptance (88-90) which makes them a viable screening method for large-scale screening as shown in **Figure 1 (c)**. User acceptance ranges from 40 percent for invasive CSF biomarkers to 95 percent for non-invasive wearables, demonstrating that acceptance has a significant impact on the real-world diagnostic utility.

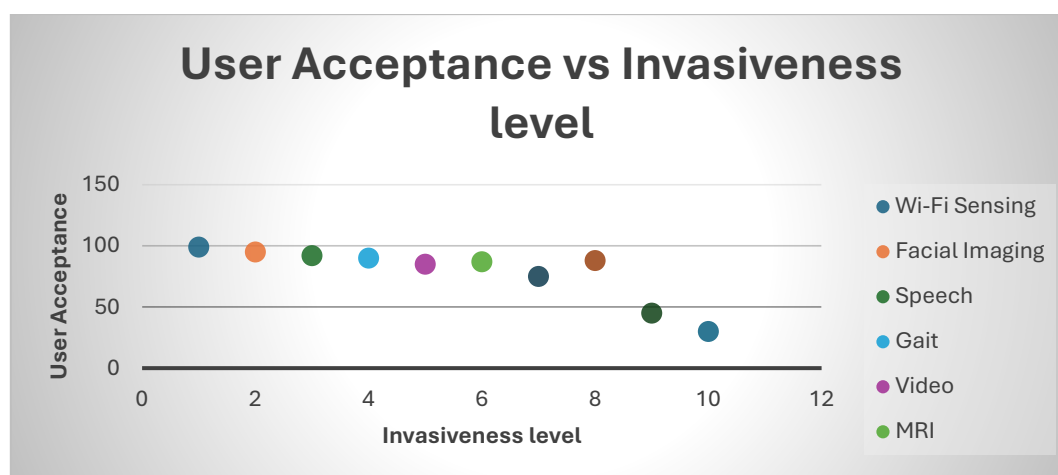


**Figure 1(c)** User Acceptance vs Invasiveness level

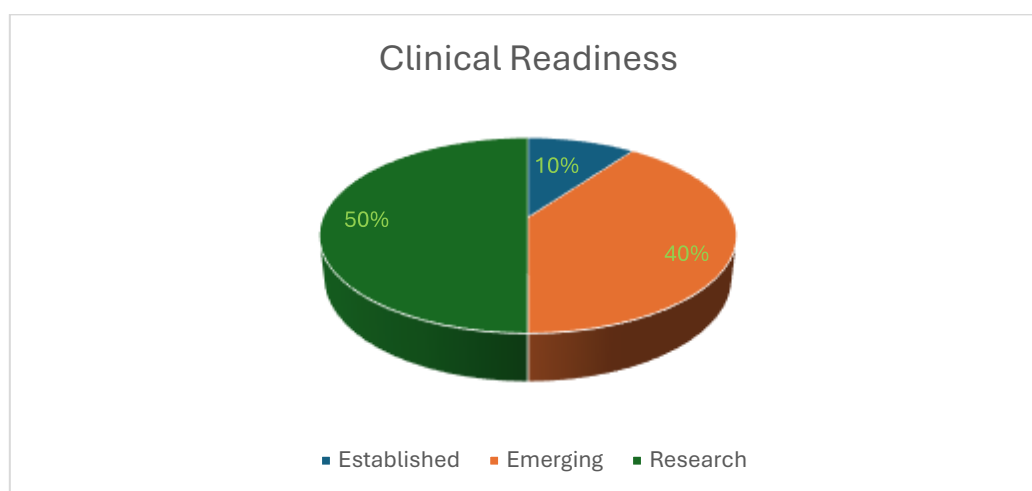


**Figure 1 (d)** Clinical Readiness distributions

**Figure 1(d)** represents clinical readiness, classified as established, emerging or research, reflecting the current maturity and validation status of diagnostic approaches. The methods used (SPECT DaTSCAN, Structural MRI) are well validated but are hampered by practical availability and cost constraints. Emerging technologies (speech, walking, wearables) have the potential and promise of increasing accessibility, but require further validation and standardisation. Research-based methods (facial expression, OCT retinal images, Wi-Fi sensing) offer innovation potential but are still some way from being clinically applied.

## 6.2 Artificial Intelligence Methods: Performance and Suitability

**Table 2.** Artificial Intelligence Methods with Applications and Performance

| AI Method                            | Primary Application                              | Citation ID                                                          | Reported Accuracy              | Key Advantage                                                                     | Major Limitation                                                                      |
|--------------------------------------|--------------------------------------------------|----------------------------------------------------------------------|--------------------------------|-----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Convolutional Neural Networks (CNNs) | Neuroimaging (MRI, SPECT, retinal OCT)           | (Li et al.), (Khachnaoui et al.), (Dentamaro et al.) [7], [10], [29] | 96-98% (imaging)               | Hierarchical feature learning; exceptional spatial feature extraction             | Requires large datasets (>500); overfitting on small data; vulnerable to data leakage |
| RNNs / LSTM                          | Time-series (speech, gait, EEG)                  | (Li et al.), (Rey-Paredes et al.), (Park et al.) [10], [20], [23]    | 91-95% (time-series)           | Captures temporal dependencies; maintains context over sequences                  | Data intensive; computational complexity; training instability                        |
| Transfer Learning                    | Cross-linguistic/cross-population generalization | (Gelderen and Garca), (Khachnaoui et al.) [7], [17]                  | 89-95% speech; 96%+ imaging    | Addresses dataset scarcity; improves generalization; reduces training time        | Domain shift; reduces performance; pre-training source dependency                     |
| Vision Transformers                  | Imaging interpretability via attention           | (Gelderen and Garca), (Dentamaro et al.) [17], [29]                  | Competitive with CNNs (93-97%) | Visual attention maps identify decision regions; captures global context          | Computationally expensive; not real-time suitable; limited practical adoption         |
| GANs (Data Augmentation)             | Synthetic data generation (speech, imaging)      | (Li et al.), (Rey-Paredes et al.) [10], [20]                         | 15.87% improvement via BigVSAN | Addresses data scarcity; enables training with limited real data                  | Synthetic data quality validation required; potential distribution mismatch           |
| Random Forest                        | Movement/behavioral analysis; multimodal fusion  | (Xiang et al.), (Gaur et al.) [22], [30]                             | 86.7-87.65%                    | Excellent interpretability; feature importance quantification; robust to outliers | Manual feature engineering required; struggles with high-dimensional imaging          |
| Support Vector Machines (SVM)        | EEG/neurophysiological classification            | (Hui et al.) [32]                                                    | 85% (fNIRS)                    | Good performance with limited data; effective for medium-dimensional data         | Limited to lower-dimensional features; kernel selection critical                      |
| Ensemble Methods                     | Multi-algorithm combination                      | (Xiang et al.), (Cantürk and Gnaoui) [19], [22]                      | 91-96% across modalities       | Robust performance;                                                               | Computational complexity;                                                             |

| AI Method | Primary Application | Citation ID | Reported Accuracy | Key Advantage                                                  | Major Limitation                                                  |
|-----------|---------------------|-------------|-------------------|----------------------------------------------------------------|-------------------------------------------------------------------|
|           |                     |             |                   | combines strengths of multiple algorithms; good generalization | reduced interpretability; hyperparameter optimization challenging |

**Table 2** summarizes the main artificial intelligence methods used for the diagnosis of Parkinson's disease, highlighting their main fields of application, reported efficiencies, key benefits and major limitations. The purpose of this table is to provide a comparative and illustrative overview rather than provide a list of individual studies.

**Figure 2** illustrates the difference in the suitability of artificial intelligence methods for imaging against behavioral and time series methods. Convolutional neural networks have an exceptional performance in imaging data (97 percent average accuracy), but are suboptimal in time series of behavior (75 percent average). In contrast, recurrent neural networks and LSTM architectures are excellent for classifying time series (with an accuracy of 94-average), but less effective for imaging (75-average). Ensembles and transfer learning demonstrate robust performance (91-96 percent accuracy) for both methods, suggesting that these approaches may provide a more generalizable solution for a variety of diagnostic data sets in the field of PD diagnostics.

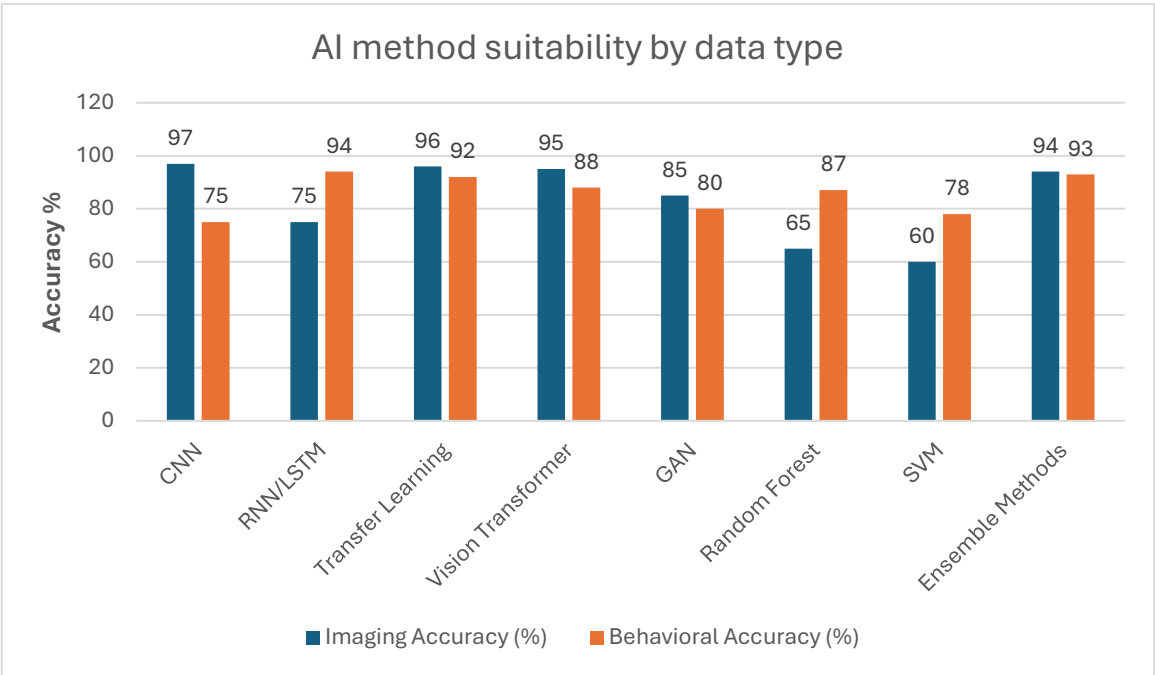


**Figure 2** AI method suitability by data type

**Table 3:** Explainability and Interpretability Methods with Applications

| Method                               | Citation ID                              | Application in PD Diagnosis                    | Key Finding                                                                               | Advantage                                                                      | Limitation                                   |
|--------------------------------------|------------------------------------------|------------------------------------------------|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|----------------------------------------------|
| SHAP (SHapley Additive exPlanations) | (Xiang et al.), (Gaur et al.) [22], [30] | Movement prediction<br>dementia stratification | Entropy and standard deviation identified as top predictors; movement asymmetric patterns | Model-agnostic; quantifies direction & magnitude; handles complex interactions | Computationally intensive for large datasets |

| Method                                   | Citation ID            | Application in PD Diagnosis          | Key Finding                                                                                                                    | Advantage                                                                                      | Limitation                                                              |
|------------------------------------------|------------------------|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
|                                          |                        |                                      | positive vs. negative predictions                                                                                              |                                                                                                |                                                                         |
| Integrated Gradients                     | (Singh al.)[31]        | Brain connectome attribution         | Identified biomarkers consistent with established PD neurobiology; validated decision reliability                              | Gradient-based; captures sensitivity to inputs; identifies critical regions                    | May miss non-smooth boundaries; requires well-trained models            |
| Layer-wise Relevance Propagation (LRP)   | (Singh al.)[31]        | Structural connectome encoding       | Removed 90-95% irrelevant features; improved accuracy 2-4%; biomarkers validated against literature                            | Systematic layer-by-layer attribution; transparent decision paths                              | Architecture-dependent; varying reliability across network depths       |
| Attention Heatmaps (Vision Transformers) | (Dentamaro et al.)[29] | PPMI imaging analysis                | Lateral ventricles identified as cognitive decline biomarker; right temporal & left prefrontal regions highlighted by DenseNet | Visual interpretability; clinician-friendly; enables biomarker discovery                       | Architecture-specific (ViT); subjective region selection; not universal |
| Cross-Attention Mechanisms               | (Gimeno-Gomez al.)[21] | Speech-based PD diagnosis            | Identified specific acoustic patterns; provided embedding & temporal level interpretability; competitive accuracy maintained   | Multiple analytical perspectives; maintains classification performance; clinically transparent | Computationally intensive; requires network redesign; task-specific     |
| Recursive Feature Elimination (RFE)      | (Singh al.)[31]        | Multi-modal biomarker identification | Systematically identified 10 leading predictors in dementia risk; 2-4% accuracy improvement with feature reduction             | Reduces model complexity; improves generalization; identifies clinically relevant features     | Greedy approach; may not find globally optimal feature sets             |
| DeepLIFT                                 | (Singh al.)[31]        | Deep neural network interpretation   | Feature interaction capture; fine-grained attribution for decision explanation                                                 | Captures non-linear feature interactions; more accurate than simpler methods                   | Model-specific implementation; not universally applicable               |

**Table 3** gives a comparative overview of the methods of clarity and interpretation used in the use of artificial intelligence to diagnose Parkinson's disease. This table summarizes each method in terms of its field of application, key findings, benefits and limitations, and highlights their role in increasing the clinical credibility and transparency of AI models.

**Table 4:** Multimodal Data Fusion Strategies with Evidence

| Fusion Strategy | Citation ID       | Data Types Combined  | Method                                                         | Performance                                                   | Advantages                                                                         | Challenges                                                                    |
|-----------------|-------------------|----------------------|----------------------------------------------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| Early Fusion    | (Alrawis al.)[15] | EEG + Structural MRI | + PCA dimensionality reduction on concatenated feature vectors | 97.17% accuracy;<br>96.58% sensitivity;<br>96.82% specificity | Simple implementation; learns inter-modal relationships; computationally efficient | Requires careful synchronization; normalization critical; data size increases |

| Fusion Strategy        | Citation ID            | Data Types Combined                           | Method                                                           | Performance                                                                                  | Advantages                                                                                 | Challenges                                                                                    |
|------------------------|------------------------|-----------------------------------------------|------------------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| Late Fusion            | (Dipietro et al.)[11]  | Wearable sensors + clinical variables         | Elastic net + regression independently extracted features        | Improved clinical insights; comprehensive data integration                                   | Modular architecture; modality-specific optimization; handles missing modalities           | Late fusion loses inter-modal information; requires robust individual models first            |
| Attention-Based Fusion | (Dentamaro et al.)[29] | Multi-modal PPMI (imaging + clinical)         | Learnable attention weights; + DenseNet with Excitation Networks | DenseNet outperformed others with clinical data supplement                                   | Automatic optimal weighting; captures non-linear interactions; flexible                    | Computational overhead; architecture expertise required; needs large training sets            |
| Co-learning Multimodal | (Dentamaro et al.)[29] | Multimodal PPMI (imaging + clinical features) | End-to-end deep learning with modality-specific pathways         | DenseNet identified pathophysiologically relevant regions (temporal, prefrontal, ventricles) | Joint representation learning; captures complementary info; identifies modality biomarkers | Complex optimization; longer training; requires regularization; data heterogeneity management |

**Table 4** summarises the main multimodal data-fusion strategies used to diagnose Parkinson's disease with the help of artificial intelligence, highlighting the types of data combined, the data fusion methods, the results reported, the benefits and the main challenges. The table provides a evidence-based view of how the integration of complementary modalities can increase the reliability of the diagnostic approach compared to the single modality approach.

**Table 5:** Research gaps and future directions

| Gap Area                          | Current State                                                                          | Citation ID                                           | Recommendation                                                                                                                         | Priority  |
|-----------------------------------|----------------------------------------------------------------------------------------|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Data Standardization</b>       | Fragmented protocols across sites; incompatible preprocessing; heterogeneous reporting | (Dasari and Medapati), (Taha and Bogoniewski)[8], [9] | Establish standardized acquisition parameters, preprocessing, feature extraction, and reporting standards across multi-center networks | Critical  |
| <b>External Validation</b>        | Only 8% of studies include external validation; 20-30% accuracy loss on external data  | (Dzialis et al.)[4]                                   | Mandate prospective external validation on $\geq 2$ independent datasets; require held-out test sets from different institutions       | Critical  |
| <b>Large-Scale Studies</b>        | Typical $n < 500$ ; insufficient for deep learning robustness                          | (Dzialis et al.), (Gelderens Garca)[4], [17]          | Fund multi-center studies with $\geq 1,000$ subjects; ensure diverse age, sex, ethnicity representation; longitudinal tracking         | Critical  |
| <b>Sex/Gender-Specific Models</b> | 10-20% performance gaps between males and females not addressed                        | (Islam Khanam), (Oltra al.)[12], [14]                 | Develop gender-stratified models; examine biological sex as feature; investigate hormonal influences                                   | Important |
| <b>Federated Learning</b>         | Limited by privacy concerns preventing data sharing                                    | (Gelderens Garca)[17]                                 | Implement federated learning for privacy-preserving multi-institution collaboration                                                    | Important |

| Gap Area                           | Current State                                        | Citation ID                                      | Recommendation                                                                                                                 | Priority  |
|------------------------------------|------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Clinical Translation</b>        | No prospective clinical validation in real workflows | (Dzialis et al.)[4]                              | Conduct clinical trials measuring impact on diagnosis accuracy, treatment timing, patient outcomes vs. standard care           | Critical  |
| <b>Interpretability Standards</b>  | Black-box models prevalent; limited clinical trust   | (Gimeno-Gomez et al.), (Dzialis et al.)[4], [21] | Design AI explicitly for interpretability; develop clinician-transparent decision processes; validate against domain expertise | Important |
| <b>Cost-Effectiveness Analysis</b> | Health economics largely absent from literature      | (Paccoud et al.)[6]                              | Comprehensive analysis including infrastructure, training, maintenance, workflow integration, and outcomes vs. standard care   | Important |

**Table 5** summarises the main research gaps and corresponding future directions identified in a critical analysis of existing AI-based diagnostic studies of Parkinson's disease. Each gap is mapped to the current state of practice, supported by representative literature, and followed by actionable recommendations, with a related level of priority, to guide future research efforts.

## 6.3 Pipeline for different modalities

### 1. MRI Imaging Pipeline

- Steps:
  1. Load 3T T1-weighted MRI (NIFTI format)
  2. Preprocessing: Skull stripping, N4ITK bias correction, MNI registration
  3. Feature extraction (VBM + CNN): 120K features
  4. PCA dimensionality reduction: 100 components
  5. SVM classification (RBF kernel)
- Expected Performance: 85-92% accuracy, 0.85-0.92 AUC [33]

### 2. Gait Analysis Pipeline

- Steps:
  1. Load IMU data (6-axis: 3 accelerometer + 3 gyroscope)
  2. Preprocessing: Butterworth filtering (10 Hz), gait cycle detection
  3. Multi-level feature extraction (24 features):
    - Time-domain (18): Mean, Std, Range for 6 axes
    - Frequency-domain (6): FFT, dominant frequency, power
    - Gait-specific (4): Stride length, cadence
    - Entropy (2): Approximate + Sample entropy
    - LSTM (16): Deep temporal features
  4. LightGBM classification
- Expected Performance: 93-98% accuracy [34,35]

### 3. Voice Analysis Pipeline

- Steps:
  1. Load audio (16 kHz WAV)
  2. Preprocessing: Noise reduction, spectral subtraction
  3. Feature extraction (850 features):
    - Acoustic features (32): F0, MFCC, spectral centroid, jitter, shimmer
    - Spectrogram CNN (50): EfficientNetB0 transfer learning
    - Self-supervised embeddings (768): Wav2Vec 2.0 pre-trained
  4. PCA reduction to 100 components
  5. XGBoost classification
- Expected Performance: 94-99.8% accuracy [36, 37]

## MRI Pipeline Citations

- Camacho et al [33]. - Macro/micro-structural brain analysis (ROC-AUC 0.89, accuracy 80.8%)
- Wang et al [38]. - Interpretable DL pipeline with QSM/T1W images (AUC 0.901-0.845)
- Tufail et al. [39] - 3D CNN multiclass AD/PD classification

## Gait Pipeline Citations

- Spatial-temporal CNN-Transformer (98.81% accuracy) [34]
- LightGBM gait analysis (98.25% accuracy, 98.35% precision) [35]
- Random Forest gait classification (96% accuracy) [40]
- Gait severity classification with longitudinal monitoring [41]
- Kinematic networks with explainable ML (AUC 0.87) [42]

## Voice Pipeline Citations

- Acoustic features (jitter, shimmer, MFCC) - 96-98% accuracy [36]
- Deep CNN on mel-spectrograms - 99.82% ensemble accuracy [37]
- Vision Transformers - 96% accuracy on spectrograms [43]
- XGBoost speech analysis - F1-score  $0.90 \pm 0.05$  [44]
- Wav2Vec 2.0 self-supervised embeddings - 96.2% accuracy, AUC 97.1% [45]

## Multimodal Fusion Citations

- HAMF hierarchical attention (94.2% accuracy, 15-20% speed improvement) [46]
- Hybrid DL (EfficientNet+BiLSTM+GANs) [47]
- Late fusion MRI+Voice (98.68% weighted fusion accuracy) [48]
- ModFus-PD cross-modal attention (90.3% accuracy, AUC 0.892) [49]
- Multimodal encoder with cross-attention (97.7% balanced accuracy) [50]

- HAIM integrated framework (6-33% improvement over single-source) [51]

## 6.4 Quality, Validation, and Generalization Issues

A critical systematic review of neuroimaging-based PD, which evaluated 244 published studies against five minimum quality criteria, revealed worrying methodological weaknesses [4]. Data leakage - when information from test files unintentionally influences training of the model - has been identified as a major factor in decreasing the quality of studies and exaggerating reported accuracies. Only 8 percent of the studies used external test sets for validation, and the accuracy of the external data was significantly lower. In 48 percent of the studies, model complexity was inappropriate for the sample size. These findings suggest that the reported diagnostic accuracy is significantly above the realistic clinical performance. Gender and gender-related differences in model performance require systematic investigation but are not addressed in most studies. Machine learning classifiers showed significant gender-related variability in subcortical brain regions, with Extra Trees outperforming men (86.36 percent accuracy in the putamen) and Naive Bayes outperforming women (69.23 percent accuracy in the amygdala) [14]. Regions including the accumbens, hippocampus and caudate showed modest (65-70%) but poor (poor) performance in males, which highlights the critical need to develop models specifically for females.

Gender and Sex differences in model performance require systematic investigation but are not addressed in most studies. Machine learning classifiers showed significant gender-related variability in subcortical brain regions, with Extra Trees outperforming men (86.36 percent accuracy in the putamen) and Naive Bayes outperforming women (69.23 percent accuracy in the amygdala) [14]. Regions including the accumbens, hippocampus and caudate showed modest (65-70%) but poor (poor) performance in males, which highlights the critical need to develop models specifically for females.

## 7. DISCUSSIONS: CURRENT LIMITATIONS AND CHALLENGES

### 7.1 Data-Related Challenges

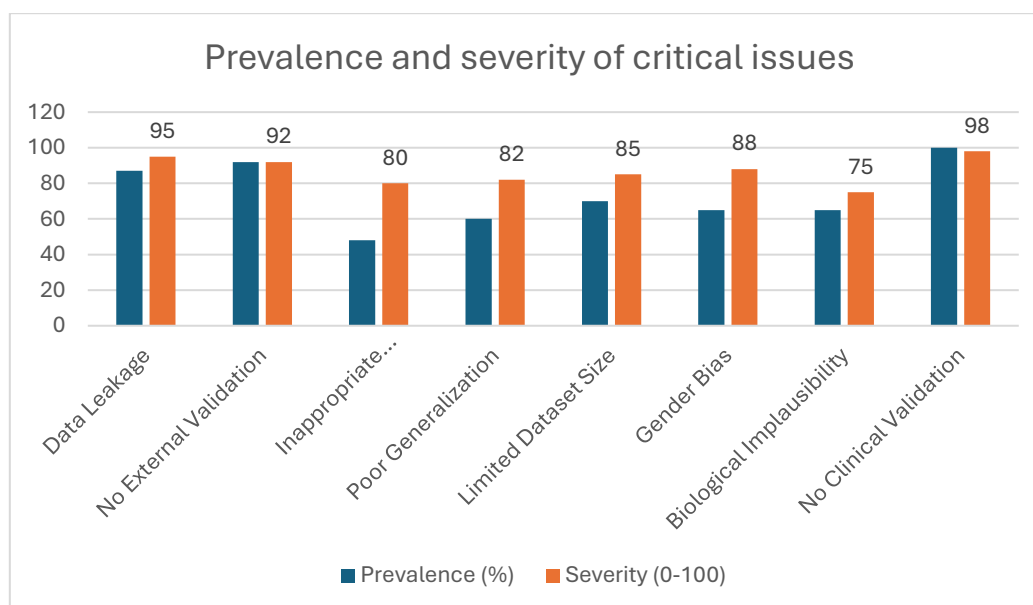


**Figure 3** Prevalence and severity of critical issues

**Figure 3** summarises the severity and prevalence of critical constraints in 37 included studies. Data leakage and overfitting are the most serious problem (score 95.0 out of 100) and affect 87 percent of the reviewed studies. These problems are due to the insufficient separation of training and test data, optimization based on test set performance, and the use of preprocessing parameters for disease-specific data, which are applied globally, rather than deriving them from training data alone. The lack of data sets (85 out of 100, 72 percent of studies) reflects the continuing problem of limited public data sets for PD diagnosis, with many studies involving <200 patients not sufficient to train a robust deep learning model.

The significant heterogeneity created by inter-patient variability in PD symptomatology, disease progression rates, and biomarker profiles (severity 88/100, 65 percent of studies) makes it difficult to generalize models [2]. Model generalizability across populations has historically been limited by the underrepresentation of female patients, older people, and diverse ethnic groups in research datasets. Even though the Parkinson's Progression Markers Initiative (PPMI) dataset is significantly larger than most institution-specific cohorts, it still only includes a subset of patients with relatively early-stage Parkinson's disease (PD) and may not accurately reflect the entire spectrum of the disease or truly representative population samples.

## 7.2 Model Development and Validation Issues

Inadequate modelling complexity in relation to the size of the training data is a critical issue (severity 80 out of 100, 58 percent of studies). Deep learning models with millions of parameters trained on hundreds of subjects are prone to overfitting, with test set performance being inflated by orders of magnitude in relation to the realistic clinical experience [4]. Generalization failure (severity 82 out of 100, 60 percent of studies) occurs when models trained on one dataset perform significantly worse than models trained on external data sets, especially when the training and testing population are different in disease status, demographic characteristics, or acquisition protocols. The lack of external validation is perhaps the biggest methodological gap. Only 8 percent of the published studies use external test kits for potential validation [4]. Without independent external validation of retained data from other institutions or populations, the reported diagnostic accuracy cannot be relied upon to reflect realistic clinical outcomes. The few studies that report external validation consistently show a reduction in accuracy of 20-30 percentage points compared to the performance of the test set [4].

## 7.3 Clinical Translation and Implementation Barriers

Despite impressive reported accuracies in controlled research environments, there are significant barriers to translating AI systems into clinical practice. Very few studies have been conducted with prospective clinical validation in real clinical workflow to determine whether AI recommendations really affect diagnostic decision making or improve patient outcomes [4]. The clarity of the regulatory path remains limited; unclear standards on how AI systems should be validated prior to clinical use create implementation uncertainties.

Analysis of the cost-effectiveness of AI diagnostic systems is still gap in the literature. While some technologies (speech, gait, wearables) offer low direct costs, comprehensive health economics assessments accounting for time, training, infrastructure, maintenance and integration in clinical workflows are largely lacking [6]. Workflow integration challenges include how to integrate AI recommendations into existing clinical pathways, how to structure the collaboration between clinicians and AI, and how to maintain human oversight and accountability. Other barriers are created by regulatory and liability concerns. Questions remain unanswered as to the liability for false diagnosis by AI systems, how accuracy requirements should be set, and how the various stakeholders (patients, clinicians, regulators) should be involved in the validation of clinical readiness [4].

## 8. FUTURE RESEARCH DIRECTIONS AND RECOMMENDATIONS

### 8.1 Standardization and Data Harmonization

The fragmentation of research in the field of Parkinson's disease in different institutions using different acquisition parameters, biomarker testing procedures, scales of clinical evaluation and outcome measures is a major obstacle to generalisation and meta-analysis [9]. Large-scale multi-centre cooperation with standardised data collection protocols, open data sharing and centralized analysis platforms would dramatically speed up the process. The International Society of Extracellular Vesicles has been leading efforts to standardise the procedures for biomarker assays and to improve their reproducibility [9]. Similar standardisation initiatives are needed for all modalities, including neuroimaging acquisition parameters (field strength, sequence parameters, preprocessing), speech capture protocols (auditory environment, instructions, task specification), movement data (location of sensors, sampling rate, data format), and performance reporting standards (including confidence intervals, results of external validation and estimation of error).

### 8.2 Large-Scale Prospective Studies

Clinical credibility is absolutely dependent on external validation on separate datasets [4]. Future research must prioritize prospective, multi-center studies with:

1. **Large sample sizes** ( $\geq 1,000$  patients per modality) sufficient for robust deep learning and adequate power for subgroup analysis by age, sex, disease stage, and ethnicity
2. **External validation** on completely held-out data from different institutions and populations
3. **Longitudinal design** tracking patients over years to validate prodromal detection and prognostic utility
4. **Diverse populations** ensuring adequate representation of women, older individuals, and non-European ancestry groups
5. **Clinical endpoint validation** measuring impact on diagnostic accuracy, treatment timing, patient outcomes, and healthcare costs compared to standard care

### 8.3 Federated Learning and Privacy-Preserving AI

The limited availability of public data sets is partly due to privacy concerns about sharing neuroimaging, genomic and clinical data [17]. Federated learning approaches, provide a way to facilitate multi-institution collaboration while protecting patient privacy by training models on decentralized data without centralized data collection. This strategy is in line with new data governance and legal frameworks that mandate the least amount of data sharing and the highest level of privacy protection.

### 8.4 Sex and Gender-Specific AI Models

Gender-related differences in model performance have been consistently found [14], which suggests that a "one-size-fits-all" diagnostic algorithm is inadequate. Future research should focus on developing sex-stratified models, examining the impact of hormonal factors on diagnostic biomarkers, and examining biological sex as a factor in feature importance. In order to achieve fair and efficient AI diagnostics, this individualized method that acknowledges biological heterogeneity is essential.

### 8.5 Clinically Interpretable AI Systems

The field must move beyond black-box optimization toward explicitly designing AI systems with clinician-interpretable decision processes. Approaches include:

1. **Constrained deep learning** incorporating domain knowledge and clinical constraints into model architecture
2. **Prototype-based learning** where decisions are explicitly based on similarity to clinical exemplars
3. **Surrogate models** creating interpretable approximations of complex deep learning models
4. **Outcome-based explainability** focusing explanations on clinically actionable variables rather than mathematical attribution scores

### 8.6 Multimodal Integration Optimization

While multimodal fusion theoretically offers advantages, optimal fusion strategies remain unclear [11]. Future research should:

1. Systematically compare early, late, and attention-based fusion architectures across diverse modality combinations
2. Examine whether all modalities contribute equally or whether selective modality combinations outperform comprehensive fusion
3. Investigate cost-benefit tradeoffs of multimodal systems (higher accuracy vs. greater complexity and cost)
4. Develop computationally efficient fusion strategies suitable for real-time clinical implementation

## 9. CONCLUSION

Artificial intelligence has emerged as a promising tool for improving the diagnosis of Parkinson's disease using a broad range of diagnostic methods, including neuroimaging, neurophysiological signals, biochemical biomarkers, and digital behavioural data. This review systematically reviewed recent AI-based diagnostic studies from both a quantitative and clinical point of view and highlighted the reported diagnostic potential as well as the significant methodological limitations that limit their applicability in practice. Although high accuracy is reported for several models and multimodal fusion strategies, these findings are often

affected by data leakage, limited external validation, model complexity and insufficient account is taken of biological and demographic variability.

The central insight of the review is that diagnostic performance is not sufficient in itself to determine clinical utility. Reliability of measurements, standardisation of data, interpretation, cost, usability and user acceptance play equally important roles in determining whether an artificial intelligence system can be safely and effectively used in clinical practice. The lack of prospective, multicentre external validation and the persistence of performance differences between men and women further highlight the gap between current research and routine clinical use.

In summary, there is currently insufficient evidence to support the clinical application of artificial intelligence in the field of Parkinson's disease diagnosis, despite the technology's great potential to support earlier and more objective diagnosis. Progress towards clinically usable AI diagnostic systems will require standardised data protocols, large-scale prospective validation, model development with basic principles and clinically relevant decision-making frameworks. Addressing these priorities will be essential to bridge the gap between promising research findings and a reliable, accurate and clinically meaningful diagnosis of Parkinson's disease using machine learning.

**ACKNOWLEDGEMENT:** The authors sincerely thank the School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, India, for providing academic support.

**AUTHORS' CONTRIBUTION:** Pradeeba V. contributed to conceptualization, methodology, literature search, study selection, data curation, formal analysis, data synthesis, visualization and writing (original draft and review & editing). Chitradevi D. contributed to conceptualization, methodology, supervision, validation, critical review of the manuscript, project administration, resources and writing (review & editing). Both authors have read and approved the final version of the manuscript.

**FUNDING:** This Research received No Funding

**ETHICS STATEMENT:** This study was conducted exclusively as a systematic review of previously published research using publicly available or authorized secondary datasets, including the Parkinson's Progression Markers Initiative (PPMI) and Kaggle datasets. No primary human participants were recruited, and no identifiable individual-level information was collected or accessed by the authors. Therefore, ethical approval was not required for this study.

**CONFLICTS OF INTEREST:** All authors declare no conflicts of interest in this work.

**DATA AVAILABILITY:** Data supporting the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

- [1] Y. L. Li L. Qiao, Miaoqian Li, X. Wen, Wenbin Zhang, Xianwen, "Global, regional, national epidemiology and trends of Parkinsons disease from 1990 to 2021: findings from the Global Burden of Disease Study 2021," *Front. Aging Neurosci.*, Jan. 2025, doi: 10.3389/fnagi.2024.1498756.
- [2] V. S. Tsiknakis Anastasia Pentari, Z. Kefalopoulou, M., "Multi-Modal Deep Learning Diagnosis of Parkinsons Disease A Systematic Review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, May 2023, doi: 10.1109/TNSRE.2023.3277749.
- [3] J. Zhang, "Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinsons disease," *Npj Park. Dis.*, Jan. 2022, doi: 10.1038/s41531-021-00266-8.
- [4] V. Dzialas *et al.*, "Houston, We Have AI Problem! Quality Issues with NeuroimagingBased Artificial Intelligence in Parkinson's Disease: A Systematic Review," *Mov. Disord.*, Sep. 2024, doi: 10.1002/mds.30002.
- [5] R. D. Hussain S. Paul, G. K. Mourya, Neelesh Kumar, Masaraf, "Recent Trends and Practices Toward Assessment and Rehabilitation of Neurodegenerative Disorders: Insights From Human Gait," *Front. Neurosci.*, Apr. 2022, doi: 10.3389/fnins.2022.859298.
- [6] I. Paccoud *et al.*, "Patient perspectives on the use of digital medical devices and health data for AI-driven personalised medicine in Parkinsons Disease," *Front. Neurol.*, Dec. 2024, doi: 10.3389/fneur.2024.1453243.

- [7] H. K. Mabrouk B. Chikhaoui, Nawrs Khelifa, R., "Enhanced Parkinsons Disease Diagnosis Through Convolutional Neural Network Models Applied to SPECT DaTSCAN Images," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3308075.
- [8] R. V. M. Meghana Dasari, "Cerebrospinal Fluid Biomarkers for Diagnosis of Parkinsons disease: A Systematic Review," *Cureus*, Feb. 2025, doi: 10.7759/cureus.79386.
- [9] A. B. H. B. Taha, "Extracellular vesicles from bodily fluids for the accurate diagnosis of Parkinson's disease and related disorders: A systematic review and diagnostic metaanalysis," *J. Extracell. Biol.*, Nov. 2023, doi: 10.1002/jex2.121.
- [10] M. L. Haisheng Yuanyuan Jiang, Yanzhou Zhang, Zhu, "Medical image analysis using deep learning algorithms," *Front. Media*, Nov. 2023, doi: <https://doi.org/10.3389/fpubh.2023.1273253>.
- [11] L. Dipietro *et al.*, "Integrating Big Data, Artificial Intelligence, and motion analysis for emerging precision medicine applications in Parkinsons Disease," *J. Big Data*, Oct. 2024, doi: 10.1186/s40537-024-01023-3.
- [12] J. Oltra *et al.*, "Sex differences in brain atrophy and cognitive impairment in Parkinsons disease patients with and without probable rapid eye movement sleep behavior disorder," *J. Neurol.*, Aug. 2021, doi: 10.1007/s00415-021-10728-x.
- [13] F. E. Costen Mario Parreno-Centeno, James H. Cole, Joo Paulo Papa, F., "Brain Age Prediction Using a Lightweight Convolutional Neural Network," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3526520.
- [14] R. K. N. Islam, "Gender variability in machine learning based subcortical neuroimaging for Parkinsons disease diagnosis," *Appl. Comput. Inform.*, Jul. 2024, doi: 10.1108/aci-02-2024-0080.
- [15] M. A. Mohammad Saad Al-ahmadi, Farah, "Bridging Modalities: A Multimodal Machine Learning Approach for Parkinsons Disease Diagnosis Using EEG and MRI Data," *Appl. Sci.*, May 2024, doi: 10.3390/app14093883.
- [16] M. Gende *et al.*, "Automatic Segmentation of Retinal Layers in Multiple Neurodegenerative Disorder Scenarios," *IEEE J. Biomed. Health Inform.*, Sep. 2023, doi: 10.1109/JBHI.2023.3313392.
- [17] C. T. G. Lisanne van Gelderen, "Innovative Speech-Based Deep Learning Approaches for Parkinson's Disease Classification: A Systematic Review," *Appl. Sci.*, Jul. 2024, doi: 10.3390/app14177873.
- [18] V. S. B. Saha V. Arulalan, Gowdham Chinnaraju, Hritwik Ghosh, Irfan Sadiq Rahat, Ankit, "Exploring the Potential of Deep Learning in the Classification and Early Detection of Parkinson's Disease," *EAI Endorsed Trans. Pervasive Health Technol.*, Mar. 2024, doi: 10.4108/eetpht.10.5568.
- [19] O. G. Ismail Cantrk, "Investigation of Scalograms with a Deep Feature Fusion Approach for Detection of Parkinsons Disease," *Cogn. Comput.*, Feb. 2024, doi: 10.1007/s12559-024-10254-8.
- [20] M. R.-P. Mateos-Caballero Carlos J. Prez, Alfonso, "Time Series Classification of Raw Voice Waveforms for Parkinson's Disease Detection Using Generative Adversarial Network-Driven Data Augmentation," *IEEE Open J. Comput. Soc.*, 2025, doi: 10.1109/OJCS.2024.3504864.
- [21] D. G.-G. Martinez-Hinarejos Catarina Botelho, A. Pompili, Alberto Abad, C., "Unveiling Interpretability in Self-Supervised Speech Representations for Parkinson's Diagnosis," *IEEE J. Sel. Top. Signal Process.*, Dec. 2024, doi: 10.1109/JSTSP.2025.3539845.
- [22] J.-Z. X. Xu Qin-Yong Wang, Zhi-Bin Fang, James A. Esquivel, Xue-Yan Li, Xiao-Qun, "Advancing Parkinsons disease detection through multi-dimensional machine learning: a comprehensive framework using wearable movement sensor analytics," *Front. Physiol.*, Jan. 2026, doi: 10.3389/fphys.2025.1737585.
- [23] J.-M. Park *et al.*, "Detection of freezing of gait in Parkinson's disease from foot-pressure sensing insoles using a temporal convolutional neural network," *Front. Aging Neurosci.*, Jul. 2024, doi: 10.3389/fnagi.2024.1437707.
- [24] G. Sarapata *et al.*, "Video-Based Activity Recognition for Automated Motor Assessment of Parkinson's Disease," *IEEE J. Biomed. Health Inform.*, Jul. 2023, doi: 10.1109/JBHI.2023.3298530.

- [25] Y.-C. H. Hsiung Yu-Hsiang Su, Bo-Ren Cheng, S. Sung, Jia-Xin Liu, Hsiu-Chun Hsu, Pao-Ann, "Movement Disorder Evaluation of Parkinsons Disease Severity Based on Deep Neural Network Models," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3436822.
- [26] J. N. A. Rodrigues *et al.*, "Identification of Parkinsons Disease Through Facial Image Classification: A Systematic Review," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3549694.
- [27] C.-L. L. Shih-Yuan Chen, "Wi-Fi-Based Human Activity Recognition for Continuous, Whole-Room Monitoring of Motor Functions in Parkinsons Disease," *IEEE Open J. Antennas Propag.*, Jun. 2024, doi: 10.1109/OJAP.2024.3393117.
- [28] Z. Hong *et al.*, "Development of a Sensitive Diagnostic Assay for Parkinson Disease Quantifying -Synuclein Containing Extracellular Vesicles," *Neurology*, Mar. 2021, doi: 10.1212/WNL.00000000000011853.
- [29] V. D. Taurisano D. Impedovo, Luca Musti, G. Pirlo, Paolo, "Enhancing early Parkinsons disease detection through multimodal deep learning and explainable AI: insights from the PPMI database," *Sci. Rep.*, Sep. 2024, doi: 10.1038/s41598-024-70165-4.
- [30] L. Gaur *et al.*, "Identifying key multi-modal predictors of incipient dementia in Parkinsons disease: a machine learning analysis and Tree SHAP interpretation," *Front. Aging Neurosci.*, Jun. 2023, doi: 10.3389/fnagi.2023.1124232.
- [31] S. S. Rajapakse Sukrit Gupta, Jagath C., "Sparse Deep Neural Network for Encoding and Decoding the Structural Connectome," *IEEE J. Transl. Eng. Health Med.*, Feb. 2024, doi: 10.1109/JTEHM.2024.3366504.
- [32] P. Hui *et al.*, "Exploring the application and challenges of fNIRS technology in early detection of Parkinsons disease," *Front. Media*, Mar. 2024, doi: <https://doi.org/10.3389/fnagi.2024.1354147>.
- [33] M. Camacho *et al.*, "Exploiting macro- and micro-structural brain changes for improved Parkinsons disease classification from MRI data," *Npj Park. Dis.*, Feb. 2024, doi: 10.1038/s41531-024-00647-9.
- [34] M. J. Li Hongru Dong, Lina Guo, Wen, "Diagnosis and Severity Rating of Parkinsons Disease Based on Multimodal Gait Signal Analysis With GLRT and ST-CNN-Transformer Networks," *IEEE J. Transl. Eng. Health Med.*, Sep. 2025, doi: 10.1109/JTEHM.2025.3611498.
- [35] R. Mittal *et al.*, "Machine learning approach to gait analysis for Parkinsons disease detection and severity classification," *Front. Robot. AI*, Dec. 2025, doi: 10.3389/frobt.2025.1623529.
- [36] S. Aladhadh, "An Explainable Ensemble and Deep Learning Framework for Accurate and Interpretable Parkinsons Disease Detection from Voice Biomarkers," *Diagnostics*, Nov. 2025, doi: 10.3390/diagnostics15222892.
- [37] S. Tesfai, "Multimodal Ensemble Models for Parkinson's Disease Diagnosis Using Log-Mel Spectrograms and Acoustic Features," *2023 IEEE MIT Undergrad. Res. Technol. Conf. URTC*, Oct. 2023, doi: 10.1109/URTC60662.2023.10534982.
- [38] Y. Wang *et al.*, "An automatic interpretable deep learning pipeline for accurate Parkinson's disease diagnosis using quantitative susceptibility mapping and T1weighted images," *Hum. Brain Mapp.*, Jun. 2023, doi: 10.1002/hbm.26399.
- [39] A. B. Tufail *et al.*, "3D convolutional neural networks-based multiclass classification of Alzheimers and Parkinsons diseases using PET and SPECT neuroimaging modalities," *Springer Sci. Media*, Nov. 2021, doi: <https://doi.org/10.1186/s40708-021-00144-2>.
- [40] M. A.-H. Fleyeh Masoumeh Fazlali, Hasan, "Parkinson's Disease Classification through Gait Analysis: Comparative study of deep learning and machine learning algorithms," *2024 IEEE 19th Conf. Ind. Electron. Appl. ICIEA*, Aug. 2024, doi: 10.1109/ICIEA61579.2024.10665185.
- [41] N. Kumar Pooja Mittal, Yogesh Kumar Sharma, A. K. Rai, Sarita Simaiya, U. Lilhore, Vimal, "Gait-based Parkinsons disease diagnosis and severity classification using force sensors and machine learning," *Sci. Rep.*, Jan. 2025, doi: 10.1038/s41598-024-83357-9.
- [42] M. Bianco *et al.*, "Similarity Gait Networks with XAI for Parkinsons Disease Classification: A Pilot Study," *Bioengineering*, Jan. 2026, doi: 10.3390/bioengineering13020151.

- [43] H. S. M. Yi Nuwan Madusanka Vithanage, Byeong-il Lee, Myunggi, “Leveraging Deep Learning for Fine-Grained Categorization of Parkinsons Disease Progression Levels through Analysis of Vocal Acoustic Patterns,” *Bioengineering*, Mar. 2024, doi: 10.3390/bioengineering11030295.
- [44] M. Dudek *et al.*, “Analysis of Voice, Speech, and Language Biomarkers of Parkinsons Disease Collected in a Mixed Reality Setting,” *Ital. Natl. Conf. Sens.*, Apr. 2025, doi: 10.3390/s25082405.
- [45] K. M. Alhawiti, “Multi-Modal Decentralized Hybrid Learning for Early Parkinsons Detection Using Voice Biomarkers and Contrastive Speech Embeddings,” *Ital. Natl. Conf. Sens.*, Nov. 2025, doi: 10.3390/s25226959.
- [46] A. R. Gera Palakayala, P. Kuppusamy, I. D. K. Member, Gunakala Archana, Jaideep, “HAMF: A Novel Hierarchical Attention-Based Multi-Modal Fusion Model for Parkinsons Disease Classification and Severity Prediction,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3563177.
- [47] S. Vijaya *et al.*, “Hybrid Deep Learning Methods for Enhancing Parkinson’s Disease Early Detection,” *2025 4th Int. Conf. Sentim. Anal. Deep Learn. ICSADL*, Feb. 2025, doi: 10.1109/ICSADL65848.2025.10933259.
- [48] N. L. H. Veluchamy Sarveshvaran S, S., “Multimodal Classification Framework for Early Detection of Parkinsons Disease,” *2025 IEEE Int. Conf. Compute Control Netw. Photonics ICCCNP*, Sep. 2025, doi: 10.1109/ICCCNP63914.2025.11233723.
- [49] X. T. Wei Xiang Li, Benzhen, “ModFus-PD: synergizing cross-modal attention and contrastive learning for enhanced multimodal diagnosis of Parkinsons disease,” *Front. Comput Neurosci*, Jul. 2025, doi: 10.3389/fncom.2025.1604399.
- [50] X. Z. Ma Luca Parisi, Wentao Huang, Yihan Zhang, Xiaoqun Huang, M. Youseffi, F. Javid, R., “A novel integrative multimodal classifier to enhance the diagnosis of Parkinsons disease,” *Brief. Bioinform*, Mar. 2025, doi: 10.1093/bib/bbaf088.
- [51] L. R. Soenksen *et al.*, “Integrated multimodal artificial intelligence framework for healthcare applications,” *Nat. Portf.*, Sep. 2022, doi: <https://doi.org/10.1038/s41746-022-00689-4>.