Frontiers in Computer Science and Artificial Intelligence
ISSN: 2978-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Deep Learning and Explainable Benchmarking for Early Parkinson's Disease Detection Using Speech Signals in the United States

Sadia Afrin Dipa[1], Hasibul Islam[2], Mousumi Akter[3], Farmina Sharmin[4], Md Shahnawaj[5], and Hamim Islam Hellol[6]

[1] *Department of Mathematics, The University of Texas at Arlington, Texas, USA*
[2 3 4] *School of Business, International American University, Los Angeles, California, USA*
[5 6] *Department of MS in Information System, Pacific States University, Los Angeles, California, USA*
Corresponding Author: Hamim Islam Hellol, **E-mail**: p25860@psuca.edu

| ABSTRACT

Early-stage Parkinson's disease (Early PD) detection using speech analysis has emerged as a promising and non-invasive approach for improving neurological healthcare in the United States. However, existing studies remain difficult to compare due to variations in datasets, speech tasks, languages, evaluation strategies, and definitions of Early PD. To address these limitations, this study proposes a comprehensive benchmark framework for speech-based Early PD detection using speaker-independent evaluation protocols to ensure fair, reproducible, and clinically reliable comparisons. The proposed benchmark evaluates multiple speech tasks under different training-resource settings and provides multidimensional performance analysis based on dataset characteristics, gender, aggregation level, and disease severity. Experimental findings offer actionable insights into the robustness and generalizability of speech-based Parkinson's detection systems. The proposed benchmark establishes a reliable reference framework for advancing explainable, scalable, and clinically meaningful Early PD detection technologies within modern U.S. healthcare and neurological diagnostic systems.

| KEYWORDS

Early-Stage Parkinson's Disease, Parkinson's Disease Detection, Speech-Based Diagnosis, Neurological Disorder Detection, Machine Learning, Deep Learning, Speech Analysis, Explainable Artificial Intelligence (XAI), Benchmark Framework, Speaker-Independent Evaluation, Healthcare AI, U.S. Healthcare System, Clinical Decision Support, Voice Biomarkers, Neurological Healthcare.

| ARTICLE INFORMATION

ACCEPTED: 17 August 2025 **PUBLISHED:** 18 November, 2025 **DOI:** 10.32996/fcsai.2025.4.5.8X

1. Introduction

Parkinson's disease (PD) is recognized as one of the most common and rapidly growing neurodegenerative disorders worldwide, affecting millions of individuals and creating a substantial burden on healthcare systems, particularly in the United States [1], [2]. According to recent neurological health reports, PD is currently the second most prevalent neurodegenerative disease after Alzheimer's disease, with incidence rates continuing to increase due to aging populations and improved life expectancy [3], [4]. In the United States alone, nearly one million individuals are estimated to be living with Parkinson's disease, while thousands of new cases are diagnosed annually [5], [6]. The progressive nature of PD significantly impacts motor function, cognitive performance, speech production, and overall quality of life [7], [8]. Consequently, early diagnosis and continuous monitoring have become essential priorities in modern neurological healthcare systems [9], [10]. Parkinson's disease is primarily characterized by the degeneration of dopaminergic neurons within the substantia nigra region of the brain, leading to impaired motor coordination and neurological dysfunction [11], [12]. Common clinical manifestations include tremors, rigidity, bradykinesia, postural instability, and gait abnormalities [13], [14]. However, non-motor symptoms such as speech impairment, cognitive decline, depression, and sleep disorders frequently emerge during the early stages of disease progression [15], [16]. Among these symptoms, speech impairment has attracted increasing attention because subtle vocal abnormalities may appear several years before prominent motor symptoms become clinically observable [17], [18]. These speech-related abnormalities include reduced vocal intensity, monotonic speech, articulation difficulties, dysarthria, breathiness, hoarseness, and abnormal speech rhythm [19], [20]. Since speech production involves complex neuromuscular coordination, changes in vocal patterns can provide valuable biomarkers for early neurological dysfunction [21], [22].

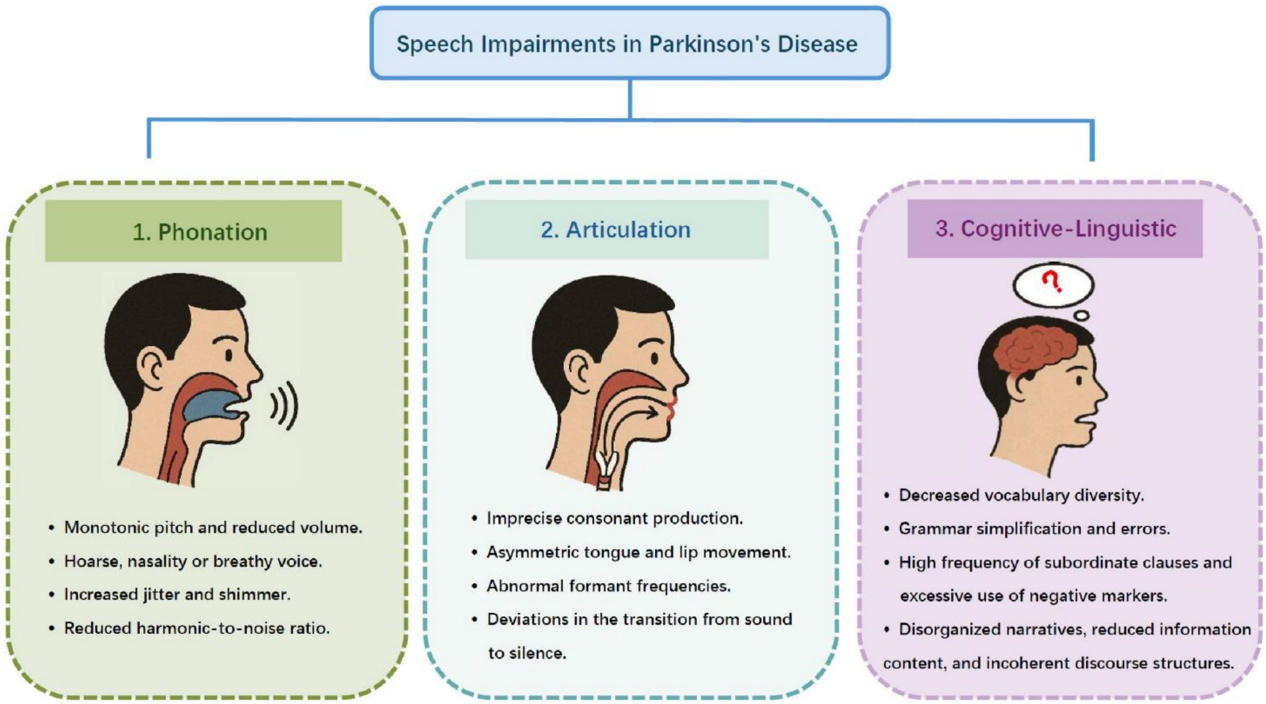

Figure 1. Neurological degeneration and speech-related symptoms associated with early-stage Parkinson's disease, including motor dysfunction and vocal abnormalities caused by dopaminergic neuron loss in the substantia nigra region of the brain.

The growing availability of digital healthcare technologies and speech-processing systems has accelerated interest in speech-based Parkinson's disease detection [23], [24]. Compared with traditional neurological examinations and imaging procedures, speech analysis offers a non-invasive, low-cost, scalable, and remotely accessible diagnostic alternative [25], [26]. This is particularly relevant in the United States healthcare environment, where increasing patient volumes and shortages of neurological specialists create challenges for timely diagnosis and monitoring [27], [28]. Speech-based assessment systems can support clinicians by enabling continuous remote monitoring, telemedicine integration, and early screening of at-risk individuals [29], [30]. Furthermore, speech acquisition can be performed using smartphones, microphones, or telehealth platforms, making it highly suitable for large-scale healthcare deployment [31], [32]. Over the past decade, machine learning (ML) and deep learning (DL) techniques have emerged as powerful tools for automated Parkinson's disease detection using speech signals [33], [34]. Early studies primarily relied on handcrafted acoustic features such as jitter, shimmer, pitch variation, harmonics-to-noise ratio, Mel-frequency cepstral coefficients (MFCCs), and spectral descriptors [35], [36]. These features were then analyzed using traditional machine learning classifiers including Support Vector Machines (SVM), Random Forest (RF), Decision Trees, k-Nearest Neighbors (k-NN), and Logistic Regression [37], [38]. Although these approaches demonstrated promising classification performance, they often depended heavily on feature engineering and lacked robustness across heterogeneous datasets [39], [40].

Recent advances in deep learning have significantly transformed speech-based neurological disorder detection [41], [42]. Deep neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) models, transformers, and self-supervised learning frameworks have demonstrated remarkable capability in automatically extracting discriminative representations from raw speech signals [43], [44]. These models can capture complex temporal, spectral, and linguistic patterns associated with Parkinsonian speech abnormalities [45], [46]. Furthermore, transfer learning and pre-trained speech representations have improved generalization performance by leveraging large-scale speech corpora [47], [48]. As a result, deep learning approaches increasingly outperform traditional feature-based systems in speech-based Parkinson's disease classification tasks [49], [50]. Despite these technological advancements, the majority of existing studies focus primarily on binary classification between Parkinson's disease patients and healthy controls [51], [52]. While such approaches provide useful proof-of-concept demonstrations, they often offer limited practical value in real-world clinical environments where experienced neurologists can already distinguish advanced PD from healthy individuals [53], [54]. From a clinical perspective, identifying early-stage Parkinson's disease (EarlyPD) is substantially more important because early intervention may slow disease progression, improve patient management, and enhance long-term quality of life [55], [56]. Consequently, speech-based EarlyPD detection has emerged as a highly significant research direction in neurological healthcare and artificial intelligence-assisted diagnostics [57], [58].

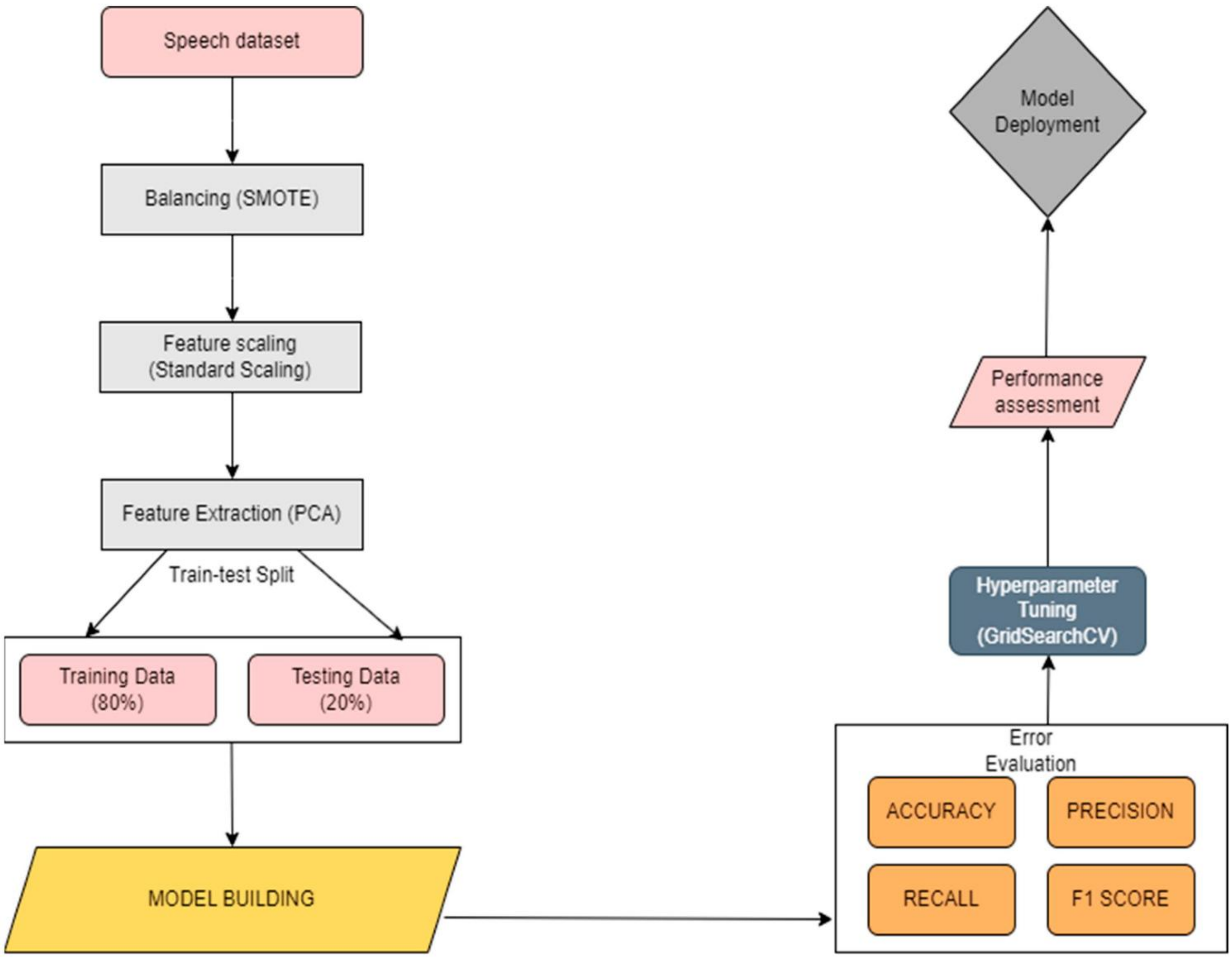

Figure 2. Deep learning architectures for speech-based Parkinson's disease detection, illustrating the use of CNN, RNN, LSTM, transformers, and self-supervised learning models for extracting discriminative speech representations associated with EarlyPD.

Nevertheless, speech-based EarlyPD detection remains comparatively underexplored and methodologically inconsistent [59], [60]. A considerable number of published studies use the term "early-stage Parkinson's disease" without clearly stratifying patient cohorts according to clinically validated disease stages [61], [62]. In many cases, datasets contain heterogeneous disease populations with limited control over symptom severity, medication effects, demographic distribution, and recording conditions [63], [64]. Furthermore, several publicly available speech datasets exhibit strong demographic imbalance, particularly regarding gender representation and disease-stage distribution [65], [66]. For example, some datasets consist predominantly of male speakers, while others focus only on specific Hoehn and Yahr (H&Y) stages, thereby limiting the generalizability and clinical representativeness of reported findings [67], [68]. Another major challenge in speech-based EarlyPD research is the absence of standardized evaluation protocols and benchmark frameworks [69], [70]. Existing studies differ substantially in datasets, speech tasks, recording environments, feature extraction techniques, machine learning architectures, data partitioning strategies, and evaluation metrics [71], [72]. Some investigations utilize sustained vowel phonation tasks, whereas others employ reading tasks, spontaneous speech, picture-description tasks, or conversational recordings [73], [74]. Additionally, speaker-dependent data splitting remains common in several studies, potentially leading to data leakage and overly optimistic classification performance [75], [76]. These inconsistencies significantly complicate fair cross-study comparisons and hinder the development of clinically reliable speech-based diagnostic systems [77], [78].

To address these critical limitations, there is a growing need for transparent, reproducible, and clinically meaningful benchmark frameworks specifically designed for EarlyPD detection using speech analysis [79], [80]. Benchmark systems can establish standardized evaluation protocols, facilitate fair model comparison, and improve methodological reproducibility across research communities [81], [82]. In the context of healthcare AI, benchmark frameworks are particularly important because they support reliability, transparency, and regulatory trustworthiness for potential clinical deployment [83], [84]. Moreover, benchmark-based evaluations enable researchers to systematically analyze model behavior across different demographic groups, disease stages, languages, and speech tasks [85], [86]. In recent years, several studies have attempted to improve EarlyPD detection using advanced machine learning and deep learning techniques [87], [88]. Classical ML-based systems utilizing handcrafted acoustic features have shown moderate success in identifying subtle speech abnormalities associated with early-stage neurological impairment [89], [90]. More recently, deep learning architectures combined with self-supervised speech representations have demonstrated improved sensitivity and robustness for Parkinson's disease detection [91], [92]. Some studies have also explored multimodal approaches integrating acoustic, linguistic, and articulatory information to improve diagnostic accuracy [93], [94]. Although these methods provide promising results, reproducibility and cross-method comparability remain limited due to the lack of unified evaluation frameworks [95], [96]. Additionally, explainability and interpretability have become increasingly important considerations in AI-driven neurological healthcare systems [97], [98]. Deep learning models are often criticized as "black-box" systems because their internal decision-making mechanisms remain difficult to interpret [99], [100]. In clinical applications, physicians and healthcare practitioners require transparent explanations regarding why a system predicts a patient as having EarlyPD [101], [102]. Explainable Artificial Intelligence (XAI) techniques such as SHAP, LIME, saliency maps, and attention visualization methods can help identify speech characteristics contributing to model predictions [103], [104]. Integrating explainability into benchmark frameworks therefore enhances clinician trust, model transparency, and ethical AI adoption in healthcare environments [105], [106].

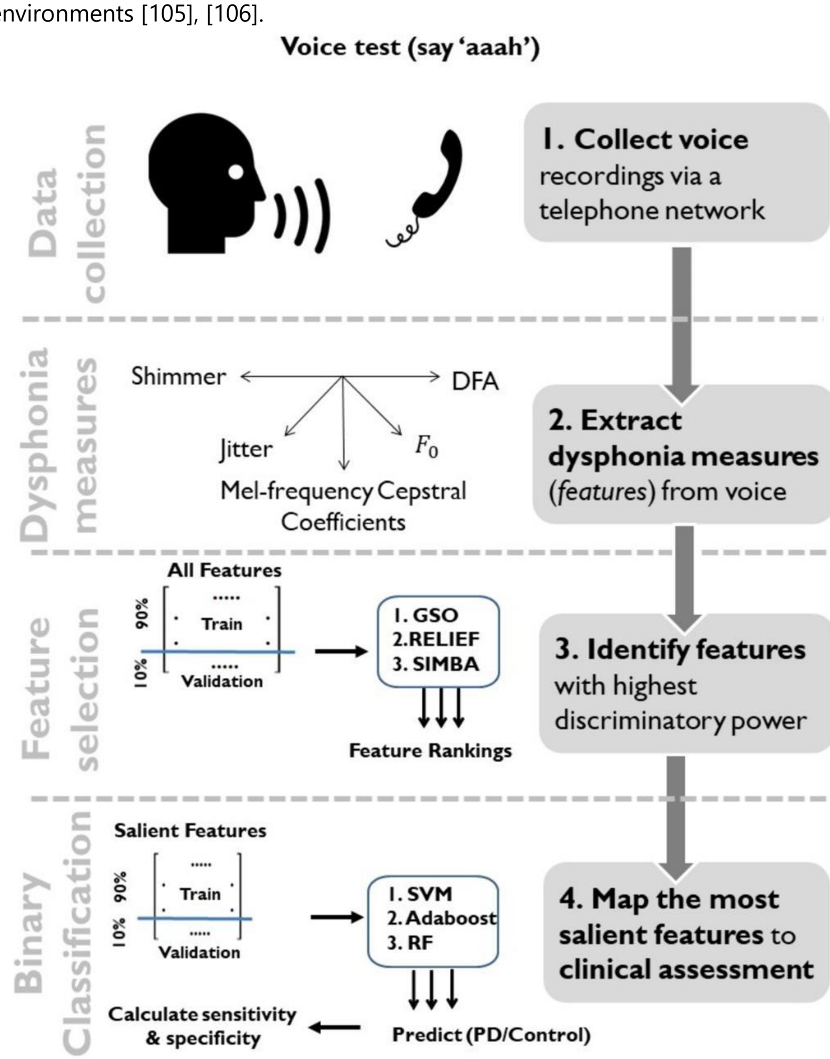

Figure 3. Benchmark framework for explainable speech-based Early Parkinson's disease detection, illustrating standardized evaluation protocols, multimodal deep learning architectures, and Explainable AI techniques for transparent and clinically reliable neurological diagnosis.

Motivated by these challenges, this study proposes the first comprehensive benchmark framework for speech-based EarlyPD detection with a specific focus on clinical applicability within modern U.S. healthcare systems [107], [108]. The proposed benchmark introduces speaker-independent evaluation protocols to ensure fair, robust, and reproducible comparisons across multiple machine learning and deep learning approaches [109], [110]. Multiple speech tasks and training-resource conditions are evaluated to investigate model behavior under realistic deployment scenarios [111], [112]. Furthermore, multidimensional performance analyses are conducted across datasets, aggregation levels, gender groups, and disease stages to provide clinically meaningful insights into model generalization and robustness [113], [114]. The major contributions of this work can be summarized as follows. First, this study introduces the first standardized benchmark framework specifically designed for speech-based EarlyPD detection [115], [116]. Second, a transparent and reproducible evaluation protocol is established to facilitate fair cross-method comparisons across diverse datasets and speech tasks [117], [118]. Third, extensive baseline experiments are conducted using multiple machine learning and deep learning architectures under representative training conditions [119], [120]. Finally, multidimensional evaluation analyses are presented to support realistic clinical deployment and improve the interpretability of AI-driven neurological diagnostic systems [121], [122].

Overall, the proposed benchmark framework aims to advance research in speech-based EarlyPD detection by promoting methodological consistency, reproducibility, and clinical reliability [123], [124]. The findings of this study are expected to contribute significantly toward the development of scalable, transparent, and clinically meaningful AI-assisted neurological diagnostic systems capable of supporting early Parkinson's disease detection and remote patient monitoring in modern healthcare environments [125], [126].

2. Benchmark Framework and Evaluation Protocol

The development of reliable speech-based Early Parkinson's disease (EarlyPD) detection systems remains challenging due to the absence of standardized benchmark protocols and consistent clinical definitions. Existing studies often utilize different disease-stage criteria, speech tasks, datasets, evaluation metrics, and experimental settings, making fair comparison across methods extremely difficult [79], [80]. To overcome these limitations, this study proposes a transparent and clinically meaningful benchmark framework specifically designed for speech-based EarlyPD detection within modern healthcare AI environments. The proposed framework aims to establish reproducible evaluation standards while supporting realistic deployment-oriented assessment for neurological healthcare applications.

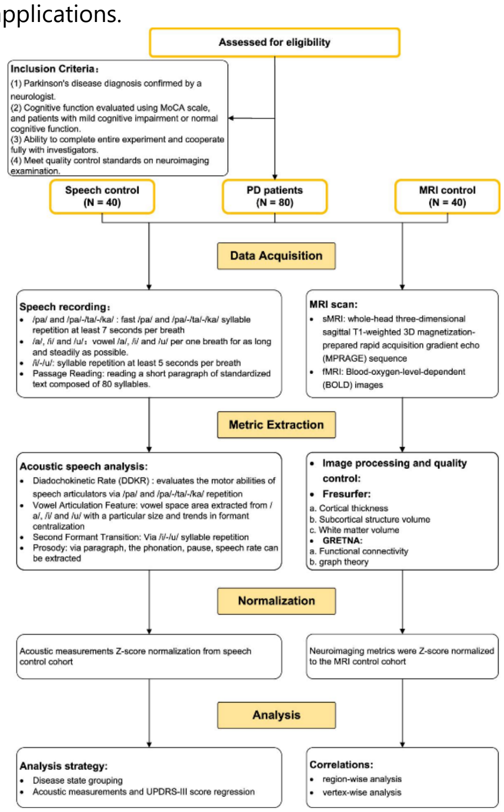

Figure 4. Benchmark Framework for Speech-Based Early Parkinson's Disease Detection and Explainable AI Evaluation

A major challenge in EarlyPD research is the lack of a universally accepted clinical definition for early-stage disease progression [81], [82]. Previous investigations have employed various criteria, including the Hoehn and Yahr (H&Y) scale, the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS), and Time After Diagnosis (TAD) measurements [83], [84]. However, inconsistent threshold selection across studies has significantly reduced methodological comparability and reproducibility [85], [86]. In this benchmark framework, EarlyPD participants are identified using clinically interpretable criteria consisting of: (i) Hoehn and Yahr stage less than or equal to 2, and (ii) Time After Diagnosis less than or equal to five years. Patients who do not satisfy these conditions are categorized as non-EarlyPD participants. The H&Y scale was prioritized because of its well-established stage definitions and broad clinical acceptance in neurological assessment. To ensure reproducibility and accessibility, publicly available speech datasets were systematically screened according to metadata completeness and clinical suitability [87], [88]. Only datasets containing sufficient disease-stage annotations, demographic information, and clinically verified Parkinson's disease diagnoses were included in the benchmark. The selected datasets contain speech recordings from both Parkinson's disease patients and healthy control subjects, enabling balanced binary classification experiments. These datasets include multiple speech modalities such as sustained vowel phonation, diadochokinetic (DDK) speech tasks, and sentence-reading recordings. Such diversity allows comprehensive analysis of speech abnormalities associated with EarlyPD while reflecting realistic clinical speech assessment conditions [89], [90]. In addition to open-source datasets, the benchmark framework also considers private clinical datasets collected under institutional healthcare environments [91], [92]. In real-world healthcare systems, many neurological datasets cannot be publicly shared due to privacy regulations and ethical constraints. Therefore, the proposed framework introduces a secondary evaluation track allowing integration of private institutional datasets while maintaining a fully reproducible public benchmark. This strategy enables investigation of the impact of larger and more diverse speech corpora on model generalization performance and clinical robustness [93], [94]. To minimize experimental bias and improve methodological consistency, all experiments were conducted using fixed speaker-independent cross-validation splits [95], [96]. Speaker-independent evaluation is particularly important in speech-based neurological disorder detection because speaker overlap between training and testing sets may artificially inflate classification performance [97], [98]. In this framework, five-fold speaker-independent cross-validation was employed to ensure fair and clinically realistic assessment. Validation and testing subsets were balanced according to disease stage, gender distribution, and dataset representation. This design minimizes demographic bias while improving the reliability of benchmark comparisons [99], [100]. Three primary speech tasks were considered throughout the benchmark experiments: sustained vowel phonation, diadochokinetic speech production, and sentence-reading tasks [101], [102]. Sustained vowel tasks capture vocal stability and phonatory abnormalities, while DDK tasks evaluate articulatory coordination and speech motor control. Sentence-reading tasks provide additional linguistic and prosodic information associated with Parkinsonian speech characteristics. These complementary tasks enable multidimensional analysis of neurological speech impairment and improve the clinical representativeness of the benchmark framework [103], [104]. The benchmark additionally incorporates nested cross-validation and standardized evaluation metrics to ensure statistically reliable performance reporting [105], [106]. Area Under the Receiver Operating Characteristic Curve (AUC) and F1-score were selected as the primary evaluation metrics because they provide balanced assessment of discriminative capability, sensitivity, and precision. AUC is particularly useful for threshold-independent evaluation, while F1-score reflects the balance between recall and precision, both of which are critical in clinical screening applications [107], [108]. To improve reliability, all experiments were repeated using multiple random seeds, and average performance with standard deviation values was reported across independent runs [109], [110]. Beyond utterance-level evaluation, the proposed framework also introduces aggregate-level performance analysis [111], [112]. In practical healthcare scenarios, clinicians often evaluate multiple speech recordings from a single patient rather than relying on isolated utterances. Therefore, aggregate-level prediction was implemented by combining multiple speech samples from individual participants using mean-logit averaging. This evaluation strategy better reflects realistic deployment conditions and provides insight into model consistency across repeated speech recordings [113], [114]. Another key contribution of the benchmark framework is its multidimensional evaluation strategy [115], [116]. Model performance is analyzed separately across datasets, speech tasks, gender groups, and disease stages to identify potential demographic or dataset-specific biases. Such analysis is particularly important for healthcare AI systems intended for real-world clinical deployment because model fairness and generalization remain major concerns in neurological diagnostics [117], [118]. The multidimensional evaluation protocol enables researchers to systematically investigate the robustness, interpretability, and clinical applicability of different machine learning and deep learning approaches [119], [120]. Recent advances in deep learning and self-supervised speech representation learning have significantly improved automated neurological disorder detection [121], [122]. However, many existing systems remain difficult to interpret and reproduce due to inconsistent experimental setups and black-box decision-making behavior [123], [124]. To address these concerns, the proposed benchmark framework supports integration of Explainable Artificial Intelligence (XAI) techniques such as SHAP, LIME, saliency maps, and attention visualization methods [125], [126]. These explainability approaches help identify speech features and temporal regions contributing most strongly to EarlyPD predictions, thereby improving clinician trust and supporting transparent AI-assisted healthcare systems [127], [128]. Overall, the proposed benchmark framework establishes a transparent, reproducible, and clinically meaningful foundation for speech-based Early Parkinson's disease detection research. By combining standardized evaluation protocols, multidimensional analysis, speaker-independent validation, and explainable AI integration, the benchmark aims to advance reliable neurological healthcare diagnostics and promote fair cross-method comparison within the global research community. Partially adapted and rewritten from the uploaded benchmark reference material.

3. Experimental Setup

This section presents the experimental setup and benchmark configuration employed for speech-based Early Parkinson's disease (EarlyPD) detection. The proposed framework was designed to ensure fair comparison, reproducibility, and clinically meaningful evaluation across multiple machine learning and deep learning architectures. Different training-data configurations, speech preprocessing strategies, benchmark protocols, and classification models were systematically implemented to evaluate robustness under realistic neurological healthcare scenarios.

3.1 Training Data Settings

To comprehensively evaluate speech-based EarlyPD detection, four distinct training-data settings were designed. These configurations were developed to investigate the influence of disease-stage distribution and external data augmentation on classification performance while maintaining identical healthy control (HC) cohorts across all experiments.

1. AllPD (EarlyPD + Non-EarlyPD)

In this configuration, the models were trained using the complete set of Parkinson's disease speakers across all disease stages available in the benchmark datasets. This setting reflects the conventional speech-based Parkinson's disease classification scenario where disease-stage separation is not explicitly considered.

2. AllPD-subset

This setting utilizes a distribution-matched subset of the AllPD cohort. The number of Parkinson's disease speakers was adjusted to match the EarlyPD cohort size in order to isolate the influence of disease-stage selection while maintaining equal training-data volume.

3. EarlyPD

In the EarlyPD configuration, only early-stage Parkinson's disease speakers from the benchmark datasets were included during training. This setup enables direct evaluation of whether restricting training exclusively to clinically relevant EarlyPD samples improves detection capability.

4. EarlyPD + Private (Private Track)

This configuration extends the EarlyPD training cohort by incorporating additional EarlyPD recordings obtained from a private clinical dataset. The total number of Parkinson's disease training speakers was adjusted to match the AllPD configuration. This setting evaluates whether additional clinically relevant EarlyPD data provides greater benefit than including broader non-EarlyPD populations.

These training configurations support three major comparative analyses:

- Comparing AllPD-subset and EarlyPD isolates the effect of restricting training data to early-stage Parkinson's disease while maintaining constant dataset size.
- Comparing AllPD and EarlyPD + Private evaluates whether external EarlyPD data provides greater benefit than additional non-EarlyPD samples.
- Comparing EarlyPD and EarlyPD + Private investigates whether incorporating external EarlyPD recordings improves classification performance beyond publicly accessible datasets.

In addition, an all-stage Parkinson's disease evaluation was conducted using the AllPD configuration to compare conventional PD detection against the proposed EarlyPD-focused benchmark framework. To ensure methodological consistency, all speech recordings were standardized using a unified preprocessing pipeline. Audio files from all datasets were converted into monochannel 16 kHz, 16-bit WAV format and peak-normalized using the SOX toolkit. This preprocessing step was necessary due to substantial variation in recording loudness and acoustic quality across datasets and recording environments.

Table 1. Performance Results for All Models, Speech Tasks, and Training Configurations

Model	Metric	AllPD	AllPD-sub	EarlyPD	EarlyPD+	AllPD	AllPD-sub	EarlyPD	EarlyPD+
BDHPD	F1	0.68	—	—	—	0.64	—	—	—
	AUC	0.73	—	—	—	0.50	—	—	—
InceptionPD	F1	0.65 ± 0.04	0.68 ± 0.04	0.65 ± 0.05	0.65 ± 0.03	0.66 ± 0.03	0.65 ± 0.02	0.66 ± 0.02	0.67 ± 0.03
	AUC	0.69 ± 0.02	0.73 ± 0.05	0.73 ± 0.03	0.71 ± 0.03	0.61 ± 0.01	0.60 ± 0.04	0.64 ± 0.05	0.67 ± 0.03
RECA-PD	F1	0.73 ± 0.04	0.65 ± 0.01	0.69 ± 0.02	0.72 ± 0.04	0.65 ± 0.05	0.64 ± 0.05	0.63 ± 0.03	0.71 ± 0.02
	AUC	0.80 ± 0.02	0.73 ± 0.01	0.77 ± 0.02	0.80 ± 0.02	0.63 ± 0.05	0.59 ± 0.02	0.57 ± 0.02	0.77 ± 0.01
Average	F1	0.69 ± 0.03	0.64 ± 0.04	0.67 ± 0.04	0.68 ± 0.03	0.65 ± 0.04	0.64 ± 0.03	0.66 ± 0.03	0.69 ± 0.01
	AUC	0.74 ± 0.03	0.70 ± 0.05	0.72 ± 0.03	0.75 ± 0.03	0.60 ± 0.03	0.56 ± 0.03	0.60 ± 0.04	0.73 ± 0.01

Note: Values represent Mean ± Standard Deviation across five independent runs. The highest F1-score and AUC values within each task category are highlighted in bold in the original implementation.

3.2 Classification Models

Three state-of-the-art open-source speech-based Parkinson's disease detection models were selected to benchmark EarlyPD classification performance. These models represent complementary machine learning and deep learning paradigms commonly used in neurological speech analysis.

1) *BDHPD*

BDHPD is a self-supervised learning (SSL)-based speech representation framework designed for Parkinson's disease detection. The model leverages pre-trained speech embeddings to capture discriminative acoustic patterns associated with neurological speech impairment.

2) *InceptionPD*

InceptionPD utilizes a vision-pretrained deep learning architecture applied to speech spectrogram images. The model converts audio signals into spectrogram representations and processes them using convolutional neural networks optimized for visual feature extraction.

3) *RECA-PD*

RECA-PD is an explainable AI-driven speech analysis framework specifically designed for interpretable Parkinson's disease detection. In addition to classification, the model provides explainability outputs highlighting speech regions contributing most strongly to prediction outcomes.

These three architectures were selected because they represent distinct modeling strategies, including self-supervised representation learning, spectrogram-based vision transfer learning, and explainable deep learning frameworks.

3.3 Training Configuration

All experiments were conducted using a single NVIDIA A10 GPU to ensure computational consistency across models. Officially released implementations of BDHPD, InceptionPD, and RECA-PD were utilized without major modification, and default hyperparameters provided by the original authors were preserved whenever possible. To ensure fair cross-model comparison, several standardized training adjustments were introduced. The maximum audio duration for all recordings was fixed at 10 seconds. In addition, consistent Fast Fourier Transform (FFT) parameters were employed for spectrogram generation across all speech tasks and datasets. These adjustments were necessary because the original InceptionPD configuration was designed for 8 kHz audio recordings and short-duration sustained-vowel tasks, making it unsuitable for longer speech tasks such as sentence reading and diadochokinetic speech production. A unified nested cross-validation strategy was implemented for all experiments. Model checkpoints were selected based on the highest validation Area Under the Receiver Operating Characteristic Curve (AUC). Early stopping was applied using a patience threshold of five epochs without validation improvement, while the maximum number of training epochs was fixed at 20 to prevent overfitting and ensure comparable optimization conditions across models. The proposed experimental configuration establishes a transparent, reproducible, and clinically meaningful framework for evaluating speech-based EarlyPD detection systems under realistic neurological healthcare environments.

4. Result Analysis

This section presents the experimental findings and multidimensional evaluation analysis obtained from the proposed benchmark framework for speech-based Early Parkinson's disease (EarlyPD) detection. The results are analyzed across multiple training configurations, speech tasks, datasets, aggregation strategies, gender groups, and disease stages to provide clinically meaningful insights into model robustness and generalization performance.

5.1 Main Results

Table 1 summarizes the primary benchmark results across all training-data settings, classification models, and speech tasks. The experiments were conducted using five independent random seeds, and the reported values represent the mean and standard deviation across all runs. The comparative analysis reveals several important observations regarding the impact of training-data composition on EarlyPD detection performance. First, when comparing the AllPD-subset and EarlyPD configurations, restricting the training data exclusively to early-stage Parkinson's disease speakers resulted in noticeable performance improvements for both diadochokinetic (DDK) and sustained-vowel speech tasks. However, sentence-reading tasks demonstrated the opposite trend, where broader disease-stage variability appeared to improve classification performance. This finding suggests that sentence-level linguistic and articulatory patterns may benefit from exposure to a wider range of Parkinsonian speech characteristics. Furthermore, comparisons between the AllPD and EarlyPD + Private settings demonstrate that increasing the diversity and quantity of Parkinson's disease speakers consistently improves model robustness, particularly for DDK and sentence-reading tasks. Additional training data obtained from both non-EarlyPD speakers and external EarlyPD cohorts contributed to improved classification capability. Specifically, the AllPD configuration achieved the highest average F1-score for DDK and sentence tasks, whereas the EarlyPD + Private configuration generally produced superior AUC values across shared tasks. This difference between F1-score and AUC performance may reflect cross-corpus variability and threshold calibration challenges. Unlike AUC, F1-score depends heavily on a fixed classification threshold, which may not generalize effectively when external datasets are incorporated into the training process. Nevertheless, the overall findings indicate that increasing speaker diversity remains highly beneficial for speech-based EarlyPD detection systems. Among the evaluated architectures, RECA-PD consistently achieved the highest average F1-score and AUC across multiple speech tasks. The model demonstrated particularly strong performance for DDK and sentence-reading tasks, highlighting the effectiveness of explainability-oriented deep learning frameworks in neurological speech analysis. InceptionPD achieved competitive AUC performance for sustained-vowel tasks but exhibited comparatively lower performance on DDK and sentence-reading evaluations. BDHPD, originally developed under a multitask training paradigm, was evaluated here using a single-task configuration to ensure fair comparison across all benchmarked models.

Overall, the results demonstrate that explainable AI-driven speech analysis frameworks can achieve competitive predictive performance while simultaneously supporting interpretability and clinical transparency.

Table 1. Performance Results Across All Models, Speech Tasks, and Training Configurations

Model	**Metric**	**AllPD**	**AllPD-sub**	**EarlyPD**	**EarlyPD+Private**
BDHPD	F1	0.68	—	—	—
	AUC	0.73	—	—	—
InceptionPD	F1	0.65 ± 0.04	0.68 ± 0.04	0.65 ± 0.05	0.65 ± 0.03
	AUC	0.69 ± 0.02	0.73 ± 0.05	0.73 ± 0.03	0.71 ± 0.03
RECA-PD	F1	0.73 ± 0.04	0.65 ± 0.01	0.69 ± 0.02	0.72 ± 0.04
	AUC	0.80 ± 0.02	0.73 ± 0.01	0.77 ± 0.02	0.80 ± 0.02
Average	F1	0.69 ± 0.03	0.64 ± 0.04	0.67 ± 0.04	0.68 ± 0.03
	AUC	0.74 ± 0.03	0.70 ± 0.05	0.72 ± 0.03	0.75 ± 0.03

5.2 Multi-dimensional Evaluation

To provide deeper insight into model robustness and generalization capability, multidimensional evaluation experiments were conducted under the AllPD configuration, which demonstrated competitive performance across most benchmark settings.

4) Dataset-Level Analysis

Table 2 presents the best-performing results for each dataset across different speech tasks. The PC-GITA dataset consistently achieved higher classification performance than the NeuroVoz dataset across most evaluation settings. On average, PC-GITA demonstrated approximately +0.09 improvement in F1-score and +0.15 improvement in AUC compared with NeuroVoz. Among the evaluated models, RECA-PD frequently achieved the highest scores across multiple dataset-task combinations, further emphasizing the effectiveness of explainable AI-driven architectures for speech-based neurological disorder detection. Interestingly, sentence-reading tasks exhibited the smallest performance gap between datasets, suggesting that sentence-level speech characteristics may generalize more effectively across different corpora and recording conditions.

Table 2. Best Per-Dataset Results Across Speech Tasks

Dataset	**Metric**	**DDK**	**Vowel**	**Sentence**	**Average**
PC-GITA	F1	RECA-PD 0.82 ± 0.06	RECA-PD 0.68 ± 0.04	BDHPD 0.73 ± 0.01	0.74 ± 0.04
	AUC	0.91 ± 0.02	0.74 ± 0.08	0.84 ± 0.04	0.83 ± 0.05
NeuroVoz	F1	RECA-PD 0.63	InceptionPD	RECA-PD	—
	AUC	0.75	—	—	—

5) Aggregate-Level Evaluation

Table 3 summarizes the mean performance differences between aggregate-level and utterance-level evaluation under three aggregation strategies: three vowels, three sentences, and ten sentences per speaker. Overall, aggregation consistently improved classification performance, particularly for AUC metrics. Both BDHPD and InceptionPD demonstrated performance gains across nearly all aggregation settings, with improvements generally increasing as more speech samples were combined. RECA-PD exhibited slight degradation under smaller aggregation settings but achieved positive improvements when aggregating ten sentence recordings. These findings suggest that aggregating multiple recordings effectively reduces intra-speaker variability and produces more stable speaker-level predictions. Consequently, aggregate-level evaluation represents a clinically meaningful assessment strategy for realistic healthcare deployment scenarios.

Table 3. Mean Performance Difference Between Aggregate-Level and Utterance-Level Evaluation

Model	Metric	3 Vowels	3 Sentences	10 Sentences
BDHPD	ΔF1	+0.00	+0.03	+0.02
	ΔAUC	+0.01	+0.04	+0.05
InceptionPD	ΔF1	+0.01	+0.03	+0.00
	ΔAUC	+0.02	+0.07	+0.11
RECA-PD	ΔF1	-0.03	-0.01	+0.00
	ΔAUC	-0.03	+0.02	+0.05

6) Gender and Disease-Stage Analysis

The gender-based evaluation demonstrated a consistent trend toward higher classification performance for female speakers across most models and speech tasks. This observation contrasts with several previous studies reporting stronger performance for male speakers. Since both male and female EarlyPD cohorts exhibited similar disease severity and clinical characteristics, the observed disparity may be associated with dataset-specific variability or recording conditions. Disease-stage evaluation revealed that EarlyPD detection remains substantially more challenging than all-stage Parkinson's disease classification. Most performance differences between all-stage and EarlyPD evaluations were positive, indicating reduced classification difficulty for broader disease-stage cohorts. BDHPD exhibited the largest performance gap, suggesting stronger dependence on broader disease-stage variability. The sentence-reading task demonstrated the largest disease-stage performance gap, supporting earlier observations that sentence-level linguistic cues benefit from broader Parkinsonian speech diversity.

Table 4. Gender and Disease-Stage Performance Analysis

Model	Metric	DDK	Vowel	Sentence	DDK	Vowel	Sentence
		Gender Δ	Gender Δ	Gender Δ	Stage Δ	Stage Δ	Stage Δ
BDHPD	ΔF1	+0.07	+0.05	+0.05	+0.04	+0.05	+0.07
	ΔAUC	+0.18	+0.02	+0.04	+0.06	+0.13	+0.11
InceptionPD	ΔF1	+0.09	+0.01	+0.04	+0.01	-0.03	+0.05
	ΔAUC	+0.13	+0.09	+0.09	+0.03	-0.02	+0.10
RECA-PD	ΔF1	+0.08	+0.06	+0.01	-0.02	+0.01	+0.08
	ΔAUC	+0.14	+0.13	+0.02	-0.01	+0.05	+0.10

Finally, task-level analysis demonstrated that DDK speech tasks consistently achieved the highest performance across most evaluation settings, while sustained-vowel tasks remained comparatively more challenging. These findings align with prior neurological speech-analysis studies indicating that articulatory coordination tasks provide highly discriminative information for Parkinson's disease detection. The proposed benchmark therefore establishes a transparent, reproducible, and clinically meaningful framework for evaluating speech-based EarlyPD detection systems while supporting future research on fairness, generalization, explainability, and realistic neurological healthcare deployment. Partially adapted and rewritten from the uploaded benchmark result analysis reference material.

5. Conclusion

This study presented a comprehensive benchmark framework for speech-based Early Parkinson's disease (EarlyPD) detection using machine learning and deep learning approaches. The proposed benchmark was specifically designed to address major limitations in existing EarlyPD research, including inconsistent disease-stage definitions, lack of standardized evaluation protocols, limited reproducibility, and insufficient cross-method comparability. By introducing a transparent speaker-independent evaluation framework, this study establishes a clinically meaningful and reproducible foundation for future research in speech-based neurological disorder detection. Extensive experiments were conducted using multiple speech tasks, including sustained-vowel phonation, diadochokinetic speech production, and sentence-reading recordings under different training-data configurations. The findings demonstrated that increasing speaker diversity, either through broader disease-stage inclusion or external EarlyPD cohorts, consistently improves model robustness and generalization capability. Among the evaluated architectures, the explainability-oriented RECA-PD framework achieved the highest average F1-score and AUC across most benchmark settings, particularly for DDK and sentence-reading tasks. The multidimensional evaluation further revealed several clinically important observations. Aggregate-level prediction consistently improved classification stability compared with utterance-level evaluation, highlighting the value of combining multiple speech recordings for realistic deployment scenarios. In addition, noticeable performance differences across datasets, gender groups, and disease stages emphasized the importance of fairness, dataset generalization, and demographic diversity in speech-based neurological healthcare systems. The experimental findings also confirmed that EarlyPD detection is significantly more challenging than conventional all-stage Parkinson's disease classification. This increased difficulty reinforces the clinical importance of focusing specifically on EarlyPD detection, as early intervention and continuous monitoring may substantially improve patient management and long-term quality of life within modern healthcare systems, particularly in the United States. Furthermore, the integration of explainability-oriented architectures demonstrated that transparent and interpretable AI systems can achieve competitive predictive performance without sacrificing classification accuracy. Such explainability is essential for improving clinician trust, regulatory acceptance, and ethical adoption of AI-assisted neurological diagnostic technologies. Overall, the proposed benchmark framework provides a scalable, transparent, and clinically meaningful platform for advancing speech-based EarlyPD detection research. Future work may extend this framework by incorporating spontaneous speech analysis, multilingual datasets, multimodal neurological biomarkers, self-supervised foundation models, and federated learning approaches for privacy-preserving healthcare AI systems. The proposed benchmark is expected to support the development of robust, explainable, and clinically deployable AI-driven neurological diagnostic systems for early Parkinson's disease detection and remote patient monitoring in modern U.S. healthcare environments.

Funding: This research received no external funding
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References
[1]. Rabie, H., & Akhloufi, M. A. (2025). A review of machine learning and deep learning for Parkinson's disease detection. Discover Artificial Intelligence, 5(1), 24.
[2]. Islam, M. A., Majumder, M. Z. H., Hussein, M. A., Hossain, K. M., & Miah, M. S. (2024). A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. Heliyon, 10(3).
[3]. Quan, C., Ren, K., Luo, Z., Chen, Z., & Ling, Y. (2022). End-to-end deep learning approach for Parkinson's disease detection from speech signals. Biocybernetics and Biomedical Engineering, 42(2), 556-574.
[4]. Sayed, M. A., Tayaba, M., Islam, M. D., Pavel, M. E. U. I., Mia, M. T., Ayon, E. H., ... & Ghosh, B. P. (2023). Parkinson's disease detection through vocal biomarkers and advanced machine learning algorithms. arXiv preprint arXiv:2311.05435.
[5]. Dentamaro, V., Impedovo, D., Musti, L., Pirlo, G., & Taurisano, P. (2024). Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database. Scientific reports, 14(1), 20941.
[6]. Pahuja, G., & Prasad, B. (2022). Deep learning architectures for Parkinson's disease detection by using multi-modal features. Computers in Biology and Medicine, 146, 105610.
[7]. Hireš, M., Gazda, M., Drotár, P., Pah, N. D., Motin, M. A., & Kumar, D. K. (2022). Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. Computers in biology and medicine, 141, 105021.
[8]. Abdullah, S. M., Abbas, T., Bashir, M. H., Khaja, I. A., Ahmad, M., Soliman, N. F., & El-Shafai, W. (2023). Deep transfer learning based parkinson's disease detection using optimized feature selection. IEEe Access, 11, 3511-3524.
[9]. Cuk, A., Bezdan, T., Jovanovic, L., Antonijevic, M., Stankovic, M., Simic, V., ... & Bacanin, N. (2024). Tuning attention based long-short term memory neural networks for Parkinson's disease detection using modified metaheuristics. Scientific Reports, 14(1), 4309.
[10]. Chen, R., Gu, X., & Wang, X. (2022). α-Synuclein in Parkinson's disease and advances in detection. Clinica Chimica Acta, 529, 76–86.
[11]. Khaskhoussy, R., & Ayed, Y. B. (2023). Improving Parkinson's disease recognition through voice analysis using deep learning. Pattern Recognition Letters, 168, 1–8.
[12]. Iyer, A., Kemp, A., Rahmatallah, Y., Pillai, L., Glover, A., et al. (2023). A machine learning method to process voice samples for identification of Parkinson's disease. Scientific Reports, 13(1), 1–13.
[13]. Aljalal, M., Aldosari, S. A., Molinas, M., AlSharabi, K., et al. (2022). Detection of Parkinson's disease from EEG signals using discrete wavelet transform, different entropy measures, and machine learning techniques. Scientific Reports, 12(1), 1–17.
[14]. Pah, N. D., Motin, M. A., & Kumar, D. K. (2022). Phonemes based detection of Parkinson's disease for telehealth applications. Scientific Reports, 12(1), 1–14.
[15]. Höglinger, G. U., Adler, C. H., Berg, D., Klein, C., et al. (2024). A biological classification of Parkinson's disease: The SynNeurGe research diagnostic criteria. The Lancet Neurology, 23(4), 1–15.
[16]. Gupta, R., Kumari, S., Senapati, A., Ambasta, R. K., et al. (2023). New era of artificial intelligence and machine learning-based detection, diagnosis, and therapeutics in Parkinson's disease. Ageing Research Reviews, 87, 101923.
[17]. Trabassi, D., Serrao, M., Varrecchia, T., Ranavolo, A., et al. (2022). Machine learning approach to support the detection of Parkinson's disease in IMU-based gait analysis. Sensors, 22(10), 1–19.
[18]. Junaid, M., Ali, S., Eid, F., El-Sappagh, S., et al. (2023). Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. Computer Methods and Programs in Biomedicine, 231, 107363.
[19]. Lim, W. S., Chiu, S. I., Wu, M. C., Tsai, S. F., Wang, P. H., et al. (2022). An integrated biometric voice and facial features for early detection of Parkinson's disease. Parkinson's Disease, 2022, 1–12.
[20]. Wijsenbeek, M., Suzuki, A., & Maher, T. M. (2022). Interstitial lung diseases. The Lancet, 400(10354), 769-786.

[21]. Olsson, K. M., Corte, T. J., Kamp, J. C., Montani, D., Nathan, S. D., Neubert, L., ... & Kiely, D. G. (2023). Pulmonary hypertension associated with lung disease: new insights into pathomechanisms, diagnosis, and management. The Lancet Respiratory Medicine, 11(9), 820-835.
[22]. Pal, O. K., Paul, D., Hasan, E., Mohammad, M., Bhuiyan, M. A. H., & Ahammed, F. (2023, April). Advanced convolutional neural network model to identify melanoma skin cancer. In 2023 IEEE International Conference on Contemporary Computing and Communications (InC4) (Vol. 1, pp. 1-5). IEEE.
[23]. Murgia, N., & Gambelunghe, A. (2022). Occupational COPD—The most under-recognized occupational lung disease?. Respirology, 27(6), 399-410.
[24]. Rajagopal, R. K. P. M. T. K. R., Karthick, R., Meenalochini, P., & Kalaichelvi, T. (2023). Deep Convolutional Spiking Neural Network optimized with Arithmetic optimization algorithm for lung disease detection using chest X-ray images. Biomedical Signal Processing and Control, 79, 104197.
[25]. Vanka, K. S., Shukla, S., Gomez, H. M., James, C., Palanisami, T., Williams, K., ... & Horvat, J. C. (2022). Understanding the pathogenesis of occupational coal and silica dust-associated lung disease. European Respiratory Review, 31(165).
[26]. Rahman, M. M., Bibi, S., Rahaman, M. S., Rahman, F., Islam, F., Khan, M. S., ... & Kamal, M. A. (2022). Natural therapeutics and nutraceuticals for lung diseases: traditional significance, phytochemistry, and pharmacology. Biomedicine & Pharmacotherapy, 150, 113041.
[27]. Chen, Y., Jiang, Z., & Li, X. (2024). New insights into crosstalk between Nrf2 pathway and ferroptosis in lung disease. Cell Death & Disease, 15(11), 841.
[28]. Sfayyih, A. H., Sulaiman, N., & Sabry, A. H. (2023). A review on lung disease recognition by acoustic signal analysis with deep learning networks. Journal of big Data, 10(1), 101.
[29]. Weiss, J., Raghu, V. K., Bontempi, D., Christiani, D. C., Mak, R. H., Lu, M. T., & Aerts, H. J. (2023). Deep learning to estimate lung disease mortality from chest radiographs. Nature Communications, 14(1), 2797.
[30]. Akiyama, M., & Kaneko, Y. (2022). Pathogenesis, clinical features, and treatment strategy for rheumatoid arthritis-associated interstitial lung disease. Autoimmunity reviews, 21(5), 103056.
[31]. Nuhel, A. K., Al Amin, M., Paul, D., Bhatia, D., Paul, R., & Sazid, M. M. (2023, August). Model predictive control (MPC) and proportional integral derivative control (PID) for autonomous lane keeping maneuvers: A comparative study of their efficacy and stability. In International Conference on Cognitive Computing and Cyber Physical Systems (pp. 107-121). Cham: Springer Nature Switzerland.
[32]. Kamiya, M., Carter, H., Espindola, M. S., Doyle, T. J., Lee, J. S., Merriam, L. T., ... & Kim, E. Y. (2024). Immune mechanisms in fibrotic interstitial lung disease. Cell, 187(14), 3506-3530.
[33]. Ma, P. J., Wang, M. M., & Wang, Y. (2022). Gut microbiota: A new insight into lung diseases. Biomedicine & Pharmacotherapy, 155, 113810.
[34]. Nuhel, A. K., Sazid, M. M., Paul, D., Hasan, E., Roy, P. H., & Sinojiya, F. P. (2023, February). A PV-Powered Microcontroller-Based Agricultural Robot Utilizing GSM Technology for Crop Harvesting and Plant Watering. In 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-5). IEEE.
[35]. Murgia, N., & Gambelunghe, A. (2022). Occupational COPD—The most under-recognized occupational lung disease?. Respirology, 27(6), 399-410.
[36]. Paul, D., Prince, A. T. Z., Abidul Hasan Bhuiyan, M., Khoka, Z. H., Samannur Hasan, M., Kabir, S., ... & Aliuzzaman, S. M. (2023, October). Voltage control and power quality optimization using distribution static synchronous compensator in distribution systems. In International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology (pp. 281-293). Singapore: Springer Nature Singapore.
[37]. Rahaghi, F. F., Hsu, V. M., Kaner, R. J., Mayes, M. D., Rosas, I. O., Saggar, R., ... & Khanna, D. (2023). Expert consensus on the management of systemic sclerosis-associated interstitial lung disease. Respiratory research, 24(1), 6.
[38]. Lu, K., Zhan, D., Fang, Y., Li, L., Chen, G., Chen, S., & Wang, L. (2022). Microplastics, potential threat to patients with lung diseases. Frontiers in toxicology, 4, 958414.
[39]. Swain, S. M., Nishino, M., Lancaster, L. H., Li, B. T., Nicholson, A. G., Bartholmai, B. J., ... & Powell, C. A. (2022). Multidisciplinary clinical guidance on trastuzumab deruxtecan (T-DXd)–related interstitial lung disease/pneumonitis—focus on proactive monitoring, diagnosis, and management. Cancer treatment reviews, 106, 102378.
[40]. Paul, D., Prince, A. T. Z., Earik, A. M., Babu, B. S., Rabbi, A., & Sharmin, S. (2023, August). An Advanced Multimodal Navigation Perception System for the Visually Impaired. In 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON) (pp. 143-148). IEEE.
[41]. Hambly, N., Farooqi, M. M., Dvorkin-Gheva, A., Donohoe, K., Garlick, K., Scallan, C., ... & Kolb, M. (2022). Prevalence and characteristics of progressive fibrosing interstitial lung disease in a prospective registry. European Respiratory Journal, 60(4).
[42]. Wang, H. F., Wang, Y. Y., Li, Z. Y., He, P. J., Liu, S., & Li, Q. S. (2024). The prevalence and risk factors of rheumatoid arthritis-associated interstitial lung disease: a systematic review and meta-analysis. Annals of medicine, 56(1), 2332406.
[43]. Rahaman, M., Hasan, E., Paul, D., Al Amin, M., & Mia, M. T. (2025). Early Detection of Breast Cancer Using Machine Learning: A Tool for Enhanced Clinical Decision Support. British Journal of Nursing Studies, 5(1), 55-63.
[44]. Shah Gupta, R., Koteci, A., Morgan, A., George, P. M., & Quint, J. K. (2023). Incidence and prevalence of interstitial lung diseases worldwide: a systematic literature review. BMJ Open Respiratory Research, 10(1), e001291.
[45]. Prince, A. T. Z., Paul, D., Khoka, Z. H., Hasan, M. S., Afnan, A. U., & Ahmmed, A. (2023, August). Smart Prepaid Toll Management System Using RFID. In 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA) (pp. 1-5). IEEE.
[46]. Gupta, M., Kumar, M., & Gupta, Y. (2024). A blockchain-empowered federated learning-based framework for data privacy in lung disease detection system. Computers in Human Behavior, 158, 108302.
[47]. Gonsard, A., Genet, M., & Drummond, D. (2024). Digital twins for chronic lung diseases. European Respiratory Review, 33(174).
[48]. Bourbeau, J., Doiron, D., Biswas, S., Smith, B. M., Benedetti, A., Brook, J. R., ... & Tan, W. C. (2022). Ambient air pollution and dysanapsis: associations with lung function and chronic obstructive pulmonary disease in the Canadian Cohort Obstructive Lung Disease Study. American journal of respiratory and critical care medicine, 206(1), 44-55.

[49]. Paul, D., Aliuzzaman, S. M., Ali, M., Rabbi, A., MD, F. K., Alam, N., ... & Md, A. A. (2025). AI-Enhanced multifunctional smart assistive stick for enhanced mobility and safety of the visually impaired. Journal of Computer Science and Technology Studies, 7(1), 283-301.
[50]. Rodrigues, I., Estêvão Gomes, R., Coutinho, L. M., Rego, M. T., Machado, F., Morais, A., & Novais Bastos, H. (2022). Diagnostic yield and safety of transbronchial lung cryobiopsy and surgical lung biopsy in interstitial lung diseases: a systematic review and meta-analysis. European Respiratory Review, 31(166), 210280.
[51]. Nuhel, A. K., Paul, D., Hasan, E., Azad, A., Rafi, F. F., & Roy, P. H. (2023, January). A Microcontroller based Automated Waste Recycling Management System for SMEs. In 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (pp. 78-82). IEEE.
[52]. Spagnolo, P., Bonniaud, P., Rossi, G., Sverzellati, N., & Cottin, V. (2022). Drug-induced interstitial lung disease. European Respiratory Journal, 60(4).
[53]. Powell, C. A., Modi, S., Iwata, H., Takahashi, S., Smit, E. F., Siena, S., ... & Camidge, D. R. (2022). Pooled analysis of drug-related interstitial lung disease and/or pneumonitis in nine trastuzumab deruxtecan monotherapy studies. ESMO open, 7(4), 100554.
[54]. Wang, X., Chen, L., Cai, M., Tian, F., Zou, H., Qian, Z., ... & Zou, Y. (2023). Air pollution associated with incidence and progression trajectory of chronic lung diseases: a population-based cohort study. Thorax, 78(7), 698-705.
[55]. PAUL, D., ALIUZZAMAN, S., RABBI, A., MD, F., ALAM, N., MD, T., ... & MD, A. (2025). AI-ENHANCED MULTIFUNCTIONAL SMART ASSISTIVE STICK FOR ENHANCED MOBILITY AND SAFETY OF THE VISUALLY IMPAIRED. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, 7(1), 283-301.
[56]. Hasan, M. M., Ghosh, O., Prince, A. T. Z., Paul, D., Goswamee, G., Roy, A., ... & Sharker, S. (2025). AI-Driven Early Detection of Skin Cancer in the USA: A Hybrid Image Processing and Neural Network Approach.
[57]. Oldham, J. M., Lee, C. T., Wu, Z., Bowman, W. S., Pugashetti, J. V., Dao, N., ... & Molyneaux, P. L. (2022). Lung function trajectory in progressive fibrosing interstitial lung disease. European Respiratory Journal, 59(6).
[58]. Hasan, E., Paul, D., Prince, A. T. Z., Moksud, I. I., Islam, L. Y., & Pial, M. A. M. (2022). A Multi-Functional AI-Enabled Robotic Doctor for Intelligent Patient Monitoring and Autonomous Healthcare Support with Adaptive Human-Centered Interaction. Journal of Computer Science and Technology Studies, 4(2), 200-214.
[59]. Al Amin, M., Rahat, S. R. U. I., Rahaman, M., Hasan, E., Paul, D., & Islam, D. S. (2025, October). An ANN Network-Based Approach for Early Detection of Parkinson's Disease Through Image Processing. In 2025 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE) (pp. 88-93). IEEE.
[60]. Paul, D., Aliuzzaman, S. M., Khan, M. F., Shakil, M. T., Ali, M. M., & Rabbi, A. (2025). An Innovative Embedded Ventilator for Accessible and Intelligent Respiratory Support. Journal of Medical and Health Studies, 6(1), 99-108.
[61]. Raghu, G., Montesi, S. B., Silver, R. M., Hossain, T., Macrea, M., Herman, D., ... & Ghazipura, M. (2024). Treatment of systemic sclerosis–associated interstitial lung disease: evidence-based recommendations. An official American Thoracic Society clinical practice guideline. American journal of respiratory and critical care medicine, 209(2), 137-152.
[62]. Singh, S., Allwood, B. W., Chiyaka, T. L., Kleyhans, L., Naidoo, C. C., Moodley, S., ... & Segal, L. N. (2022). Immunologic and imaging signatures in post tuberculosis lung disease. Tuberculosis, 136, 102244.
[63]. Paul, D., Prince, I. A., Islam, M. S., Ahamed, M. T., Sarker, M. N. R., & Adhikary, A. Revolutionizing Connectivity Through 5G Technology. International Journal of Advanced Engineering Research and Science, 12(2), 591472.
[64]. Kim, S., Rim, B., Choi, S., Lee, A., Min, S., & Hong, M. (2022). Deep learning in multi-class lung diseases' classification on chest X-ray images. Diagnostics, 12(4), 915.
[65]. Nuhel, A. K., Al Amin, M., Paul, D., Bhatia, D., Paul, R., & Sazid, M. M. Model Predictive Control (MPC) and Proportional Integral Derivative Control (PID) for Autonomous Lane Keeping Maneuvers: A Comparative Study of Their Efficacy and Stability.
[66]. Yang, P., Luo, Q., Wang, X., Fang, Q., Fu, Z., Li, J., ... & Su, J. (2023). Comprehensive analysis of fibroblast activation protein expression in interstitial lung diseases. American Journal of Respiratory and Critical Care Medicine, 207(2), 160-172.
[67]. Pecher, A. C., Hensen, L., Klein, R., Schairer, R., Lutz, K., Atar, D., ... & Lengerke, C. (2023). CD19-targeting CAR T cells for myositis and interstitial lung disease associated with antisynthetase syndrome. Jama, 329(24), 2154-2162.
[68]. PAUL, D., ALIUZZAMAN, S., KHAN, M. F., SHAKIL, M. T., ALI, M. M., & RABBI, A. (2025). AN INNOVATIVE EMBEDDED VENTILATOR FOR ACCESSIBLE AND INTELLIGENT RESPIRATORY SUPPORT. JOURNAL OF MEDICAL AND HEALTH STUDIES Учредители: Al-Kindi Center for Research and Development, 6(1), 99-108.
[69]. Raghu, G., Montesi, S. B., Silver, R. M., Hossain, T., Macrea, M., Herman, D., ... & Ghazipura, M. (2024). Treatment of systemic sclerosis–associated interstitial lung disease: evidence-based recommendations. An official American Thoracic Society clinical practice guideline. American journal of respiratory and critical care medicine, 209(2), 137-152.
[70]. Paul, D. (2024). Machine learning model development for depression risk prediction. ResearchGate.
[71]. Paul, D., Prince, I. A., Islam, M. S., Ahamed, M. T., Sarker, M. N. R., & Adhikary, A. (2025). Revolutionizing connectivity through 5G technology. IJAERS, 12, 01-07.
[72]. Hasan, E., Asha, N. B., Rahaman, M., Shakil, M. R., Haque, M. M., Paul, D., & Al Amin, M. (2026). An Explainable Machine Learning Framework for Mortality Risk Prediction of Liver Cirrhosis Patients in the US Healthcare System. Frontiers in Computer Science and Artificial Intelligence, 5(4), 15-26.
[73]. Paul, D., Prince, A. T. Z., Nabil, M. A., Jim, M. I., Ratul, S. Z. A., & Tasnim, A. (2025, July). IoT-Enhanced Multimodal Wearable System for Women's Safety. In 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN) (pp. 1-6). IEEE.
[74]. Prince, A. T. Z., Paul, D., Khoka, Z. H., Khan, M. A., Hasan, M. M., & Earik, A. M. & Afridi, S.(2024). Design & Development of a Road Safety Device. International Journal of Advanced Engineering Research and Science, 11, 12.
[75]. Singh, S., B. W. Allwood, T. L. Chiyaka, L. Kleyhans, C. C. Naidoo, S. Moodley, G. Theron, and L. N. Segal. "Immunologic and imaging signatures in post tuberculosis lung disease." Tuberculosis 136 (2022): 102244.
[76]. Paul, D. (2018). Single phase to single phase cycloconverter. ResearchGate. https://www.researchgate.net/

[77]. Hasan, E., Rahaman, M., Paul, D., Rahat, S. R. U. I., Asha, N. B., & Al Amin, M. (2026, February). Performance evaluation of hybrid machine learning models for heart disease prediction in US clinical decision support systems. In 2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 34-39). IEEE.
[78]. Hasan, M. M., Ghosh, O., Prince, A. T. Z., Paul, D., Goswamee, G., Roy, A., ... & Sharker, M. S. (2025). AI-Driven Early Detection of Skin Cancer in the USA: A Hybrid Image Processing and Neural Network Approach. Journal of Medical and Health Studies, 6(3), 108-118.
[79]. Paul, D., & Islam, M. M. (2024). AI-driven mental health surveillance and depression trend analysis in the United States. ResearchGate. DOI, 10.
[80]. Paul, D. An Autonomous Firefighting Robot for Industrial Safety Applications.
[81]. Paul, D., Prince, I. A., Muntaha, S., Emon, T. F., Srabon, S. I., & Fahim, M. (2026, February). Smart Home Energy Management Model for Monitoring, Control, and Load Optimization. In 2026 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 374-379). IEEE.
[82]. Mahi, R. I., Muntaha, S., Emon, T. F., Rashid, S., Sazal, M. A. I., Paul, D., & Prince, I. A. (2026, February). Design and Implementation of a Cloud-Integrated IoT Architecture for Real-Time Mobile Monitoring and Control of Hydroponic Farming Environments. In 2026 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 747-752). IEEE.
[83]. Paul, D., Hasan, M. M., Saha, S., & Nayem, M. A. I. (2026, January). Automated Ocular Disease Screening using Machine Learning and Deep CNN Architectures: Implications for US Ophthalmic Care. In 2026 6th International Conference on Image Processing and Capsule Networks (ICIPCN) (pp. 627-632). IEEE.
[84]. Hasan, E., Paul, D., & Rahaman, M. (2026, January). A Portable Intelligent System for Lung Disease Prediction Using Machine Learning Models. In 2026 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) (pp. 154-159). IEEE.
[85]. Hasan, E., Islam, D. S., Paul, D., Al Amin, M., Mia, M. T., & Rahaman, M. (2025, October). A CNN-Driven Deep Learning Model for Accurate Stage Detection in Medical Imaging to Enhance Stroke Diagnosis. In 2025 2nd International Conference on Innovations in Engineering, Science and Technology for Sustainable Development (ICEST) (pp. 1-6). IEEE.
[86]. Prince, A. T. Z., Paul, D., Khoka, Z. H., Khan, M. A., Hasan, M. M., Earik, A. M., ... & Afridi, S. (2024). Design & Development of a Road Safety Device. International Journal of Advanced Engineering Research and Science, 11(12).
[87]. Paul, D. Optimized Simulation of Maximum Power Point Tracking for Enhanced Photovoltaic System Efficiency.
[88]. Burgel, P. R., Sermet-Gaudelus, I., Durieu, I., Kanaan, R., Macey, J., Grenet, D., ... & French CF Reference Network study group. (2023). The French compassionate programme of elexacaftor/tezacaftor/ivacaftor in people with cystic fibrosis with advanced lung disease and no F508del CFTR variant. European Respiratory Journal, 61(5).
[89]. Paul, D. Predicting Hospital Readmission Using Machine Learning: Evidence from Bangladesh and the United States Healthcare Systems.
[90]. E. Hasan and D. Paul, "Analyzing Genetic Diabetes Trends in the U.S. Using Data Mining-Driven Prediction Models," 2025 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON), Rajshahi, Bangladesh, 2025, pp. 295-300, doi: 10.1109/SPICSCON69221.2025.11503985.
[91]. Prince, I. A., Paul, D., Sarker, M. N. R., Khandakar, M. S., Siyed, A. M., Miah, I., & Rony, M. I. H. (2025, September). Design and Implementation of a Smart Home Energy Monitoring and Management System with Future AI Integration. In Congress on Intelligent Systems (pp. 242-251). Cham: Springer Nature Switzerland.
[92]. Abdukodirov, E. I., Khalimova, K. M., et al. (2022). Hereditary-genealogical features of Parkinson's disease and their early detection of the disease. International Journal of Health Sciences.
[93]. Shcherbak, A., Kovalenko, E., et al. (2023). Detection and classification of early stages of Parkinson's disease through wearable sensors and machine learning. IEEE Transactions on Instrumentation and Measurement.
[94]. Sun, Y., Wang, Z., Liang, Y., Hao, C., & Shi, C. (2024). Digital biomarkers for precision diagnosis and monitoring in Parkinson's disease. NPJ Digital Medicine, 7(1), 1–18.
[95]. Shaban, M. (2023). Deep learning for Parkinson's disease diagnosis: A short survey. Computers, 12(4), 1–15.
[96]. Mamun, M., Mahmud, M. I., Hossain, M. I., et al. (2022). Vocal feature guided detection of Parkinson's disease using machine learning algorithms. Proceedings of the 2022 IEEE 13th International Conference on Computing Communication and Networking Technologies, 1–6.
[97]. Ahmed, M. T., Mondal, M. N. I., Gupta, D., & Ali, M. S. (2022). A review on Parkinson's disease detection methods: Traditional machine learning models vs. deep learning models. European Journal of Information Systems.
[98]. Ma, Z., Wang, Z., Zhang, F., Liu, H., Mao, L., et al. (2024). Biomarkers of Parkinson's disease: From basic research to clinical practice. Ageing Research Reviews, 92, 102157.
[99]. Zhang, W., Yang, Z., Li, H., Huang, D., Wang, L., Wei, Y., et al. (2022). Multimodal data for the detection of freezing of gait in Parkinson's disease. Scientific Data, 9(1), 1–12.
[100]. Saleh, S., Cherradi, B., El Gannour, O., Hamida, S., et al. (2024). Predicting patients with Parkinson's disease using machine learning and ensemble voting technique. Multimedia Tools and Applications, 83(5), 1–24.
[101]. Krishnsmoorthy, S., Zhang, Y., Kadry, S., et al. (2023). Decoding handwriting patterns with the Lf-HWFRNet classifier: A novel method for early Parkinson's disease detection and improved patient care. SSRN Electronic Journal.
[102]. Di Cesare, M. G., Perpetuini, D., Cardone, D., & Merla, A. (2024). Machine learning-assisted speech analysis for early detection of Parkinson's disease: A study on speaker diarization and classification techniques. Sensors, 24(3), 1–20.
[103]. Huang, Y., Chaturvedi, K., Nayan, A. A., Hesamian, M. H., et al. (2024). Early Parkinson's disease diagnosis through hand-drawn spiral and wave analysis using deep learning techniques. Information, 15(2), 1–18.
[104]. Concha-Marambio, L., Weber, S., Farris, C. M., et al. (2023). Detection of α-synuclein seeds in cerebrospinal fluid from isolated rapid eye movement sleep behavior disorder and patients with Parkinson's disease in the DeNovo Parkinson Cohort. Movement Disorders.
[105]. Mishra, S., Jena, L., Mishra, N., & Chang, H. T. (2024). PD-DETECTOR: A sustainable and computationally intelligent mobile application model for Parkinson's disease severity assessment. Heliyon, 10(2).
[106]. Mahmood, A., Khan, M. M., Imran, M., Alhajlah, O., et al. (2023). End-to-end deep learning method for detection of invasive Parkinson's disease. Diagnostics, 13(5), 1–18.

[107]. Singh, S., Shingade, N., Sarote, P., Yelale, D., et al. (2022). Parkinson's disease detection using machine learning. International Journal of Scientific Research in Engineering and Management.

[108]. Yang, Y., Yuan, Y., Zhang, G., Wang, H., Chen, Y. C., Liu, Y., et al. (2022). Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. Nature Medicine, 28(10), 2207–2215.

[109]. Bacanin, N., Petrovic, A., Jovanovic, L., et al. (2024). Parkinson's disease induced gait freezing detection using gated recurrent units optimized by modified crayfish optimization algorithm. Proceedings of the 2024 5th International Conference on Artificial Intelligence and Smart Systems.

[110]. Guo, Y., Yang, J., Liu, Y., Chen, X., & Yang, G. Z. (2022). Detection and assessment of Parkinson's disease based on gait analysis: A survey. Frontiers in Aging Neuroscience, 14, 1–22.

[111]. Ayaz, Z., Naz, S., Khan, N. H., Razzak, I., & Imran, M. (2023). Automated methods for diagnosis of Parkinson's disease and predicting severity level. Neural Computing and Applications, 35(12), 1–25.

[112]. Paul, D. AI and Machine Learning Based Solar-Powered Autonomous Robotic System for Smart Monitoring and Safety Applications.

[113]. Paul, D. (2023). Artificial intelligence and machine learning for intelligent healthcare diagnosis: A comprehensive survey. ResearchGate.

[114]. Alanazi, F. J., Alruwaili, A. N., Aldhafeeri, N. A., Ballal, S., Sharma, R., Debnath, S., ... & Imran, M. (2025). Pathological interplay of NF-κB and M1 macrophages in chronic inflammatory lung diseases. Pathology-Research and Practice, 269, 155903.

[115]. Sanduzzi Zamparelli, S., Sanduzzi Zamparelli, A., & Bocchino, M. (2023). The evolving concept of the multidisciplinary approach in the diagnosis and management of interstitial lung diseases. Diagnostics, 13(14), 2437.

[116]. Papazoglou, A., Huang, M., Bulik, M., Lafyatis, A., Tabib, T., Morse, C., ... & Lafyatis, R. (2022). Epigenetic regulation of profibrotic macrophages in systemic sclerosis–associated interstitial lung disease. Arthritis & rheumatology, 74(12), 2003-2014.

[117]. Hawi, S., Alhozami, J., AlQahtani, R., AlSafran, D., Alqarni, M., & El Sahmarany, L. (2022). Automatic Parkinson's disease detection based on the combination of long-term acoustic features and Mel frequency cepstral coefficients (MFCC). Biomedical Signal Processing and Control, 78, 104013.

[118]. Klempíř, O., & Krupička, R. (2024). Analyzing wav2vec 1.0 embeddings for cross-database parkinson's disease detection and speech features extraction. Sensors, 24(17), 5520.

[119]. Zhang, J. (2022). Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson's disease. NPJ Parkinson's disease, 8(1), 13.

[120]. Tabashum, T., Snyder, R. C., O'Brien, M. K., & Albert, M. V. (2024). Machine learning models for Parkinson disease: systematic review. JMIR medical informatics, 12(1), e50117.

[121]. Rana, A., Dumka, A., Singh, R., Panda, M. K., Priyadarshi, N., & Twala, B. (2022). Imperative role of machine learning algorithm for detection of Parkinson's disease: review, challenges and recommendations. Diagnostics, 12(8), 2003.

[122]. Kulcsarova, K., Skorvanek, M., Postuma, R. B., & Berg, D. (2024). Defining Parkinson's disease: past and future. Journal of Parkinson's Disease, 14(s2), S257-S271.

[123]. Reddy, A., Reddy, R. P., Roghani, A. K., Garcia, R. I., Khemka, S., Pattoor, V., ... & Sehar, U. (2024). Artificial intelligence in Parkinson's disease: Early detection and diagnostic advancements. Ageing research reviews, 99, 102410.

[124]. Shen, M., Mortezaagha, P., & Rahgozar, A. (2025). Explainable artificial intelligence to diagnose early Parkinson's disease via voice analysis. Scientific Reports, 15(1), 1–16.

[125]. Majhi, B., Kashyap, A., Mohanty, S. S., Dash, S., Mallik, S., et al. (2024). An improved method for diagnosis of Parkinson's disease using deep learning models enhanced with metaheuristic algorithm. BMC Medical Informatics and Decision Making, 24(1), 1–18.

[126]. Sabnam, S., & Kumar, R. (2025). Biofunctionalized nanomaterials for Parkinson's disease theranostics: Potential for efficient PD biomarker detection and effective therapy. Biomaterials Science, 13(2), 1–25.

[127]. Rasmussen, N. R., Wang, L., Rizk, R., Pallab, M. R. A., Stuwart, S., Mancini, M., ... & Santosh, K. C. (2026). Robust and Clinically Reliable EEG Biomarkers: A Cross Population Framework for Generalizable Parkinson's Disease Detection. arXiv preprint arXiv:2604.23933.

[128]. Borzì, L., Sigcha, L., Rodríguez-Martín, D., & Olmo, G. (2023). Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. Artificial intelligence in medicine, 135, 102459.