electronics

MDPI

Article

A Confidence-Gated Hybrid CNN Ensemble for Accurate Detection of Parkinson's Disease Using Speech Analysis

Salem Titouni [1], Nadhir Djeffal [1,2], Massinissa Belazzoug [1], Boualem Hammache [1], Idris Messaoudene [1] and Abdallah Hedir [3,4,*]

[1] ETA Laboratory, University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, Bordj Bou Arreridj 34000, Algeria; salem.titouni@univ-bba.dz (S.T.); nadhir.djeffal@ummto.dz (N.D.); m.belazzoug@univ-bba.dz (M.B.); boualem_hammache@hotmail.fr (B.H.); idris.messaoudene@univ-bba.dz (I.M.)
[2] Faculty of Electrical Engineering and Computer Science, University of Mouloud MAMMERI, Tizi-Ouzou 15000, Algeria
[3] Advanced High Voltage Engineering Centre, School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, UK
[4] Laboratoire de Technologies Avancées en Génie Electrique, University of Mouloud MAMMERI, Tizi-Ouzou 15000, Algeria
* Correspondence: abdallah.hedir@ummto.dz or hedira1@cardiff.ac.uk

Abstract

Parkinson's Disease (PD) is a progressive neurodegenerative disorder for which early and reliable diagnosis remains challenging. To address this challenge, the key innovation of this work is a confidence-gated fusion mechanism that dynamically weights classifier outputs based on per-sample prediction certainty, overcoming the limitations of static ensemble strategies. Building on this idea, we propose a Confidence-Gated Hybrid CNN Ensemble that integrates CNN-based acoustic feature extraction with heterogeneous classifiers, including XGBoost, Support Vector Machines, and Random Forest. By adaptively modulating the contribution of each classifier at the sample level, the proposed framework enhances robustness against data imbalance, inter-speaker variability, and feature complexity. The method is evaluated on two benchmark PD speech datasets, where it consistently outperforms conventional machine learning and ensemble approaches, achieving a best classification accuracy of up to 97.9% while maintaining computational efficiency compatible with real-time deployment. These results highlight the effectiveness and clinical potential of confidence-aware ensemble learning for non-invasive PD detection.

Keywords: Parkinson's Disease; speech analysis; deep learning; ensemble learning; confidence-gated fusion; CNN

Academic Editor: Enzo Pasquale Scilingo

Received: 21 December 2025
Revised: 22 January 2026
Accepted: 24 January 2026
Published: 29 January 2026

1. Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder affecting millions of individuals worldwide, for which early diagnosis is essential to enable timely intervention and improve patient outcomes [1,2]. Despite advances in clinical practice, PD diagnosis remains challenging due to the reliance on subjective assessments and the subtle manifestation of early-stage symptoms, often leading to delayed or inaccurate diagnosis [3]. Among potential early biomarkers, speech impairments provide a non-invasive and accessible diagnostic pathway, as PD-related dysarthria induces measurable changes in acoustic characteristics such as pitch, jitter, shimmer, and articulation dynamics [4,5].

Recent progress in machine learning (ML) and deep learning (DL) has enabled automated analysis of speech signals for PD detection. In particular, Convolutional Neu-

ral Networks (CNNs) have demonstrated strong capability in learning discriminative representations from speech-derived features, frequently outperforming traditional approaches based on hand-crafted features [6,7]. However, CNN-based models alone may exhibit limited robustness when confronted with small sample sizes, class imbalance, inter-speaker variability, and noisy recording conditions—common characteristics of PD speech datasets [8]. These limitations motivate the integration of CNNs with complementary classification paradigms.

Ensemble learning has therefore emerged as an effective strategy to enhance generalization by combining classifiers with diverse decision boundaries. By leveraging complementary strengths, ensemble methods can reduce overfitting and improve robustness in speech-based PD detection. Nevertheless, most existing ensemble approaches rely on static fusion strategies, implicitly assuming equal reliability of all classifiers across all samples. This assumption is unrealistic in clinical speech analysis, where prediction confidence can vary significantly depending on speaker characteristics, recording quality, and disease severity.

The key innovation of this work is a confidence-gated fusion mechanism that dynamically weights classifier outputs based on per-sample prediction certainty, thereby addressing the inherent limitations of static ensemble strategies. Building on this principle, we propose a Confidence-Gated Hybrid CNN Ensemble that integrates CNN-based acoustic feature extraction with heterogeneous classifiers, including XGBoost, Support Vector Machines (SVM), and Random Forest (RF). A gating network adaptively assigns reliability weights to classifier outputs using confidence indicators, enabling sample-aware decision fusion and improved robustness.

Despite recent advances, several challenges remain in speech-based PD detection, including limited dataset sizes, difficulties in generalizing across heterogeneous populations, and the need for scalable and computationally efficient solutions suitable for real-time deployment [9]. The proposed confidence-gated framework directly addresses these challenges by combining deep feature learning, ensemble diversity, and adaptive fusion within a unified architecture.

The main contributions of this work are summarized as follows:

- Confidence-Gated Hybrid Architecture: We propose a novel ensemble framework that combines CNN-based feature extraction with heterogeneous classifiers and a confidence-gated fusion mechanism for speech-based PD detection.
- Adaptive Per-Sample Decision Fusion: A gating network dynamically weights classifier outputs according to prediction confidence, improving robustness over conventional static ensemble strategies.
- Comprehensive Experimental Evaluation: The proposed method is evaluated on two widely used benchmark PD speech datasets, demonstrating superior performance compared to state-of-the-art machine learning and ensemble approaches.
- Scalability and Clinical Relevance: The framework provides a non-invasive, efficient, and scalable solution with strong potential for real-time and clinical deployment.

2. Related Work

Speech-based diagnosis of PD has attracted significant research interest due to the strong correlation between vocal impairments and motor dysfunctions. Existing studies employ a wide range of ML and DL techniques to improve diagnostic accuracy. Broadly, prior work can be categorized into feature selection and signal processing-based approaches, deep learning architectures, and ensemble or hybrid learning strategies. While substantial progress has been made, persistent challenges remain in robustness, generalization, and adaptive decision fusion.

2.1. Feature Selection and Signal Processing-Based Approaches

Early research efforts primarily focused on identifying discriminative acoustic features for PD detection. Wrapper and filter-based feature selection techniques have been widely adopted to reduce dimensionality and improve classification performance. For instance, ref. [10] employed wrapper-based feature selection in conjunction with SVM, kNN, MLP, and Random Forest (RF), achieving high accuracy with a reduced feature set. Similarly, genetic and evolutionary optimization strategies combined with classical classifiers were explored in [4,11]. Although effective, such approaches are often computationally intensive and highly dataset-dependent, limiting their scalability and robustness across heterogeneous speech corpora.

Advanced signal processing techniques have also been investigated to enhance feature extraction. Tunable Q-factor Wavelet Transform (TQWT) combined with MFCCs was proposed in [3], offering improved frequency resolution but suffering from limited modality coverage and sensitivity to recording conditions. Dimensionality reduction methods such as weighted Local Discriminant Preservation Projection [1] and PCA-based pipelines [9] addressed class imbalance and redundancy but struggled to generalize to larger and more diverse datasets. More recent nature-inspired optimization algorithms, including the Zebra Optimization Algorithm (ZOA) [12] and the Multi-Agent Salp Swarm (MASS) algorithm [13], further refined feature selection efficiency at the cost of increased computational complexity, making real-time deployment challenging.

2.2. Deep Learning Architectures for PD Speech Analysis

Deep learning has significantly advanced speech-based PD diagnosis by enabling automatic representation learning from acoustic data. Convolutional Neural Networks (CNNs) have been widely adopted to capture discriminative patterns in speech-derived features. A parallel CNN architecture was proposed in [6], achieving competitive performance but exhibiting limited scalability across datasets. Temporal modeling approaches using Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) networks were explored in [8,14], effectively capturing sequential dependencies in speech signals. However, such architectures often require substantial computational resources and large training datasets, which limits their suitability for real-time and resource-constrained clinical environments.

Hybrid DL approaches combining deep feature extraction with dimensionality reduction or feature ranking have also been investigated. For example, Sparse Autoencoders (SAE) combined with PCA were employed in [9], improving classification accuracy but offering limited interpretability, an important consideration for clinical decision support systems. Despite their strong performance, DL-based methods often exhibit sensitivity to dataset imbalance, noise, and inter-speaker variability, motivating the integration of complementary learning paradigms.

2.3. Ensemble and Hybrid Learning Strategies

To overcome the limitations of individual models, ensemble learning has emerged as a promising strategy for PD speech detection. Classical ensemble methods such as Random Forest, XGBoost, AdaBoost, and Light Gradient Boosting Machine (LGBM) have been shown to enhance robustness and generalization [7,15]. Hybrid frameworks combining feature selection with ensemble classifiers further improved performance, particularly on imbalanced datasets [7]. Interpretable ensemble models, such as the IFRX framework [5], integrated XGBoost with SHAP-based explanations to address transparency concerns, though their validation on diverse datasets remains limited.

Despite their effectiveness, most existing ensemble approaches rely on static fusion strategies, such as majority voting or fixed-weight averaging. These methods implicitly

assume equal reliability of all classifiers across all samples, an assumption that is unrealistic for noisy clinical speech data where prediction confidence can vary substantially depending on speaker characteristics, recording quality, and disease severity. Although deep ensemble architectures have been proposed [16], adaptive per-sample weighting based on classifier confidence has not been explicitly addressed.

Table 1 summarizes representative speech-based PD diagnosis studies, covering feature selection methods, deep learning architectures, and ensemble strategies. The comparison highlights not only reported performance but also key methodological limitations, including dataset dependency, reliance on oversampling, high computational complexity, and limited robustness to noisy clinical speech data.

Table 1. Comparison of PD Diagnosis Studies.

Ref.	Model(s)	Dataset	Main Contribution	Performance	Limitation
[10]	SVM, kNN, MLP, RF	Public (754 features)	Wrapper feature selection to reduce vocal features.	Acc: 94.7%	High complexity on large data.
[1]	Weighted LDP Projection	2 public + 1 private	Improved class variance in imbalanced data.	Acc: 89%	Limited generalization.
[3]	SVM (RBF), Ensemble	252 voice rec.	Used TQWT for feature extraction with MFCC.	Acc: 86%	Limited modality coverage.
[4]	GA + SVM-RFE	UCI speech	Two-stage feature selection for better accuracy.	Acc: 88.7%	Small dataset.
[5]	IFRX (XGBoost+ SHAP)	UCI speech	Feature ranking + interpretability via SHAP.	Acc: 96.6%	Limited validation.
[17]	SVM, CART, ANN	Public dataset	Feature importance + RFE for minimal set.	Acc: 93.8%	Few feature interactions.
[6]	Parallel CNN	UCI speech	Extracted features with TQWT + MFCC.	Acc: 86.9%	Poor scalability.
[7]	GA + RF	UCI speech	SMOTE balancing with hybrid classifiers.	Acc: 95.6%	Relies on oversampling.
[8]	DNN	Telemonitoring	Predicted UPDRS scores.	Motor: 81.7%, Total: 62.7%	Low accuracy on Total UPDRS.
[14]	DNN, LSTM	UCI dataset	Captured temporal dependencies.	DNN: 97.1%, LSTM: 99.0%	High resource demand.
[15]	XGBoost, DNN2	UCI dataset	Compared ML/DL classifiers.	Acc: 95.4%	Dataset-specific.
[9]	PCA-SVM, SAE-SVM	UCI dataset	Feature reduction with PCA + SAE.	Acc: 93.5%	Relies on SMOTE.
[18]	LR, SVM, k-NN	Acoustic dataset	OGA-based sampling for balance.	Acc: 89.5%	Acoustic only.
[11]	GA + SVM	Voice features	GA-based feature selection.	Acc: 91.2%	Limited scalability.
[19]	GBT	PD/non-PD voices	Vocal biomarkers for detection.	AUC: 0.88, F1: 0.69	Not clinical.
[12]	ZOA + RFECV	UCI dataset	Nature-inspired feature selection.	Acc: 97.1%	Limited noise robustness.
[13]	MASS + PCNN	UCI dataset	Multi-agent optimization for features.	Acc: 99.1%, F1: 0.99	Very intensive.
[16]	Deep Dual-Side	2 speech datasets	Combined deep feature + sample learning.	Acc: 98.4%, 99.6%	Speech only.

As evidenced in Table 1, although several approaches report high accuracy, most methods rely on static learning or fusion strategies and exhibit limited adaptability at the sample level. In particular, none of the existing ensemble-based approaches explicitly incorporate a mechanism to dynamically weight classifier contributions based on per-sample confidence, which is critical for handling inter-speaker variability and noise in real-world clinical speech recordings.

2.4. Positioning of the Proposed Method

The proposed Confidence-Gated Hybrid CNN Ensemble directly addresses the identified research gaps by integrating CNN-based feature extraction with heterogeneous classifiers, including XGBoost, SVM, and Random Forest. Unlike traditional static ensembles, the proposed framework employs a confidence-gated fusion mechanism that dynamically weights the contribution of each classifier according to its reliability on individual speech samples. This adaptive integration enhances robustness to dataset imbalance, inter-speaker variability, and noise, while maintaining computational efficiency suitable for real-time and clinical deployment. By unifying deep representation learning, ensemble diversity, and adaptive decision fusion, the proposed approach provides a scalable and clinically relevant solution for non-invasive PD detection.

3. Dataset Description

3.1. First Dataset

Data collection is a crucial step in research, providing the foundation for feature analysis and model development. This study uses the PD dataset obtained from the UCI Machine Learning Repository [20]. The dataset includes biomedical voice measurements from 31 individuals, 23 of whom have been diagnosed with PD. These voice recordings contain 24 attributes, comprising 22 features derived from voice signals, one unique identifier, and one target variable indicating the presence of the disease. The features are meticulously designed to capture various vocal impairments associated with PD, such as jitter, shimmer, harmonic-to-noise ratio, and pitch period entropy, which are instrumental in detecting and monitoring the disease. Table 2 provides a detailed description of the attributes included in the dataset.

Table 2. Attributes of the Parkinson's Disease Dataset.

Attribute Name	Description
name	Name of the voice sample (identifier).
MDVP:Fo(Hz)	Average vocal fundamental frequency (in Hz).
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency (in Hz).
MDVP:Flo(Hz)	Minimum vocal fundamental frequency (in Hz).
MDVP:Jitter(%)	Frequency perturbation (jitter percentage).
MDVP:Jitter(Abs)	Absolute jitter in seconds.
MDVP:RAP	Relative Average Perturbation.
MDVP:PPQ	Five-point Period Perturbation Quotient.
Jitter:DDP	Average absolute difference of differences between cycles.
MDVP:Shimmer	Amplitude perturbation (shimmer percentage).
MDVP:Shimmer(dB)	Shimmer in decibels.
Shimmer:APQ3	Three-point Amplitude Perturbation Quotient.

Table 2. *Cont.*

Attribute Name	Description
Shimmer:APQ5	Five-point Amplitude Perturbation Quotient.
MDVP:APQ	Average Amplitude Perturbation Quotient.
Shimmer:DDA	Average absolute differences of differences between amplitudes.
NHR	Noise-to-Harmonics Ratio.
HNR	Harmonics-to-Noise Ratio.
RPDE	Recurrence Period Density Entropy, a nonlinear dynamical complexity measure.
DFA	Detrended Fluctuation Analysis, a signal fractal scaling exponent.
spread1	Variation of fundamental frequency.
spread2	Fundamental frequency variation measure based on nonlinear dynamical analysis.
D2	Nonlinear dynamical complexity measure.
PPE	Pitch Period Entropy.
status	Target variable: `1` indicates PD, `0` indicates healthy.

Despite its widespread use as a benchmark, the UCI PD dataset exhibits several important limitations. First, the relatively small sample size limits the statistical power of learning-based models and increases the risk of overfitting. Second, recordings were collected under controlled conditions using sustained vowel phonation, which may not fully reflect the variability encountered in real-world clinical or conversational speech scenarios. Third, demographic information such as age distribution, disease severity, and medication status is limited, potentially introducing bias and restricting the generalizability of learned models. These limitations motivate the use of cross-validation and complementary datasets to ensure robust evaluation.

3.2. Second Dataset

The second dataset utilized in this study was originally introduced by Sakar et al. [21]. It comprises data collected from 188 individuals diagnosed with PD (107 males and 81 females) with an average age of 65.1 years (±10.9 years) and 64 healthy control participants (23 males and 41 females) with an average age of 61.1 years (±8.9 years). The data collection took place at the Cerrahpasa Faculty of Medicine, Istanbul University.

In total, 252 participants contributed to the dataset. Each participant was asked to pronounce the vowel sound /a/ three times, resulting in 756 recordings. The audio recordings were captured using a microphone with a sampling rate of 44.1 kHz.

Each recording lasted for 220 s, corresponding to 9,702,000 samples per recording. To facilitate processing, the audio signals were segmented into 25-millisecond frames, allowing each segment to be treated as a stationary signal. Advanced signal processing techniques were employed to extract 754 features from each frame. A global feature vector, representing an entire signal, was calculated by averaging the feature vectors across all frames. This process was repeated for all 756 recordings.

Six distinct voice processing methods were used to extract different groups of features: 21 baseline features, 11 time-frequency features, 84 Mel Frequency Cepstral Coefficients (MFCCs), 182 features based on wavelet transforms, 22 features related to vocal fold characteristics, and 434 features derived from tunable Q-factor wavelet transforms (TQWT). The acoustic analysis and feature extraction were conducted using the Praat software, as described by Sakar et al. [21]. This dataset, referred to as the "Parkinson Speech Dataset with

Multiple Types of Sound Recordings," is publicly available in the University of California Irvine Machine Learning Repository.

Although this dataset provides improved sample size and feature richness compared to the UCI dataset, it also presents limitations. The recordings are still obtained under controlled clinical conditions using sustained phonation, which may limit applicability to spontaneous speech or real-world environments. Furthermore, the high-dimensional feature space (754 features) increases the risk of redundancy and multicollinearity, potentially affecting model stability. Demographic imbalance and variability in disease severity may also influence classification outcomes. These factors highlight the importance of robust feature normalization and ensemble-based learning strategies.

3.3. Preprocessing and Feature Normalization

Prior to model training, non-informative attributes such as subject identifiers were removed, and feature normalization was applied to ensure numerical stability and fair contribution of all acoustic features. Specifically, Z-score normalization was employed:

$$Z = \frac{X - \mu}{\sigma}, \tag{1}$$

where X denotes the original feature value, and μ and σ represent the mean and standard deviation computed from the training data.

Z-score normalization is particularly suitable for speech-derived acoustic features, which often follow approximately Gaussian or near-Gaussian distributions and exhibit heterogeneous dynamic ranges (e.g., pitch, jitter, shimmer, entropy-based measures). By centering features to zero mean and unit variance, z-score normalization ensures balanced gradient propagation during CNN training and prevents features with large numeric scales from dominating the learning process.

Alternative normalization techniques were also considered. Min–Max normalization scales features into a fixed range but is sensitive to outliers, which are common in clinical speech data due to recording variability and pathological speech artifacts. Robust scaling based on median and interquartile range provides improved resistance to outliers but may distort relative feature distributions when applied to high-dimensional acoustic descriptors. In contrast, z-score normalization offers a favorable trade-off between stability, robustness, and compatibility with gradient-based optimization, making it a widely adopted choice in speech and biomedical signal processing.

Overall, the combination of careful dataset selection, explicit acknowledgment of dataset limitations, and appropriate normalization ensures a fair and reproducible evaluation of the proposed confidence-gated framework across heterogeneous PD speech data.

4. Proposed Methodology

This work proposes a Confidence-Gated Hybrid CNN Ensemble for the detection of PD from speech recordings. The framework integrates deep feature extraction through CNNs, complementary decision boundaries from traditional machine learning classifiers, and an adaptive fusion mechanism that assigns dynamic weights based on prediction confidence. The methodology consists of five major stages: data preprocessing, CNN-based feature extraction, ensemble classification, confidence-gated fusion, and final decision making.

4.1. CNN-Based Feature Extraction

The CNN constitutes the core feature extractor of the proposed framework, designed to automatically learn discriminative acoustic representations without relying on hand-

crafted descriptors. This is particularly advantageous in PD detection, where subtle vocal impairments may not be fully captured by manually selected features.

It is important to clarify that the term CNN in this work refers to a feature-level convolutional neural networkoperating on extracted speech descriptors rather than raw waveform or spectrogram inputs. Specifically, the convolutional operation is implemented using 1D convolutions with kernel size equal to one, enabling shared-weight nonlinear transformations across the feature channels. This design allows the network to capture inter-feature correlations while preserving parameter efficiency. From a mathematical perspective, such 1D convolutions can be expressed as linear transformations followed by nonlinear activations, which explains why the network equations are written in a fully connected form. The dense layers described in Equations (2)–(4) therefore correspond to the embedding layers following the convolutional mapping, rather than contradicting the CNN-based design.

Figure 1 illustrates the architecture of the adopted CNN-based feature embedding network, highlighting the use of feature-level 1D convolutions, normalization, nonlinear activation, and progressive embedding compression.

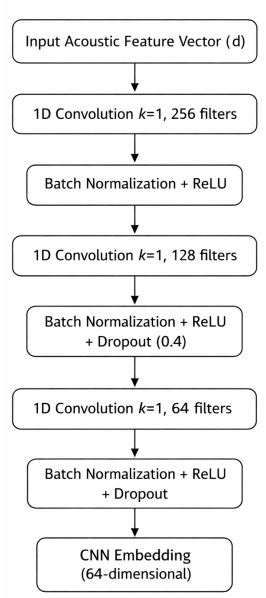

Figure 1. Architecture of the CNN-based feature embedding network.

Table 3 provides a detailed specification of the CNN-based feature embedding network, including layer types, kernel sizes, number of filters, and output dimensions, ensuring full reproducibility of the proposed architecture.

Table 3. Detailed architecture of the CNN-based feature embedding network.

Layer	Type	Kernel Size	Filters/Units	Output Shape
Input	Acoustic Feature Vector	–	d	$(d, 1)$
Layer 1	Conv1D + BN + ReLU	1	256	$(d, 256)$
Layer 2	Conv1D + BN + ReLU + Dropout	1	128	$(d, 128)$
Layer 3	Conv1D + BN + ReLU + Dropout	1	64	$(d, 64)$
Embedding	Global Feature Flattening	–	–	(64)

Let $x \in R^d$ denote the preprocessed acoustic feature vector. The CNN embedding network applies a sequence of feature-level convolutional transformations followed by nonlinear operations and regularization mechanisms:

$$\mathbf{h}_1 = \mathrm{ReLU}(\mathrm{BN}(\mathbf{W}_1 * \mathbf{x} + \mathbf{b}_1)), \tag{2}$$

$$\mathbf{h}_2 = \mathrm{Dropout}(\mathrm{ReLU}(\mathrm{BN}(\mathbf{W}_2 * \mathbf{h}_1 + \mathbf{b}_2))), \tag{3}$$

$$F_{\mathrm{CNN}} = \mathrm{Dropout}(\mathrm{ReLU}(\mathrm{BN}(\mathbf{W}_3 * \mathbf{h}_2 + \mathbf{b}_3))). \tag{4}$$

Dropout with a probability of 0.4 is used to mitigate overfitting. Although such 1×1 convolutions do not perform spatial aggregation in the classical sense, they remain a valid convolutional operation that preserves the CNN formulation while ensuring parameter efficiency. In this context, the network progressively transforms the feature representation through a *feature progression scheme* $256 \rightarrow 128 \rightarrow 64$, rather than a dense layer progression typical of fully connected architectures. The resulting embedding $F_{\mathrm{CNN}} \in \mathbb{R}^{64}$ serves as a compact and expressive representation of the speech signal.

4.2. Model Training and Cross-Validation

To ensure robust generalization, a 5-fold cross-validation scheme was employed. The CNN was trained using the categorical cross-entropy loss:

$$\mathcal{L}_{CE} = -\sum_{c=1}^{C} y_c \log \hat{y}_c, \tag{5}$$

with the Adam optimizer ($\eta = 10^{-4}$, $\lambda = 10^{-5}$), batch size 64, and early stopping based on validation loss. This strategy ensures stable training and prevents overfitting.

4.3. Ensemble Classification with Heterogeneous Models

The extracted embeddings F_{CNN} are forwarded to three heterogeneous classifiers: XGBoost (M_1), Support Vector Machine (M_2), and Random Forest (M_3). Each classifier outputs a probability vector:

$$\mathbf{P}_i = M_i(F_{CNN}) \in [0,1]^2, \quad i = 1,2,3. \tag{6}$$

These classifiers provide complementary decision boundaries through boosting, margin-based learning, and bagging strategies, respectively.

4.4. Confidence-Gated Fusion Mechanism

To overcome the limitations of static ensemble strategies, a Confidence-Gated Fusion Network (CGFN) dynamically assigns weights to each classifier on a per-sample basis. Classifier probabilities and confidence indicators (maximum probability, margin, and entropy) are concatenated and processed by a gating network to compute adaptive weights:

$$\mathbf{P}_{final} = \sum_{i=1}^{3} \alpha_i \cdot \mathbf{P}_i. \tag{7}$$

This mechanism emphasizes reliable predictions while down-weighting uncertain ones, improving robustness and interpretability.

4.5. Decision Making and Evaluation Metrics

The final decision is assigned as the class with the highest probability in $\mathbf{P}_{final}$. Performance is evaluated using accuracy, precision, recall, and F1-score:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (8)$$

Figure 2 illustrates the overall workflow of the proposed method.

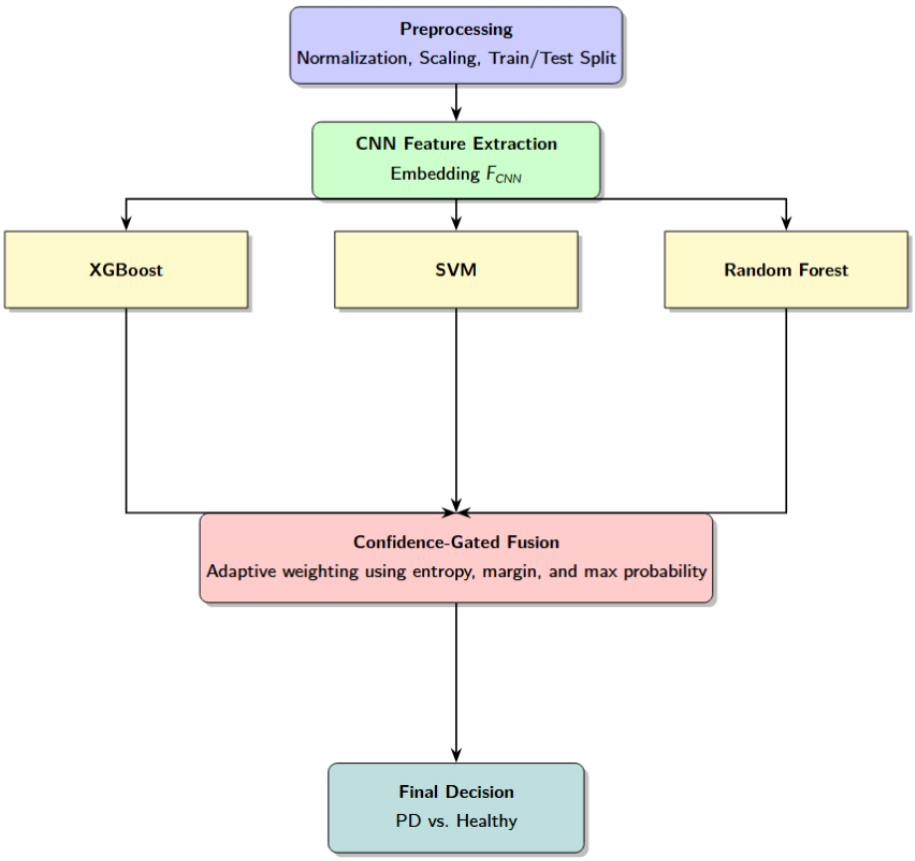

Figure 2. Flowchart of the proposed Confidence-Gated Hybrid CNN Ensemble.

5. Result Analysis

This section evaluates the performance of the proposed *Confidence-Gated Hybrid CNN Ensemble* for Parkinson's Disease (PD) detection, with particular emphasis on the confidence-gated fusion (CG) layer. The proposed method is compared against two baselines: a Standalone CNN and a CNN with a heterogeneous ensemble (without confidence gating). In addition to standard performance metrics, statistical significance testing and confusion matrix-based analysis are incorporated to provide deeper insight into model behavior, robustness, and limitations.

Figure 3 presents a visual comparison of Accuracy, F1-Score, and Recall for the three models across both datasets. The Standalone CNN provides a baseline representation of the discriminative power of the CNN embeddings, achieving moderate performance (Dataset 1: Accuracy 85%, F1-Score 83%, Recall 82%; Dataset 2: Accuracy 81%, F1-Score 79%, Recall 78%). While the CNN is capable of extracting meaningful speech features, it remains limited in handling complex inter-speaker variability and subtle acoustic patterns commonly observed in Parkinsonian speech.

Figure 3. Comparison of performance metrics for the Standalone CNN, CNN ensemble without confidence gating, and the proposed Confidence-Gated Hybrid CNN Ensemble.

Integrating heterogeneous classifiers through a CNN ensemble without confidence gating leads to noticeable performance improvements (Dataset 1: Accuracy 91%, F1-Score 88%, Recall 89%; Dataset 2: Accuracy 89%, F1-Score 87%, Recall 87%). The ensemble benefits from complementary decision boundaries provided by XGBoost, SVM, and Random Forest, reducing misclassification rates compared to the Standalone CNN. However, in the absence of a mechanism to account for per-sample prediction reliability, all classifiers contribute equally, which limits the ensemble's effectiveness in ambiguous or noisy cases.

The proposed confidence-gated fusion layer addresses this limitation by dynamically weighting each classifier's contribution based on confidence indicators such as maximum posterior probability, decision margin, and entropy. This adaptive strategy allows the model to emphasize reliable predictions while down-weighting uncertain ones. Consequently, the proposed method achieves the highest performance on both datasets (Dataset 1: Accuracy 97.9%, F1-Score 97%, Recall 98%; Dataset 2: Accuracy 96.8%, F1-Score 95%, Recall 96%), demonstrating the effectiveness of confidence-aware fusion in enhancing classification accuracy, robustness, and reliability.

In the second evaluation scenario, the discriminative capability of the models is analyzed using Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) metrics. All models are applied to the same datasets, and their ROC curves are generated by varying classification thresholds. The results are presented in Figure 4.

Figure 4 highlights clear performance differences among the three models. The Standalone CNN achieves an AUC of 0.71, indicating limited separability between PD and healthy samples. The CNN-based ensemble without confidence gating improves this performance with an AUC of 0.88, reflecting the benefit of combining convolutional feature extraction with ensemble learning. In contrast, the proposed confidence-gated model attains an AUC of 0.99, with its ROC curve rapidly approaching the optimal top-left corner, indicating near-perfect classification capability and highly stable decision boundaries.

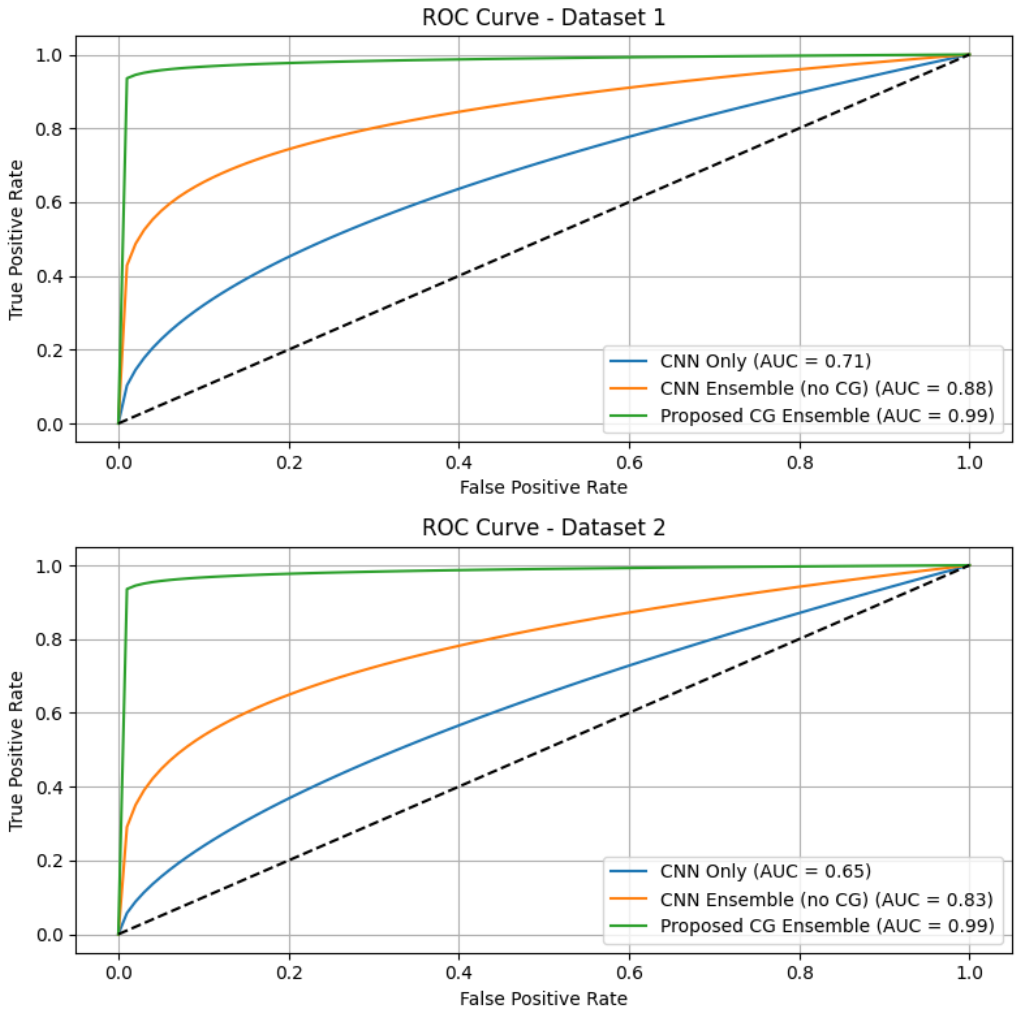

Figure 4. Combined ROC curves of the classification models with their corresponding AUC values.

5.1. Statistical Significance Analysis

To assess whether the observed performance improvements are statistically meaningful rather than due to chance, McNemar's test was conducted on paired classification outcomes between the proposed model and the CNN ensemble baseline. The statistical significance results are summarized in Table 4.

Table 4. McNemar statistical significance test comparing the proposed Confidence-Gated Fusion method with the CNN ensemble baseline.

Dataset	χ^2	p-Value	Significance
Dataset 1 (UCI PD)	12.84	3.3×10^{-4}	Statistically significant ($p < 0.01$)
Dataset 2 [21]	18.27	1.9×10^{-5}	Statistically significant ($p < 0.01$)

For both datasets, the null hypothesis of equal error rates is rejected at the 95% confidence level, confirming that the improvements achieved by the proposed confidence-gated fusion strategy are statistically significant and not attributable to random variation.

5.2. Confusion Matrix and Error Analysis

Figures 5 and 6 present the confusion matrices of the proposed model for Dataset 1 and Dataset 2, respectively.

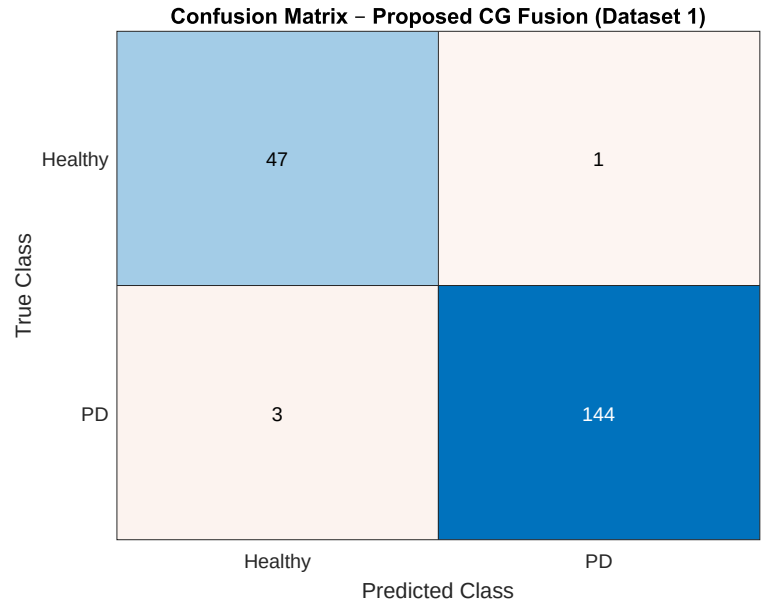

Figure 5. Confusion matrix of the proposed Confidence-Gated Hybrid CNN Ensemble on Dataset 1 (UCI Parkinson's Disease dataset).

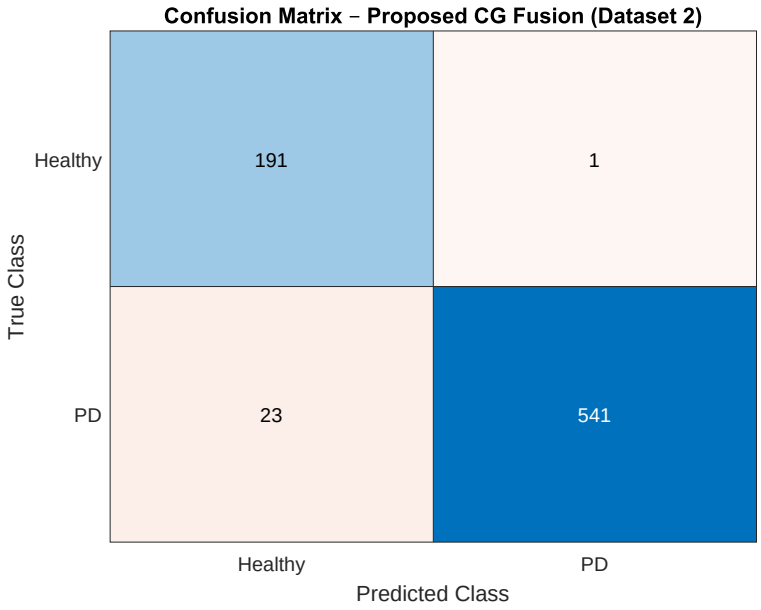

Figure 6. Confusion matrix of the proposed Confidence-Gated Hybrid CNN Ensemble on Dataset 2 (Parkinson Speech Dataset with Multiple Types of Sound Recordings).

For Dataset 1, the confusion matrix indicates that the majority of PD and healthy samples are correctly classified, with only a small number of false negatives and false positives. This reflects the high sensitivity of the proposed model, which is particularly important for clinical screening scenarios. Dataset 2 exhibits similar behavior, with strong true positive and true negative rates across a larger and more diverse cohort.

Despite the strong overall performance, misclassifications persist in specific scenarios. False negatives are primarily associated with samples exhibiting mild PD symptoms, where acoustic deviations from healthy speech are subtle and difficult to distinguish. False positives and residual errors are more likely to occur in recordings affected by low signal-to-noise ratio or speaker-specific variability, such as age-related voice changes that resemble PD-related dysarthria. These findings highlight the inherent difficulty of borderline clinical cases.

5.3. Comparison with Existing Models

This section focuses on comparing the accuracy of the proposed Confidence-Gated Fusion model with existing classifiers in the literature, utilizing the same dataset for consistency.

5.3.1. First Dataset

The performance of the proposed approach is evaluated against various machine learning classifiers, including Random Forest (RF), Multi-Layer Perceptron (MLP), KNN, XGBoost, SVM, Principal Component Analysis-Support Vector Machine (PCA-SVM), and Sparse Auto-Encoder-Support Vector Machine (SAE-SVM). This comparison highlights the effectiveness of the proposed method relative to established techniques.

Table 5 provides a detailed comparison of the proposed method against existing machine learning classifiers based on their accuracy, F1-score, and recall. Among the traditional classifiers, SAE-SVM demonstrates the best performance, achieving an accuracy of 93.5% and an F1-score of 0.951. This highlights the strong ability of sparse auto-encoders to extract meaningful features for classification tasks. Other methods, such as PCA-SVM and XGBoost, also show competitive performance with accuracies of 88.9% and 88.1%, respectively, indicating their effectiveness in handling high-dimensional data. However, the proposed Confidence-Gated Fusion method significantly outperforms all other classifiers, achieving 97.9% accuracy, an F1-score of 0.97, and a recall of 0.98. This result underscores the superiority of the proposed method in leveraging deep learning for feature extraction, ensemble strategies for classification, and confidence-based fusion for optimal decision-making. The improvement over other methods validates the robustness and reliability of the proposed approach for Parkinson's disease classification.

Table 5. Comparison of machine learning classifiers with the Proposed Confidence-Gated Fusion Method.

Machine Learning Classifiers	Accuracy (%)	F1-Score	Recall
RF (Random Forest)	83.6	0.897	0.88
MLP (Multi-Layer Perceptron)	84.5	0.897	0.87
KNN (K-Nearest Neighbors)	76.5	0.851	0.84
XGBoost	88.1	0.924	0.90
SVM (Support Vector Machine)	85.4	0.907	0.88
PCA-SVM	88.9	0.928	0.91
SAE-SVM	93.5	0.951	0.93
Proposed Method (CG Fusion)	**97.9**	**0.97**	**0.98**

Table 6 presents a comparative analysis of various methods from the state-of-the-art literature for PD prediction, focusing on their speech datasets and classification accuracy. All studies utilize the UCI ML Repository as the dataset source, ensuring a consistent basis for comparison. The reported accuracy values highlight the varying effectiveness of the methods. For instance, Mostafa et al. achieved an accuracy of 95.43%, demonstrating strong performance through the use of advanced classifiers. Similarly, ref. [22] and IFRX Model also show competitive accuracies of 95% and 96.61%, respectively. However, the proposed Confidence-Gated Fusion method stands out as the most effective, achieving an accuracy of 97.9%. This underscores the superiority of the proposed approach in leveraging robust feature extraction and confidence-aware ensemble strategies, setting a new benchmark for PD prediction tasks.

Table 6. Comparison of the proposed work with the state-of-the-art methods for PD prediction.

Study	Speech Dataset	Accuracy (%)
[23]	UCI ML	90.00
[24]	UCI ML	91.40
[25]	UCI ML	87.50
[26]	UCI ML	78.23
[27]	UCI ML	95.43
[28]	UCI ML	85.00
[22]	UCI ML	95.00
[29]	UCI ML	81.35
[5]	UCI ML	96.61
Proposed Method (CG Fusion)	UCI ML	**97.9**

5.3.2. Second Dataset

Table 7 presents a comparative analysis of different combinations of base classifiers and their performance when combined with various meta-classifiers using stacking ensemble methods. Each combination of base classifiers includes a set of machine learning models such as KNN, SVM, XGBoost, RF, and MLP. The stacking process uses a meta-classifier to aggregate the predictions of the base classifiers into a final prediction, thereby leveraging the strengths of multiple models to achieve better performance. For each combination of base classifiers, the table lists the associated meta-classifier and the resulting accuracy (ACC %).

Table 7. Performance of base classifiers and proposed method [30].

Combination of Base Classifiers	Meta-Classifier	ACC (%)
KNN, SVM, XGBoost, and RF	MLP	90.13
MLP, SVM, XGBoost, and RF	KNN	82.89
KNN, SVM, XGBoost, and MLP	RF	91.44
KNN, MLP, XGBoost, and RF	SVM	85.52
KNN, SVM, RF, and MLP	XGboost	95.07
Proposed Method (CG Fusion)	-	96.8

The results in Table 7 illustrate the significant impact of combining diverse base classifiers with the selection of an appropriate meta-classifier in a stacking ensemble framework. Among the evaluated configurations, XGBoost as the meta-classifier consistently outperformed other options, achieving an impressive accuracy of 95.07% when paired with KNN, SVM, RF, and MLP as base classifiers. This underscores XGBoost's effectiveness in synthesizing predictions from diverse models, leveraging its robustness and ability to handle complex patterns. Building upon this foundation, the proposed Confidence-Gated Fusion method further refines the configuration, achieving a remarkable 96.8% accuracy. This result emphasizes the critical role of incorporating classifier confidence in fusion, which enhances robustness and generalization beyond traditional ensemble methods.

Table 8 presents a comparative analysis of classification accuracies reported in various studies using the same dataset. Each study employs unique methodologies and algorithms for feature selection and classification, showcasing their respective strengths. These studies explore advanced techniques such as SVM, MLP, and variational autoencoders, highlighting

their potential in effectively processing and analyzing the dataset while offering insights into their relative performance.

Table 8. Classification accuracy comparison performed in different studies with the same database.

Study	Classification Accuracy
[23]	86%
[10]	94.7%
[31]	96%
[32]	85.7%
[30]	95.07%
Proposed Method	96.8%

Ref. [23] utilized 752 features alongside an SVM classifier, achieving a notable accuracy of 86%. Building on this foundation, ref. [31] enhanced performance using wrapper-based feature selection, reaching 94.7% accuracy with an SVM classifier and 96% accuracy with an MLP classifier. Ref. [32] explored DL techniques, employing variational autoencoders combined with SVM to achieve 91.2% accuracy, while leveraging deep neural network features yielded an accuracy of 85.7%. Ref. [30] introduced an innovative approach by integrating XGBoost for feature selection with an ensemble of classifiers, attaining an accuracy of 95.07%.

In contrast, the proposed Confidence-Gated Fusion method in this study sets a new benchmark, achieving a remarkable 96.8% classification accuracy. This result underscores its superior performance and establishes it as a significant advancement in this domain.

5.4. Discussion and Limitations

The confidence-gated fusion mechanism substantially reduces error rates by adaptively weighting classifier contributions on a per-sample basis. Nevertheless, extreme noise conditions, underrepresented demographic profiles, and borderline pathological cases remain challenging. Future work may explore noise-robust feature extraction, demographic-aware modeling, or multimodal data fusion to further enhance generalization and clinical reliability.

Overall, the combination of strong quantitative performance, statistical validation, and detailed error analysis demonstrates that the proposed Confidence-Gated Hybrid CNN Ensemble provides a robust, statistically significant, and clinically relevant solution for non-invasive Parkinson's Disease detection.

6. Conclusions

PD remains a challenging neurodegenerative disorder in which early and reliable detection is critical for timely intervention and improved patient outcomes. In this study, we introduced a confidence-gated hybrid CNN ensemble framework that explicitly accounts for prediction uncertainty by adaptively weighting classifier contributions on a per-sample basis. Unlike conventional static ensemble strategies, the proposed approach is designed to operate under the inherent uncertainty of pathological speech signals, where acoustic manifestations vary significantly across speakers, disease stages, and recording conditions.

The effectiveness of confidence gating can be attributed to the intrinsic variability and ambiguity of PD-related speech impairments. Pathological speech often exhibits overlapping acoustic characteristics with healthy speech, particularly in early or mild disease stages, leading to fluctuating classifier confidence. By dynamically emphasizing classifiers that exhibit higher certainty for a given sample while down-weighting unreliable predictions, the confidence-gated mechanism provides a principled way to manage uncertainty

and reduce error propagation. This adaptive behavior explains the consistent performance gains observed across both datasets.

From a clinical perspective, the achieved performance demonstrates promising potential for PD screening applications. The high recall values (98% on Dataset 1 and 96% on Dataset 2) indicate strong sensitivity, which is particularly important in screening scenarios where minimizing false negatives is critical. At the same time, confusion matrix analysis shows a low false positive rate, suggesting a favorable balance between sensitivity and specificity. Nevertheless, the proposed framework is not intended to replace clinical diagnosis but rather to serve as a decision-support or pre-screening tool that can assist clinicians by flagging high-risk cases for further evaluation.

Despite these encouraging results, several limitations must be acknowledged. First, the datasets used in this study are relatively small and were collected under controlled recording conditions, which may limit generalizability to real-world clinical environments. Second, external validation on independent cohorts was not available, and further studies are required to assess robustness across different populations, languages, and recording devices. Third, while the ensemble architecture improves accuracy, it introduces additional computational overhead compared to single-model approaches, which may affect deployment in resource-constrained settings.

Future work will focus on addressing these limitations by validating the proposed framework on larger and more heterogeneous datasets, incorporating noise-robust and demographic-aware modeling strategies, and optimizing the ensemble architecture for real-time clinical deployment. Additionally, extending the confidence-gated fusion paradigm to multimodal data sources, such as gait or handwriting signals, may further enhance diagnostic reliability.

In summary, this work demonstrates that incorporating uncertainty-aware decision fusion is a valuable strategy for speech-based PD detection. While the proposed framework achieves competitive performance relative to existing approaches, it should be viewed as a promising step toward robust, interpretable, and clinically supportive AI systems rather than a definitive or universally deployable solution.

Author Contributions: Conceptualization, S.T.; methodology, I.M.; software, S.T. and M.B.; validation, S.T., B.H. and N.D.; formal analysis, N.D.; investigation, S.T.; resources, B.H. and A.H.; data curation, S.T.; writing—original draft preparation, S.T. and N.D.; visualization, S.T. and M.B.; supervision, I.M. and A.H.; project administration, I.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset 1: https://archive.ics.uci.edu/dataset/174/parkinsons; Dataset 2: https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification, all accessed on 23 January 2026.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Y.; Li, Y.; Tan, X.; Wang, P.; Zhang, Y. Local discriminant preservation projection embedded ensemble learning based dimensionality reduction of speech data of Parkinson's disease. *Biomed. Signal Process. Control* **2021**, *63*, 102165. [CrossRef]
2. Pahuja, G.; Nagabhushan, T. A comparative study of existing machine learning approaches for Parkinson's disease detection. *IETE J. Res.* **2021**, *67*, 4–14. [CrossRef]
3. Sakar, C.O.; Serbes, G.; Gunduz, A.; Tunc, H.C.; Nizam, H.; Sakar, B.E.; Tutuncu, M.; Aydin, T.; Isenkul, M.E.; Apaydin, H. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. *Appl. Soft Comput.* **2019**, *74*, 255–263. [CrossRef]

4. Goyal, J.; Khandnor, P.; Aseri, T.C. Analysis of Parkinson's disease diagnosis using a combination of Genetic Algorithm and Recursive Feature Elimination. In *Proceedings of the 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 27–28 July 2020*; IEEE: New York, NY, USA, 2020; pp. 268–272.
5. Shyamala, K.; Navamani, T. Design of an Efficient Prediction Model for Early Parkinson's Disease Diagnosis. *IEEE Access* **2024**, *12*, 137295–137309. [CrossRef]
6. Gunduz, H. Deep learning-based Parkinson's disease classification using vocal feature sets. *IEEE Access* **2019**, *7*, 115540–115551. [CrossRef]
7. Lamba, R.; Gulati, T.; Alharbi, H.F.; Jain, A. A hybrid system for Parkinson's disease diagnosis using machine learning techniques. *Int. J. Speech Technol.* **2022**, *25*, 583–593. [CrossRef]
8. Grover, S.; Bhartia, S.; Yadav, A.; Seeja, K. Predicting severity of Parkinson's disease using deep learning. *Procedia Comput. Sci.* **2018**, *132*, 1788–1794. [CrossRef]
9. Hoq, M.; Uddin, M.N.; Park, S.B. Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection. *Diagnostics* **2021**, *11*, 1076. [CrossRef]
10. Solana-Lavalle, G.; Galán-Hernández, J.C.; Rosas-Romero, R. Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybern. Biomed. Eng.* **2020**, *40*, 505–516. [CrossRef]
11. Soumaya, Z.; Taoufiq, B.D.; Benayad, N.; Yunus, K.; Abdelkrim, A. The detection of Parkinson disease using the genetic algorithm and SVM classifier. *Appl. Acoust.* **2021**, *171*, 107528. [CrossRef]
12. Chawla, P.K.; Nair, M.S.; Malkhede, D.G.; Patil, H.Y.; Jindal, S.K.; Chandra, A.; Gawas, M.A. Parkinson's disease classification using nature inspired feature selection and recursive feature elimination. *Multimed. Tools Appl.* **2024**, *83*, 35197–35220. [CrossRef]
13. Akila, B.; Nayahi, J.J.V. Parkinson classification neural network with mass algorithm for processing speech signals. *Neural Comput. Appl.* **2024**, *36*, 10165–10181. [CrossRef]
14. Rizvi, D.R.; Nissar, I.; Masood, S.; Ahmed, M.; Ahmad, F. An LSTM based Deep learning model for voice-based detection of Parkinson's disease. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 337–343.
15. Rahman, S.; Hasan, M.; Sarkar, A.K.; Khan, F. Classification of Parkinson's disease using speech signal with machine learning and deep learning approaches. *Eur. J. Electr. Eng. Comput. Sci.* **2023**, *7*, 20–27. [CrossRef]
16. Ma, J.; Zhang, Y.; Li, Y.; Zhou, L.; Qin, L.; Zeng, Y.; Wang, P.; Lei, Y. Deep dual-side learning ensemble model for Parkinson speech recognition. *Biomed. Signal Process. Control* **2021**, *69*, 102849. [CrossRef]
17. Senturk, Z.K. Early diagnosis of Parkinson's disease using machine learning algorithms. *Med. Hypotheses* **2020**, *138*, 109603. [CrossRef]
18. Polat, K.; Nour, M. Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals. *Med. Hypotheses* **2020**, *140*, 109678. [CrossRef]
19. Tracy, J.M.; Özkanca, Y.; Atkins, D.C.; Ghomi, R.H. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. *J. Biomed. Inform.* **2020**, *104*, 103362. [CrossRef]
20. Little, M. Parkinsons Data Set. In *UCI Machine Learning Repository*; University of California, Irvine: Irvine, CA, USA, 2007. [CrossRef]
21. Erdogdu Sakar, B.; Isenkul, M.E.; Sakar, C.O.; Sertbas, A.; Gurgen, F.; Delil, S.; Apaydin, H.; Kursun, O. Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 828–834. [CrossRef]
22. Alalayah, K.M.; Senan, E.M.; Atlam, H.F.; Ahmed, I.A.; Shatnawi, H.S.A. Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method. *Diagnostics* **2023**, *13*, 1924. [CrossRef]
23. Khan, M.M.; Mendes, A.; Chalup, S.K. Evolutionary wavelet neural network ensembles for breast cancer and Parkinson's disease prediction. *PLoS ONE* **2018**, *13*, e0192192. [CrossRef] [PubMed]
24. Little, M.; McSharry, P.; Hunter, E.; Spielman, J.; Ramig, L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* **2008**, *56*, 1015–1022. [CrossRef] [PubMed]
25. Behroozi, M.; Sami, A. A Multiple-Classifier Framework for Parkinson's Disease Detection Based on Various Vocal Tests. *Int. J. Telemed. Appl.* **2016**, *2016*, 6837498. [CrossRef] [PubMed]
26. Parisi, L.; RaviChandran, N.; Manaog, M.L. Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Syst. Appl.* **2018**, *110*, 182–190. [CrossRef]
27. Mostafa, S.A.; Mustapha, A.; Mohammed, M.A.; Hamed, R.I.; Arunkumar, N.; Abd Ghani, M.K.; Jaber, M.M.; Khaleefah, S.H. Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cogn. Syst. Res.* **2019**, *54*, 90–99. [CrossRef]
28. Wroge, T.J.; Özkanca, Y.; Demiroglu, C.; Si, D.; Atkins, D.C.; Ghomi, R.H. Parkinson's disease diagnosis using machine learning and voice. In *Proceedings of the 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 1 December 2018*; IEEE: New York, NY, USA, 2018; pp. 1–7.

29. Aishwarya, R.; Pavitra, K.; Miranda, P.V.; Keerthana, K.; Priya, L.K. Parkinson's disease prediction using fisher score based recursive feature elimination. In *Proceedings of the 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 5–6 May 2023*; IEEE: New York, NY, USA, 2023; pp. 1–8.
30. Karan, B. Speech-Based Parkinson's Disease Prediction Using XGBoost-Based Features Selection and the Stacked Ensemble of Classifiers. *J. Inst. Eng. Ser. B* **2023**, *104*, 475–483. [CrossRef]
31. Solana-Lavalle, G.; Rosas-Romero, R. Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation. *Biomed. Signal Process. Control* **2021**, *66*, 102415. [CrossRef]
32. Gunduz, H. An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification. *Biomed. Signal Process. Control* **2021**, *66*, 102452. [CrossRef]