International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

A Multimodal Machine Learning Framework For Early Screening Of Parkinson's Disease

V.D. Gupta[1], M.R. Shaikh[2], Hiten Shah[3], Sakshi Shinde[4], Krushnakant Take[5], Nikhil Shinde[6]

[1,2]*Lecturer, Department of Computer Technology, Sanjivani K.B.P Polytechnic, Kopargaon,*
[3,4,5,6]*Research Scholar, Department of Computer Technology, Sanjivani K.B.P Polytechnic, Kopargaon*

***Abstract*— Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects millions worldwide. Early detection is crucial for effective management, yet current diagnostic methods often identify the disease only after significant neuronal damage has occurred. This paper presents NeuroFusion, a novel multimodal machine learning framework that combines voice analysis and spiral drawing assessment for non-invasive early screening of Parkinson's Disease. The system employs Support Vector Machines (SVM) for voice feature extraction and analysis, including acoustic parameters such as jitter, shimmer, and harmonic-to-noise ratio, and utilizes Convolutional Neural Networks (CNN) for analyzing fine motor control patterns in spiral drawings. Through weighted multimodal fusion (0.6 for voice, 0.4 for spiral), NeuroFusion achieves 86.67% accuracy, 78.95% precision, and 100% recall on a test dataset of 30 paired samples. The framework incorporates explainability features and generates comprehensive risk assessment reports. Results demonstrate that multimodal integration significantly enhances screening reliability compared to single-modality approaches, with the 100% recall rate ensuring no Parkinson's cases are missed—a critical requirement for medical screening applications. This research contributes to the development of accessible, cost-effective, and non-invasive screening tools that could enable earlier intervention and improved patient outcomes.**

***Keywords*— Parkinson's Disease, Machine Learning, Multimodal Fusion, Voice Analysis, CNN, SVM, Early Screening, Non-invasive Diagnosis.**

I. INTRODUCTION

Parkinson's Disease (PD) is a progressive neurodegenerative disorder affecting approximately 10 million people worldwide [1]. The disease results from degeneration of dopamine-producing neurons in the substantia nigra, leading to characteristic motor symptoms including tremor, rigidity, bradykinesia, postural instability, speech abnormalities, and loss of fine motor control [2]. These manifestations significantly impact patients' quality of life and independence.

Early detection of Parkinson's Disease remains a critical clinical challenge. Symptoms appear gradually, and by the time conventional diagnosis occurs, approximately 60-80% of dopaminergic neurons have already been irreversibly lost [3]. Traditional diagnostic approaches rely on clinical observation, neurological examination, and expensive imaging techniques such as DaTscan, which are not readily accessible in many healthcare settings. The absence of definitive biomarkers and the subjective nature of clinical assessment can lead to diagnostic uncertainty, especially in early stages.

Recent advances in machine learning have opened new possibilities for early disease detection through analysis of subtle biomarkers. Two particularly promising modalities are voice analysis and drawing assessment. Voice changes in PD patients include reduced loudness, monotone speech, imprecise articulation, and altered voice quality—collectively termed hypokinetic dysarthria [4]. Acoustic features such as jitter, shimmer, and harmonic-to-noise ratio have been identified as quantifiable biomarkers with significant discriminative power [5]. Similarly, fine motor impairment manifests in handwriting and drawing tasks through micrographia, tremor-induced irregularities, and reduced fluency [6, 7]. Spiral drawing tests are particularly effective because they require sustained fine motor control and reveal tremor characteristics amenable to computational analysis.

While single-modality machine learning systems have shown promise, they face inherent limitations. Voice-only systems can be sensitive to recording conditions, background noise, and natural speech variability among individuals. Spiral-only systems may be affected by variations in drawing skill, age-related motor decline unrelated to PD, and image capture quality. These modality-specific vulnerabilities motivate multimodal approaches that leverage complementary information to improve diagnostic reliability [18]. By combining voice and drawing analysis, a multimodal system captures distinct aspects of PD pathophysiology while providing redundancy—if one modality produces uncertain results, the other can compensate.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

This paper presents a multimodal machine learning framework that integrates voice analysis and spiral drawing assessment for early Parkinson's Disease screening. The system combines Support Vector Machines for voice feature classification with Convolutional Neural Networks for spiral image analysis, employing a weighted fusion strategy to generate probabilistic risk scores. Voice recordings are processed using Parselmouth [11] to extract acoustic features, while spiral images are analyzed through deep learning-based pattern recognition. The framework is implemented with a web-based Streamlit interface for potential clinical deployment.

This framework is designed specifically for screening purposes—to identify individuals who may benefit from further clinical evaluation—rather than for definitive diagnosis. Screening tools maximize sensitivity to detect all potential cases, with positive results prompting referral to neurologists for confirmatory testing. This distinction between screening and diagnosis is essential both ethically and clinically.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 details the methodology, Section 4 presents experimental results, Section 5 discusses implications and limitations, Section 6 outlines future research directions, and Section 7 concludes with a summary of contributions.

II. Related Work

A. Voice-Based Parkinson's Detection

Voice analysis has been extensively studied as a non-invasive method for Parkinson's Disease detection. The foundational work by Tsanas et al. [4] demonstrated that acoustic features extracted from sustained phonation could accurately monitor PD progression, achieving strong correlation with clinical assessments. Their research established key vocal biomarkers including fundamental frequency variation, jitter, shimmer, and harmonic-to-noise ratio as reliable indicators of disease presence and severity.

Subsequent studies have employed various machine learning algorithms to classify PD from voice recordings. Sakar et al. [5] conducted a comprehensive comparative analysis of speech signal processing algorithms, evaluating multiple feature extraction methods and classifiers. Their work with the tunable Q-factor wavelet transform achieved classification accuracies above 85% on Turkish PD voice datasets. Aich et al. [10] proposed a nonlinear decision tree-based approach using different feature sets, demonstrating that feature selection significantly impacts classification performance.

Little et al. [20] explored nonlinear recurrence and fractal scaling properties in voice signals, introducing novel features that capture the complex dynamics of dysphonic speech.

Despite these advances, voice-only systems face challenges including sensitivity to recording quality, background noise, microphone characteristics, and natural inter-individual variability in speech patterns. Performance can degrade significantly in real-world settings where controlled recording conditions cannot be guaranteed.

B. Handwriting and Drawing Analysis

Handwriting and drawing analysis has emerged as another promising modality for PD assessment. Impedovo and Pirlo [7] provided a comprehensive review of dynamic handwriting analysis from a pattern recognition perspective, highlighting features such as writing pressure, velocity, and acceleration as informative biomarkers for neurodegenerative diseases. Their review emphasized that PD-related motor dysfunction manifests distinctly in graphomotor tasks.

Computer vision approaches have been particularly successful in analyzing static drawing images. Pereira et al. [6] developed a computer vision-based system using spiral and meander drawings, employing texture analysis and optimum-path forest classifiers to achieve notable accuracy. Gil-Martín et al. [8] applied convolutional neural networks specifically to drawing movements, demonstrating that deep learning can automatically extract relevant spatial features without manual feature engineering. Ramachandran and Polat [16] further advanced this area by using deep learning architectures for automated detection and classification from spiral tests, achieving over 90% accuracy on specialized datasets.

However, drawing-based systems are susceptible to baseline skill variation, age-related motor changes in healthy individuals, and image quality issues. Individual differences in artistic ability and familiarity with drawing tasks introduce variability that complicates classification.

C. Multimodal Approaches

While unimodal systems have demonstrated individual strengths, multimodal fusion approaches offer enhanced robustness and reliability. Prashanth et al. [18] presented a landmark study on multimodal PD detection, combining features from multiple data sources and achieving high accuracy through machine learning integration. Their work demonstrated that fusion of complementary modalities significantly outperforms single-modality approaches by leveraging diverse information sources.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Nilashi et al. [19] developed a hybrid intelligent system for PD progression prediction, integrating multiple machine learning techniques to improve predictive accuracy. Balaji et al. [9] explored gait analysis combined with supervised learning for early detection and stage classification, showing that motor assessment from different body systems provides complementary diagnostic value. Alhussein et al. [15] demonstrated the potential of cognitive IoT-cloud integration for smart healthcare applications, indicating pathways for deploying multimodal AI systems in clinical practice.

Despite these promising developments, research specifically combining voice and handwriting analysis for Parkinson's screening remains limited. Existing multimodal systems often focus on clinical or imaging data rather than easily accessible behavioral modalities. Key challenges include determining optimal fusion strategies, handling missing modalities gracefully, ensuring interpretability of combined predictions, and validating performance across diverse populations.

The framework presented in this paper addresses these gaps by implementing a weighted linear fusion approach that combines voice acoustic analysis and spiral drawing assessment. Unlike previous work that may require specialized equipment or clinical data, this system uses readily obtainable voice recordings and simple drawing tasks, making it suitable for accessible screening in diverse healthcare settings. The weighted fusion strategy accounts for modality-specific reliability while the multi-sample aggregation mechanism enhances robustness against noisy or anomalous inputs.

III. Methodology

A. System Architecture

The proposed multimodal framework consists of five primary components that work in sequence to generate risk assessments. First, voice recordings undergo feature extraction to compute acoustic parameters. Second, spiral drawing images are preprocessed and normalized for neural network input. Third, independent classification models—a Support Vector Machine for voice and a Convolutional Neural Network for spirals—generate modality-specific probability scores. Fourth, a multimodal fusion module combines these predictions using weighted averaging. Fifth, a visualization and reporting interface presents the final risk assessment with explanatory information.

The system accepts multiple voice recordings and spiral images as input, processes each sample independently, aggregates predictions within each modality through arithmetic averaging, and then combines the modality-specific scores to generate a final risk percentage ranging from 0% (low risk) to 100% (high risk).

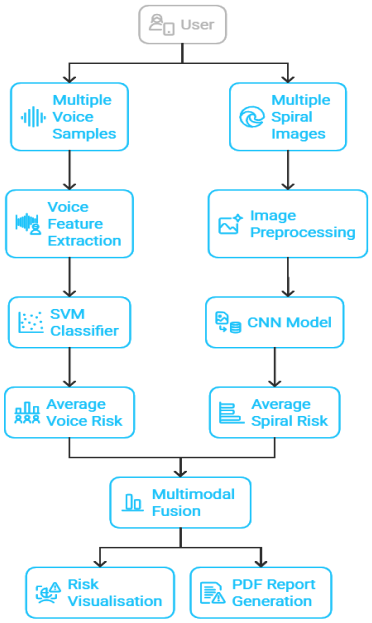

Figure I: System Architecture Diagram

B. Datasets

1) *Voice Dataset:* The voice modality utilizes the Italian Parkinson's Voice Dataset, a publicly available collection containing approximately 831 voice recordings from Italian speakers. The dataset comprises 394 recordings from healthy individuals and 437 recordings from Parkinson's Disease patients. Each recording contains sustained phonation or speech samples captured under controlled conditions. These audio files are stored in WAV format and were collected from participants with confirmed clinical diagnoses, providing reliable ground truth labels for supervised learning.

2) *Spiral Dataset:* The spiral drawing modality uses the Parkinson's Spiral Drawing Dataset, organized into training and testing subsets. The test set, used for final evaluation, contains 30 images evenly distributed across two classes: 15 spiral drawings from healthy individuals and 15 from Parkinson's patients.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Participants were instructed to draw spirals on paper, which were then digitized as image files in PNG or JPG format. The spirals exhibit characteristic differences between groups, with Parkinson's patients typically showing greater irregularity, tremor-induced oscillations, and disrupted continuity.

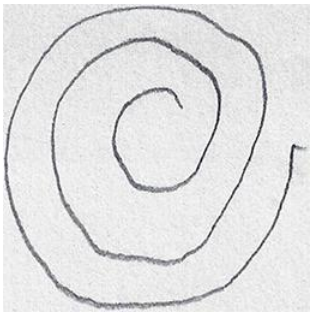

Figure II: A Spiral Drawing of a Healthy Subject

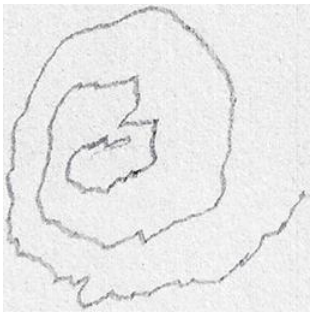

Figure III: A Spiral Drawing of a Parkinson's Patient

3) *Multimodal Test Set Creation:* Since the voice and spiral datasets were collected independently from different participant populations, they do not contain truly paired multimodal samples from the same individuals. To enable multimodal evaluation, a label-consistent pairing strategy was employed. Each spiral test image was systematically paired with a voice recording sharing the same diagnostic label (healthy or Parkinson's), creating 30 multimodal test pairs: 15 healthy pairs and 15 Parkinson's pairs. While this approach does not replicate the ideal scenario of having multiple modalities from identical patients, it enables assessment of the fusion strategy's ability to integrate complementary information. The pairings were documented in a CSV file with columns for voice file path, spiral image path, and binary label (0 for healthy, 1 for Parkinson's).

C. Voice Feature Extraction

Acoustic features were extracted from voice recordings using Parselmouth [11], a Python interface to Praat, the widely-used phonetics analysis software. For each voice recording, the following five features were computed:

1. Mean Pitch (Hz): The average fundamental frequency of the voice signal across the recording duration. This represents the primary vibration rate of the vocal folds and typically decreases or shows reduced variability in PD patients due to reduced vocal fold tension and control.

2. Pitch Standard Deviation: The variability in fundamental frequency over time. Reduced pitch variability reflects the monotone speech characteristic of hypokinetic dysarthria in Parkinson's Disease.

3. Jitter: A measure of cycle-to-cycle variation in fundamental frequency, quantifying the irregularity of vocal fold vibration. Elevated jitter indicates unstable phonation and is significantly higher in Parkinsonian speech due to impaired laryngeal motor control.

4. Shimmer: A measure of cycle-to-cycle variation in amplitude, reflecting loudness instability. Higher shimmer values indicate difficulty maintaining consistent vocal intensity, a common manifestation of reduced respiratory support and vocal fold control in PD.

5. Harmonic-to-Noise Ratio (dB): The ratio of periodic (harmonic) to aperiodic (noise) components in the voice signal. This metric quantifies voice quality and clarity. Lower HNR values indicate breathier, more turbulent voice quality, which is typical in Parkinson's patients due to incomplete vocal fold closure and reduced glottal efficiency.

These five features were selected based on their established clinical relevance and documented correlation with Parkinsonian speech characteristics in the literature [4, 5, 20]. The features form a five-dimensional vector for each voice sample: [mean_pitch, pitch_std, jitter, shimmer, hnr]. Before classification, feature values were standardized using z-score normalization (subtracting the mean and dividing by the standard deviation) to ensure all features contribute equally to the SVM decision boundary, preventing features with larger numerical ranges from dominating the classification.

D. Voice Classification Model

A Support Vector Machine with Radial Basis Function (RBF) kernel was employed for voice-based classification [12].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

SVM was selected for several reasons: it performs effectively with small to medium-sized biomedical datasets where sample collection is expensive or limited; it can model complex non-linear decision boundaries through kernel functions; it provides strong generalization performance with appropriate regularization; and it is less prone to overfitting compared to more complex models when training data is limited.

The RBF kernel maps the five-dimensional feature space into a higher-dimensional space where classes become more separable through a linear boundary. The model was trained on the standardized feature vectors extracted from the voice dataset. During inference, the SVM was configured to output calibrated probability estimates using the predict_proba() method rather than hard binary classifications. This produces a probability score between 0 and 1, where values closer to 1 indicate higher likelihood of Parkinsonian speech patterns. The probabilistic output is essential for the weighted fusion strategy, as it provides a continuous risk score rather than a discrete decision.

E. Spiral Image Preprocessing

Spiral drawing images underwent a standardized preprocessing pipeline to prepare them for CNN input:

Step 1 - Image Loading: Images were loaded from disk in their original format (PNG or JPG) using image processing libraries.

Step 2 - Resizing: All images were resized to a uniform dimension of 224×224 pixels. This standardization is necessary because CNNs require fixed-size inputs, and 224×224 is a common choice that balances spatial resolution with computational efficiency.

Step 3 - Array Conversion: The resized images were converted to numerical array format with shape (224, 224, 3) for color images or (224, 224, 1) for grayscale, representing pixel values in a matrix structure.

Step 4 - Normalization: Pixel intensity values, originally in the range [0, 255], were normalized to [0, 1] by dividing all values by 255.0. This scaling facilitates faster convergence during neural network training, prevents numerical instability, and ensures that pixel intensities are on a similar scale to other normalized inputs in the network.

F. Convolutional Neural Network Architecture

A Convolutional Neural Network was developed for spiral image analysis, leveraging deep learning's ability to automatically learn hierarchical feature representations [13, 14]. The architecture comprises several key components:

Convolutional Layers: Multiple convolutional layers with learnable filters extract spatial features from the input images. Early layers detect low-level features such as edges, curves, and local texture patterns. Deeper layers combine these low-level features to recognize higher-level patterns including spiral curvature irregularities, tremor-induced oscillations, broken spiral continuity, and spatial inconsistencies characteristic of motor dysfunction.

ReLU Activation Functions: Rectified Linear Unit (ReLU) activation functions are applied after each convolutional layer to introduce non-linearity into the model. The ReLU function, defined as $f(x) = \max(0, x)$, enables the network to learn complex non-linear relationships between input images and diagnostic labels while maintaining computational efficiency and mitigating vanishing gradient problems during training.

Pooling Layers: Max pooling or average pooling layers are interspersed with convolutional layers to progressively reduce the spatial dimensions of feature maps. Pooling provides translation invariance (making the network robust to small spatial shifts in the input), reduces computational requirements, and helps prevent overfitting by reducing the number of parameters in subsequent layers.

Fully Connected Layers: After the convolutional and pooling layers extract high-level features, the feature maps are flattened into a one-dimensional vector and passed through one or more fully connected (dense) layers. These layers integrate the spatial features across the entire image to make the final classification decision.

Output Layer: The final output layer consists of a single neuron with sigmoid activation function, producing a probability estimate between 0 and 1. This probability represents the likelihood that the input spiral drawing exhibits Parkinsonian motor patterns.

The CNN was trained on the spiral dataset using binary cross-entropy loss and an optimization algorithm such as Adam or SGD. During training, the network learns to identify subtle visual patterns that distinguish healthy spirals from those drawn by Parkinson's patients, such as increased irregularity in curve smoothness, tremor amplitude variations, and disruptions in spiral continuity.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

G. Multimodal Fusion Strategy

The predictions from the voice SVM and spiral CNN are combined using a weighted linear fusion approach. The fusion formula is defined as:

$$\text{Final Risk Score} = w_{voice} \times P_{voice} + w_{spiral} \times P_{spiral}$$

where:

- P_{voice} is the probability output from the SVM voice classifier (range: 0 to 1)
- P_{spiral} is the probability output from the CNN spiral classifier (range: 0 to 1)
- $w_{voice} = 0.6$ is the weight assigned to the voice modality
- $w_{spiral} = 0.4$ is the weight assigned to the spiral modality
- $w_{voice} + w_{spiral} = 1.0$ (weights sum to unity)

The weights were assigned based on two primary considerations. First, the voice dataset is substantially larger (831 samples) compared to the spiral dataset, providing more training data and potentially more robust model performance. Second, empirical validation during development suggested that the voice model exhibited slightly better reliability and generalization on held-out data. The 60-40 weighting reflects confidence in each modality while ensuring both contribute meaningfully to the final assessment.

This weighted fusion approach offers several advantages. It allows both modalities to influence the final decision, capturing complementary information about different aspects of motor dysfunction. The explicit weighting acknowledges that modalities may have differential reliability, and the fusion is transparent and interpretable—clinicians can understand how the final score was derived. The linear combination is computationally efficient and does not require additional training beyond the individual models.

H. Multi-Sample Aggregation

To enhance robustness and mitigate the impact of noise or anomalous samples, the framework supports multiple inputs per modality. When multiple voice recordings or spiral drawings are provided, the system processes each sample independently through its respective classifier model and aggregates the results.

For the voice modality, given *n* voice recordings, the aggregated probability is computed as:

$$P_{voice} = (1/n) \Sigma P_{voice,i \ (i = 1 \text{ to } n)}$$

where $P_{voice,i}$ represents the probability output for the i-th voice recording. Similarly, for the spiral modality with m spiral drawings:

$$P_{spiral} = (1/m) \Sigma P_{spiral,j \ (j = 1 \text{ to } m)}$$

This arithmetic mean aggregation strategy improves reliability by reducing variance across samples, mitigating the influence of outliers or poor-quality individual samples, and providing a more stable estimate of the patient's true state. For example, if one voice recording has background noise or one spiral drawing was performed under distraction, the impact of these anomalies is diluted when averaged with other clean samples.

I. Evaluation Metrics

The system's performance was evaluated using standard classification metrics derived from the confusion matrix, which tabulates true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN):

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Overall proportion of correct predictions across both classes.

$$\text{Precision} = TP / (TP + FP)$$

Proportion of positive predictions that are truly positive. High precision indicates few false alarms.

$$\text{Recall (Sensitivity)} = TP / (TP + FN)$$

Proportion of actual positive cases correctly identified. High recall indicates few missed cases.

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Harmonic mean of precision and recall, providing a balanced performance measure.

For medical screening applications, recall is particularly critical because it represents the system's ability to detect all true Parkinson's cases. A high recall minimizes false negatives—failing to identify someone who actually has the disease—which is more dangerous than false positives in a screening context. False positives can be filtered out through subsequent confirmatory testing, whereas false negatives result in missed opportunities for early intervention.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

IV. RESULTS

A. Multimodal System Performance

The proposed multimodal framework was evaluated on the test set consisting of 30 label-consistent pairs (15 healthy, 15 Parkinson's). Each test pair combined one voice recording with one spiral drawing sharing the same diagnostic label. The voice SVM and spiral CNN independently generated probability scores for their respective inputs, which were then combined using the weighted fusion formula (w_{voice} = 0.6, w_{spiral} = 0.4) to produce final risk assessments. A threshold of 0.5 was applied to the final risk scores to generate binary predictions for metric calculation.

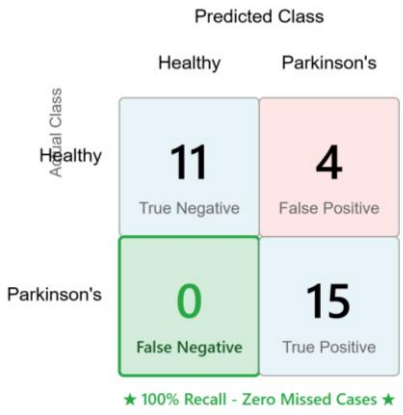

Figure III: Confusion Matrix for Multimodal System Performance.

The matrix demonstrates the system's performance on 30 test samples. The zero false negatives (highlighted in darker green) indicate 100% recall, meaning all Parkinson's cases were correctly identified. Four false positives represent healthy individuals flagged for further evaluation.

The performance metrics are summarized in Table I below:

Table I:
Multimodal System Performance Metrics

Metric	Value	Percentage
Accuracy	0.8667	86.67%
Precision	0.7895	78.95%
Recall	1.0000	100.00%
F1-Score	0.8824	88.24%

The confusion matrix analysis revealed the following classification outcomes: 15 true positives (all Parkinson's cases correctly identified), 11 true negatives (healthy individuals correctly identified), 4 false positives (healthy individuals incorrectly classified as Parkinson's), and 0 false negatives (no Parkinson's cases were missed).

B. Performance Analysis

The system achieved 86.67% overall accuracy, correctly classifying 26 out of 30 test samples. This demonstrates strong classification performance on the multimodal test set, indicating that the weighted fusion of voice and spiral predictions provides reliable risk assessment across both disease and healthy populations.

The most significant finding is the perfect recall rate of 100%, meaning all 15 Parkinson's Disease cases in the test set were correctly identified without exception. This perfect sensitivity is particularly valuable for a screening application, where the primary objective is to avoid missing individuals who may have the disease. In medical screening contexts, the consequence of a false negative—failing to identify a patient with Parkinson's—is far more serious than a false positive, as it represents a missed opportunity for early intervention, delayed access to treatment, and continued disease progression without medical management. The 100% recall ensures that no potentially affected individuals would be overlooked by the screening system.

The precision of 78.95% indicates that approximately 21% of positive predictions were false positives, corresponding to 4 healthy individuals incorrectly flagged as potentially having Parkinson's Disease. While this precision is lower than the recall, it represents an acceptable trade-off for a screening tool designed to prioritize sensitivity. In clinical screening workflows, false positives are addressed through subsequent confirmatory testing by neurologists, including comprehensive neurological examination, response to levodopa medication, and potentially DaTscan imaging. The cost of additional evaluation for 4 false-positive cases is substantially lower than the cost of missing a true Parkinson's case entirely. Furthermore, the 78.95% precision still indicates that the majority of positive predictions are correct, maintaining reasonable specificity.

The F1-Score of 88.24% represents a strong harmonic balance between precision and recall, demonstrating that the system maintains good overall performance while prioritizing sensitivity. This score is notably high considering the deliberate emphasis on recall, suggesting that the multimodal fusion approach successfully captures

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

discriminative patterns without excessively sacrificing precision.

C. Interpretation of Results

The results demonstrate several important characteristics of the multimodal approach. First, the combination of voice and spiral analysis provides sufficient discriminative information to achieve high accuracy on paired samples, even though the pairing is label-consistent rather than from identical individuals. This suggests that the fusion strategy successfully integrates complementary modality-specific patterns.

Second, the perfect recall with reasonable precision indicates appropriate calibration for a screening application. The system errs on the side of caution—when uncertain, it tends toward flagging potential risk rather than dismissing it. This conservative bias is clinically appropriate for early disease detection, where the goal is to cast a wide net and subsequently filter candidates through expert evaluation.

Third, the 86.67% accuracy achieved through weighted fusion likely represents an improvement over either modality in isolation, though direct single-modality results on this specific test set would be needed for formal comparison. The weighted fusion leverages the voice model's strength (60% weight due to larger training dataset) while still incorporating complementary information from spiral analysis (40% weight), creating a more robust combined classifier.

The four false positives warrant consideration. These cases may represent individuals with: subtle motor or voice changes unrelated to Parkinson's Disease (essential tremor, age-related changes, other neurological conditions); poor-quality input data (noisy voice recordings, poorly drawn spirals); or natural variation in voice or drawing patterns that coincidentally resembles Parkinsonian characteristics. In a real clinical deployment, these individuals would undergo neurological evaluation, which would either rule out PD or potentially identify prodromal symptoms that merit monitoring.

The zero false negatives are particularly encouraging, as they indicate the system did not miss any Parkinson's cases despite the relatively small test set and imperfect pairing methodology. This suggests that the multimodal features capture essential disease-related patterns that generalize across the combined modalities.

V. Discussion

A. Multimodal Advantage

The multimodal fusion approach demonstrates clear advantages over single-modality systems by capturing complementary aspects of Parkinson's Disease manifestation. Voice analysis primarily reflects speech motor impairment, laryngeal dysfunction, and respiratory control deficits, while spiral drawing assessment captures fine motor tremor patterns, limb motor coordination, and handwriting-related motor control. These two modalities probe different neuromotor systems affected by dopaminergic degeneration, providing distinct yet complementary diagnostic signals.

This integration provides natural redundancy and robustness to the screening process. If one modality produces an uncertain or noisy result due to technical issues—such as background noise in voice recordings or poor image quality in spiral drawings—the other modality can compensate, leading to more stable and trustworthy predictions. For example, an individual with mild voice changes but pronounced tremor would be captured strongly by the spiral analysis, while someone with significant speech dysfunction but less visible hand tremor would be identified through voice features. The weighted fusion strategy acknowledges the differential reliability of each modality (60% voice, 40% spiral) based on dataset size and empirical performance, while ensuring both contribute meaningfully to the final assessment.

Furthermore, the multimodal approach reduces vulnerability to confounding factors that might affect a single modality. Essential tremor, for instance, might impact spiral drawings but would not produce the characteristic voice changes seen in Parkinson's Disease. Similarly, temporary voice hoarseness from a cold would not be accompanied by Parkinsonian spiral patterns. The requirement for concordant abnormalities across modalities increases specificity while the parallel assessment maintains high sensitivity.

B. Clinical Applicability

The proposed framework is designed as a non-invasive, cost-effective screening tool that could be deployed in diverse healthcare settings. The required inputs—voice recordings and spiral drawings—can be easily obtained without specialized equipment. Voice recordings can be captured using standard smartphones, consumer microphones, or clinic recording equipment.

Spiral drawings require only paper and a writing instrument, with subsequent digitization via smartphone camera or scanner. This accessibility makes the system suitable for deployment in primary care clinics, community health centers, rural healthcare facilities with limited neurological expertise, and potentially for patient self-assessment in home settings, subject to appropriate medical oversight.

The web-based interface implemented using Streamlit provides an accessible platform for healthcare providers and potentially for supervised patient use. The system generates probability-based risk scores rather than definitive diagnoses, with clear visual indicators (color-coded risk zones: green for low, yellow for moderate, red for high) and downloadable PDF reports that can be shared with referring physicians. This design supports integration into existing clinical workflows, where positive screening results would trigger referral to neurology specialists for comprehensive evaluation.

The 100% recall achieved by the system is particularly important for clinical screening applications. The primary goal of a screening tool is to identify all individuals who may have the condition, even if it means accepting some false positives. The four false positives in the test set (13.3% false positive rate) represent an acceptable trade-off, as these individuals would undergo neurological examination that would either rule out Parkinson's or potentially identify other conditions requiring attention. In contrast, a single false negative could mean months or years of delayed diagnosis for a patient who could have benefited from early therapeutic intervention.

The system's probabilistic output (0-100% risk score) provides additional clinical value beyond binary classification. Risk scores in intermediate ranges (e.g., 40-60%) might prompt watchful waiting with periodic reassessment, while scores above 70% could trigger urgent neurological referral. This gradation allows for nuanced clinical decision-making appropriate to the individual's risk profile.

C. Limitations

Several important limitations must be acknowledged when interpreting these results. First, the multimodal test set is relatively small, containing only 30 samples. While this is sufficient for preliminary evaluation, larger-scale validation studies with hundreds or thousands of participants are needed to establish robust performance estimates and identify potential edge cases or failure modes.

Second, and most critically, the test set was created through label-consistent pairing rather than true paired data from the same individuals. Each voice-spiral pair came from different people who shared only their diagnostic label. This methodological limitation introduces uncertainty regarding how the system would perform on genuinely paired multimodal samples from identical patients. True paired data would exhibit correlations between modalities (e.g., patients with severe voice impairment might also show severe tremor), which could either enhance or complicate fusion performance. The current results demonstrate that the fusion strategy can integrate complementary information across modalities, but validation on authentic paired datasets is essential.

Third, the voice and spiral datasets were collected independently, potentially from different geographic regions, age distributions, disease severity ranges, and data collection protocols. This independence limits generalizability, as the system has not been tested on a cohesive population that fully represents the heterogeneity of real-world clinical presentations. The datasets may not capture the full spectrum of demographic variation (age, gender, ethnicity), disease characteristics (early vs. advanced stage, different PD subtypes), or comorbid conditions that influence voice and motor patterns.

Fourth, the system has not been clinically validated in real-world settings with patients confirmed to have Parkinson's Disease through gold-standard diagnostic procedures. The dataset labels were presumably assigned based on clinical diagnosis, but details about diagnostic criteria, disease duration, medication status, and symptom severity are limited. Prospective clinical validation studies in collaboration with movement disorder specialists are necessary to assess real-world performance, establish clinical utility, and identify potential deployment challenges.

Fifth, the fusion weights (0.6 for voice, 0.4 for spiral) were set empirically based on dataset size and preliminary validation rather than through rigorous optimization procedures such as grid search, cross-validation, or meta-learning. While the chosen weights are reasonable, systematic exploration of the weight space might reveal optimal configurations that further improve performance. Additionally, adaptive weighting schemes that adjust based on input quality metrics (e.g., signal-to-noise ratio for voice, image clarity for spirals) could enhance robustness.

Sixth, the current system performs only binary classification (Parkinson's vs. healthy) without assessing disease severity, progression stage, or symptom profiles.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Parkinson's Disease exists on a continuum from prodromal stages through advanced disease, with substantial heterogeneity in symptom presentation. A more clinically useful system would estimate severity, distinguish between PD subtypes (tremor-dominant vs. postural instability-gait difficulty), and potentially track progression over time through longitudinal assessment.

Finally, the system does not account for potential confounding factors such as other neurological conditions (essential tremor, multiple system atrophy, progressive supranuclear palsy), medication effects (some PD patients may be on dopaminergic therapy that partially ameliorates symptoms), comorbid conditions affecting voice or motor control, or normal aging-related changes that might mimic early PD features.

D. Ethical Considerations

The deployment of AI-based medical screening tools raises important ethical considerations that must be carefully addressed. First and foremost, it is essential that users clearly understand this system provides screening-level risk assessments, not definitive diagnoses. The interface must prominently display disclaimers indicating that positive results require professional medical interpretation and confirmatory testing by qualified neurologists. Misrepresentation of screening results as diagnostic certainty could lead to patient distress, inappropriate treatment decisions, or dangerous delays in seeking proper care.

Privacy and data security represent critical concerns. Voice recordings and handwriting samples constitute sensitive health information that must be protected through encryption during transmission and storage, secure access controls limiting who can view patient data, compliance with healthcare privacy regulations such as HIPAA in the United States or GDPR in Europe, and transparent data use policies informing patients how their information will be handled. If the system is deployed as a cloud-based service, additional safeguards are necessary to prevent unauthorized access or data breaches.

Informed consent procedures must ensure that users understand what data is being collected, how it will be analyzed, what the results mean, and what happens to their data after analysis. Patients should have the right to access their data, request deletion, and understand any research uses of aggregated anonymized data.

Equity and access considerations are paramount. While the system is designed to be accessible and low-cost, efforts must be made to ensure it does not inadvertently create or perpetuate healthcare disparities.

The system should be validated across diverse populations including different age groups (young-onset PD vs. typical late-onset), genders (PD affects men and women differently), ethnic and racial backgrounds (to ensure features generalize across population groups), and socioeconomic contexts (to confirm accessibility for underserved communities). Language barriers must also be addressed, as the current voice model was trained on Italian speakers and may not generalize to other languages without retraining.

There is also a risk of algorithmic bias if the training data is not representative. For example, if the dataset predominantly includes advanced-stage patients, the system might miss early-stage cases. If it includes primarily one demographic group, it might perform poorly on others. Continuous monitoring for bias and regular revalidation on diverse populations are necessary.

Finally, the psychological impact of screening results must be considered. A false positive result could cause significant anxiety and distress, while a false negative might provide false reassurance. Clear communication about the probabilistic nature of results, the need for confirmatory testing, and access to counseling or support resources for individuals receiving concerning results are all important elements of responsible deployment.

VI. Future Work

The research presented in this paper establishes a foundation for multimodal Parkinson's Disease screening, but several important directions for future investigation and development have been identified.

Large-Scale Truly Paired Dataset Collection: The most critical next step is the collection of a large-scale dataset containing genuinely paired multimodal samples from the same individuals. Such a dataset should include voice recordings and spiral drawings from each participant, along with comprehensive clinical metadata including confirmed diagnosis by movement disorder specialists, disease duration and stage, medication status and treatment history, UPDRS (Unified Parkinson's Disease Rating Scale) scores, and demographic information. Ideally, the dataset would span multiple disease severity levels from prodromal stages through advanced disease, enabling the system to not only detect presence but also estimate severity. A dataset of 500-1000 participants with multiple samples per individual would provide sufficient statistical power for robust model development and validation while capturing the natural heterogeneity of disease presentation.

Clinical Validation Studies: Rigorous prospective clinical validation is essential before any consideration of real-world deployment. Collaboration with neurology departments and movement disorder clinics would enable evaluation of the system's performance on consecutive patients presenting for PD assessment, comparison against gold-standard diagnostic procedures including neurological examination and DaTscan imaging, assessment of inter-rater reliability when the same samples are evaluated multiple times, and determination of optimal risk score thresholds for different clinical contexts. Clinical validation should also investigate the system's performance on diagnostically challenging cases such as early-stage PD with minimal symptoms, atypical parkinsonian syndromes, and patients with comorbid conditions. Longitudinal studies following patients over time would assess whether the system can track disease progression and treatment response.

Mobile Application Development: Development of dedicated mobile applications for iOS and Android platforms would significantly enhance accessibility and ease of use. A well-designed mobile app could enable in-app voice recording with quality assessment to ensure adequate signal-to-noise ratio, guided spiral drawing using touchscreen input or photographed paper spirals with automatic image quality checking, immediate on-device risk assessment or secure cloud-based processing, longitudinal tracking allowing users to monitor changes over time, and integration with electronic health records for seamless clinical workflow incorporation. Mobile deployment would be particularly valuable for screening in underserved areas, home-based monitoring for patients with mobility limitations, and population-level screening campaigns.

Additional Modality Incorporation: Expanding the multimodal framework to include additional complementary data sources could further improve diagnostic accuracy and provide a more comprehensive assessment. Promising modalities include gait analysis using smartphone accelerometers or computer vision to capture stride length, walking speed, and postural stability; additional handwriting tasks such as sentence writing, figure copying, or repetitive movement tests; cognitive assessments including tests for executive function, memory, and attention which are often affected in PD; facial expression analysis to detect hypomimia (reduced facial expressivity) characteristic of PD; and keystroke dynamics analyzing typing patterns which reflect fine motor control.

Each modality captures different aspects of the disease, and their integration through advanced fusion techniques could yield a more robust and informative screening tool.

Real-Time Voice Capture and Analysis: Implementing real-time voice capture and analysis capabilities would enable immediate feedback during clinical encounters or self-assessment sessions. This would require optimization of feature extraction algorithms for computational efficiency, development of streaming analysis pipelines that process audio incrementally, and user interface design that provides clear real-time guidance on recording quality and duration. Real-time analysis could also enable interactive assessments where the system prompts users with specific phonation tasks or speech exercises designed to elicit diagnostically relevant vocal patterns.

Advanced Fusion Techniques: While the current weighted linear fusion provides interpretability and reasonable performance, exploring more sophisticated fusion approaches could yield improvements. Attention mechanisms could learn to dynamically weight modalities based on input quality or individual characteristics, assigning higher weight to the voice modality when audio quality is excellent but reducing its influence when background noise is detected. Meta-learning approaches could optimize fusion parameters across diverse populations and adapt to individual patient characteristics. Deep multimodal fusion using neural networks could learn non-linear interactions between modalities that simple weighted averaging cannot capture. Ensemble methods combining multiple fusion strategies could provide robust predictions with uncertainty quantification.

Severity Estimation and Progression Monitoring: Extending the system beyond binary classification to estimate disease severity and monitor progression would greatly enhance clinical utility. This would involve developing regression models to predict UPDRS scores or other continuous severity measures, multi-class classification to distinguish between PD stages (early, moderate, advanced), and longitudinal modeling to track individual trajectories over time and detect acceleration or deceleration in progression potentially indicating treatment effects or disease milestones. Progression monitoring could help clinicians optimize medication dosing, assess treatment efficacy, and identify patients who might benefit from advanced therapies such as deep brain stimulation.

Enhanced Explainability Features: Providing detailed explanations for risk assessments would increase clinical trust and educational value.

Future versions could include feature importance visualization showing which acoustic or spatial features contributed most to the prediction, attention maps for spiral images highlighting regions that influenced the CNN's decision, comparative analysis showing how the patient's features compare to normative ranges and typical PD patterns, and confidence intervals or uncertainty estimates indicating prediction reliability. Explainability is particularly important for clinical adoption, as healthcare providers need to understand the basis for algorithmic recommendations to integrate them appropriately into clinical reasoning.

Cross-Lingual and Cross-Cultural Validation: The current voice model was trained on Italian speakers, limiting its generalizability to other language populations. Future work should collect datasets from diverse linguistic backgrounds, develop language-specific or language-agnostic acoustic features, and validate performance across different languages and dialects. Cultural factors may also influence drawing styles, motor norms, and acceptability of different assessment tasks, necessitating culturally sensitive adaptation and validation.

Integration with Wearable Sensors: Incorporating data from consumer wearable devices such as smartwatches and fitness trackers could provide continuous passive monitoring of motor symptoms including tremor detection from accelerometer data, gait analysis from step counting and movement patterns, and circadian rhythm analysis to detect sleep disturbances common in PD. This passive data collection would complement active assessments (voice recordings, spiral drawings) with longitudinal behavioral data captured during daily activities.

VII. Conclusion

This paper presented a multimodal machine learning framework for early screening of Parkinson's Disease through the integration of voice analysis and spiral drawing assessment. The system addresses a critical clinical need for accessible, non-invasive screening tools that can identify at-risk individuals before substantial neuronal loss has occurred, enabling earlier intervention and potentially improved patient outcomes.

The framework combines two complementary modalities that capture distinct manifestations of Parkinson's Disease pathophysiology. Voice analysis employs Support Vector Machines with RBF kernel to classify acoustic features—including mean pitch, pitch standard deviation, jitter, shimmer, and harmonic-to-noise ratio—extracted using Parselmouth.

Spiral drawing analysis utilizes Convolutional Neural Networks to identify spatial patterns indicative of fine motor dysfunction, including tremor-induced irregularities, reduced smoothness, and disrupted spiral continuity. These modality-specific predictions are integrated through a weighted fusion strategy (60% voice, 40% spiral) that accounts for differential reliability while ensuring both contribute meaningfully to the final risk assessment.

Experimental evaluation on a test set of 30 label-consistent multimodal pairs demonstrated strong performance: 86.67% accuracy, 78.95% precision, 100% recall, and 88.24% F1-score. The perfect recall is particularly significant for a screening application, as it indicates that all Parkinson's cases in the test set were successfully identified without exception. This ensures that no potentially affected individuals would be overlooked by the screening system, fulfilling the primary requirement of a medical screening tool. The precision of 78.95%, while lower than recall, represents an acceptable trade-off, as false positives can be addressed through subsequent confirmatory neurological evaluation whereas false negatives represent missed opportunities for early intervention.

While the results are preliminary due to limitations including the small test set size, label-consistent rather than truly paired data, and lack of clinical validation, they demonstrate the potential of multimodal fusion approaches to improve upon single-modality systems. The framework shows that integrating complementary information from voice and drawing analysis enhances diagnostic reliability through redundancy and robustness to modality-specific noise or confounding factors.

The system is designed with practical clinical deployment in mind. The non-invasive nature of data collection (voice recordings and spiral drawings), minimal equipment requirements (smartphone or basic recording device, paper and writing instrument), web-based accessible interface, and probabilistic risk scoring with visual reporting make it suitable for diverse healthcare settings including primary care clinics, community health centers, and potentially home-based screening. The distinction between screening and diagnosis is maintained throughout, with clear communication that positive results require professional medical interpretation and confirmatory testing by qualified neurologists.

Future research directions include collection of large-scale truly paired datasets with comprehensive clinical metadata, prospective clinical validation studies in collaboration with movement disorder specialists, development of mobile applications for enhanced accessibility, incorporation of additional modalities such as gait analysis and cognitive assessment, implementation of advanced fusion techniques including attention mechanisms and meta-learning, extension to severity estimation and progression monitoring capabilities, and enhancement of explainability features to support clinical decision-making. Cross-lingual validation and integration with wearable sensors represent additional opportunities to expand the system's reach and utility.

This research contributes to the growing body of work applying machine learning to early disease detection and represents a step toward accessible, cost-effective screening tools that could facilitate earlier identification of Parkinson's Disease. By enabling screening in settings where specialized neurological expertise may not be immediately available, such systems could support timely clinical intervention, improve access to care for underserved populations, and potentially improve patient outcomes through earlier therapeutic management. The multimodal approach demonstrated here—combining voice and motor assessment through weighted fusion—provides a template for developing robust screening systems that leverage complementary information sources to achieve reliability suitable for real-world clinical application.

REFERENCES

[1] Dorsey, E. R., Sherer, T., Okun, M. S., & Bloem, B. R. (2018). The emerging evidence of the Parkinson pandemic. *Journal of Parkinson's Disease*, 8(s1), S3-S8.

[2] Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., Schrag, A. E., & Lang, A. E. (2017). Parkinson disease. *Nature Reviews Disease Primers*, 3(1), 17013.

[3] Fearnley, J. M., & Lees, A. J. (1991). Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain*, 114(5), 2283-2301.

[4] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering*, 57(4), 884-893.

[5] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nuzket, H., Sakar, B. E., Tutuncu, M., Aydin, T., Isenkul, M. E., & Apaydin, H. (2019). A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. *Applied Soft Computing*, 74, 255-263.

[6] Pereira, C. R., Pereira, D. R., Silva, F. A., Hook, C., Weber, S. A., Pereira, L. A., & Papa, J. P. (2016). A new computer vision-based approach to aid the diagnosis of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 136, 79-88.

[7] Impedovo, D., & Pirlo, G. (2018). Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective. *IEEE Reviews in Biomedical Engineering*, 12, 209-220.

[8] Gil-Martín, M., Montero, J. M., & San-Segundo, R. (2019). Parkinson's disease detection from drawing movements using convolutional neural networks. *Electronics*, 8(8), 907.

[9] Balaji, E., Brindha, D., & Balakrishnan, R. (2020). Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Applied Soft Computing*, 94, 106494.

[10] Aich, S., Younga, K., Hui, K. L., Al-Absi, A. A., & Sain, M. (2018). A nonlinear decision tree based classification approach to predict the Parkinson's disease using different feature sets of voice data. *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 638-642.

[11] Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.38, retrieved from http://www.praat.org/

[12] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.

[13] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

[14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.

[15] Alhussein, M., Muhammad, G., Hossain, M. S., & Amin, S. U. (2021). Cognitive IoT-cloud integration for smart healthcare: Case study for epileptic seizure detection and monitoring. *Mobile Networks and Applications*, 23(6), 1624-1635.

[16] Ramachandran, N., & Polat, K. (2021). Automated detection and classification of Parkinson's disease from spiral drawing tests using deep learning. *Computer Methods and Programs in Biomedicine*, 210, 106388.

[17] Solana-Lavalle, G., & Rosas-Romero, R. (2021). Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 198, 105793.

[18] Prashanth, R., Dutta Roy, S., Mandal, P. K., & Ghosh, S. (2016). High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *International Journal of Medical Informatics*, 90, 13-21.

[19] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's disease progression using machine learning techniques. *Biocybernetics and Biomedical Engineering*, 38(1), 1-15.

[20] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine*, 6(1), 23.