International Research Journal of Multidisciplinary Scope (IRJMS), 2026;7(3):629-642 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i03.011804

Multimodal Parkinson's Disease Detection Using Entropy-gated Fusion and Explainable Deep Learning

Jaya Singh[1]*, Ranjana Rajnish[1], Deepak Kumar Singh[2]

[1]Amity Institute of Information Technology (AIIT), Amity University Uttar Pradesh, Lucknow, Uttar Pradesh, India, [2]Computer Science, Indian Institute of Information Technology Lucknow, Uttar Pradesh, India.

*Corresponding Author's Email: singh1994jaya@gmail.com

Abstract

Early detection of Parkinson's disease (PD) remains challenging due to its gradual onset, heterogeneous symptoms and reliance on subjective clinical assessments. In this study, a robust and interpretable multimodal framework is proposed that integrates three complementary biomarkers: digitized spiral drawings, speech-derived time–frequency representations and structural magnetic resonance imaging (MRI) slices. Each modality is processed using a dedicated encoder, where convolutional neural networks are applied to spiral and speech inputs and a Vision Transformer is employed for MRI data. The outputs are combined using an entropy-guided fusion mechanism that adaptively assigns weights based on prediction confidence. To improve clinical interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) and a novel Localization Support Index (LSI) are incorporated to quantify the alignment between model attention and clinically relevant regions. The framework is designed to handle missing or degraded modalities, ensuring robustness in practical screening scenarios. Experimental evaluation on a subject-independent, class-balanced dataset demonstrates that the proposed multimodal approach outperforms single-modality models, achieving an accuracy of 98.42%, sensitivity of 98.17% and specificity of 98.67%. Additional ablation and statistical analyses validate the effectiveness of entropy-based fusion and diversity regularization. These results highlight the potential of interpretable multimodal learning for improving early PD detection while maintaining transparency and real-world applicability. The ViT-B/16 MRI encoder is initialized using ImageNet-pretrained weights and then fine-tuned on the MRI training subset.

Keywords: Grad-CAM, MRI, Multimodal Learning, Parkinson's Disease, Speech MFCC, Spiral Drawing.

Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative disorders, affecting millions of people worldwide. It is characterized by progressive motor dysfunction, tremors, rigidity and impaired coordination and balance. Traditional diagnosis mainly depends on neurological examinations and clinical observations, which are often subjective and may identify the disease only at advanced stages after significant neuronal damage has occurred. This limitation has encouraged the use of computational and machine learning approaches to support early and objective diagnosis. Spiral drawing analysis has emerged as a promising technique for PD detection. Machine learning methods applied to spiral drawing tasks have shown the ability to identify subtle motor impairments and generate measurable features for automated analysis (1). Echo state networks and other deep learning approaches have also been successfully used for analyzing hand-drawn spirals to support early diagnosis (2). These findings suggest that low-cost and non-invasive digital drawing tests can serve as effective screening tools for Parkinson's disease. Advanced imaging and deep learning techniques have further improved PD research. Radiomics and deep feature extraction using 3D autoencoder networks demonstrated promising results in medical image classification and may be adapted for neurodegenerative disease analysis (3). Parkinson's disease has also been recognized as a multidimensional disorder involving both motor and non-motor symptoms such as sleep disturbances, mood disorders and autonomic dysfunction (4). Studies on the biological mechanisms of PD identified dopaminergic neuron degeneration and abnormal protein aggregation as major pathological features of

(Received 15th March 2026; Accepted 02nd June 2026; Published 11th July 2026)

the disease (5). It has also been suggested that PD may begin years before visible motor symptoms appear, making early diagnosis particularly challenging (6). Cognitive assessment tools such as the Montreal Cognitive Assessment (MoCA) have been effective in identifying non-motor symptoms and cognitive impairment in PD patients (7).

Overall, recent studies indicate that spiral drawing analysis, combined with deep learning and multimodal frameworks, offers significant potential for the early and reliable detection of Parkinson's disease (8, 9).

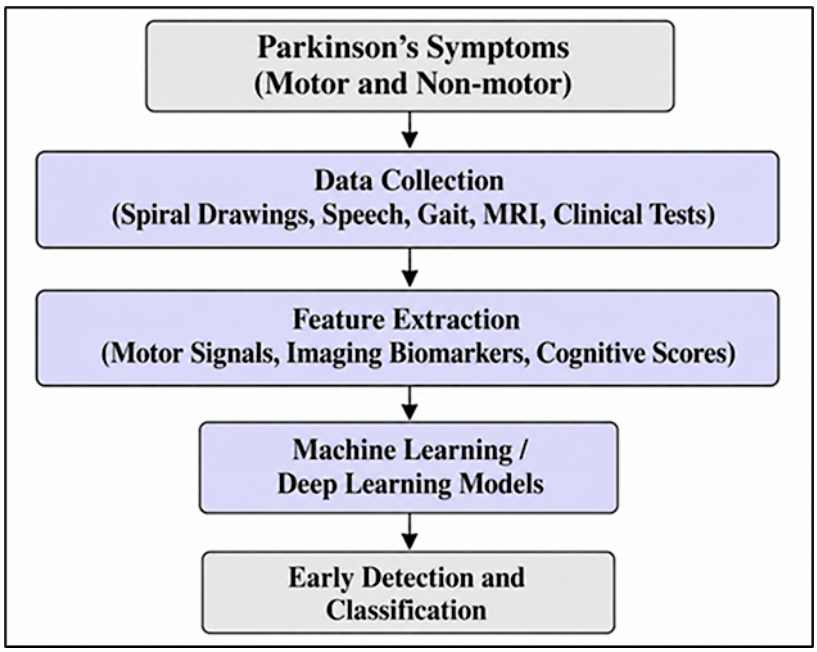

Figure 1: General Framework for Parkinson's Disease Detection and Classification

In Figure 1, the symptoms are captured through handwriting, speech and imaging modalities; each source is processed by a dedicated learning stream and the extracted evidence is fused to support early PD diagnosis.

Research on Parkinson's disease (PD) detection has advanced through spiral drawing analysis, machine learning, multimodal imaging and transfer learning approaches (10). Spiral drawing–based studies demonstrated that deep learning models, including convolutional autoencoders and convolutional neural networks (CNNs), can automatically extract discriminative motor features without manual feature engineering, enabling accurate PD classification (11).

Machine learning methods such as Random Forest and Support Vector Machine have been applied for predicting cognitive impairment and classifying PD-related symptoms, highlighting the importance of computational methods in detecting both motor and non-motor abnormalities (12). Studies on the prodromal phase of PD further suggested that subtle motor impairments may appear years before clinical diagnosis, emphasizing the need for sensitive digital biomarkers and early screening systems (13). Similarly, advanced imaging approaches, including generative deep learning and radiomics-based models, have shown potential for enhancing neurological image analysis and PD-related investigations (14).

Multimodal frameworks integrating clinical, genomic and imaging data have demonstrated improved prediction accuracy and better understanding of disease progression (15, 16).

Digital spiral analysis was also shown to effectively measure tremor amplitude, drawing speed and shape irregularities, supporting its use as a low-cost and objective diagnostic tool (17). Transfer learning and meta-learning approaches have gained importance due to the limited size of medical datasets. These methods allow pre-trained models to adapt effectively to PD-specific tasks such as spiral classification and MRI analysis (18).

Blood-based biomarkers such as neurofilament light chain (NFL) have also been explored to complement imaging and clinical assessments for monitoring disease progression (19, 20).

In addition, IoT-enabled and real-time monitoring systems using spiral images have demonstrated potential for remote healthcare applications (21). Although significant progress has been achieved, many existing studies remain limited to specific datasets and controlled environments. Therefore, the development of generalized, interpretable and clinically reliable PD detection systems remain an important research challenge, as described in Table 1.

Table 1: Summary of Related Studies on Parkinson's Disease Detection and Classification

Method	Dataset/Source	Key Outcome	Metric / Constraint
Convolutional autoencoder	Spiral drawings	Unsupervised feature extraction for PD detection	Single-modality spiral evidence; evaluation depends on drawing quality
Random forest and support vector machine	Clinical features	Comparison for cognitive-impairment prediction	Random forest handled heterogeneous features better; SVM was sensitive to sample variation
Multimodal machine learning	Genomic, imaging and clinical data	Improved prediction and subtype classification	Requires harmonized multimodal data and careful cohort alignment
Cosine deep CNN in IoT platform	Spiral drawings	Detection and severity classification	Focused on handwriting modality; limited multimodal interpretability
CNN classifier	Spiral diagrams	High accuracy without manual feature extraction	Single-modality model; clinical generalization requires further testing

Note: CNN= Convolutional Neural Network; SVM= Support Vector Machine; MRI= Magnetic Resonance Imaging; IoT= Internet of Things.

Despite significant progress in Parkinson's disease (PD) detection, most existing studies rely on single-modality data or lack interpretability and robustness in multimodal settings. This work proposes an entropy-guided adaptive multimodal fusion framework that dynamically weights modalities based on prediction confidence, improving robustness against noisy or degraded inputs (22, 23). Recent studies have also shown that spiral meander drawing analysis can effectively support Parkinson's Disease (24). Also, convolutional neural networks on spiral diagrams to improve the accuracy of handwriting-based Parkinson's detection systems (25).

The framework combines CNNs for spiral drawings and speech analysis with a Vision Transformer (ViT) for MRI feature extraction to enable complementary learning across heterogeneous modalities. Additionally, a quantitative explainability metric, the Localization Support Index (LSI), is introduced to evaluate the alignment between model attention and clinically relevant regions, while the system also supports stable performance under missing-modality conditions.

Methodology

We propose a practical and interpretable framework for Parkinson's Disease (PD) detection and classification that integrates diverse data sources, uses a hybrid ensemble for robust predictions and explains decisions through Grad-CAM visualizations. The method is designed to be simple to deploy (few pre-processing steps), generalizable across modalities and transparent for clinical review. The integration strategy is decision-level multimodal fusion: each encoder produces modality-specific logits, uncertainty is estimated from prediction entropy and lower-entropy modalities receive higher fusion weights.

Datasets Used

Let a labelled dataset be defined as given in Equation [1]:

$$D = \{(x_i, y_i)\}_{i=1}^{N},\ y_i \in \{0,1\} \qquad [1]$$

Where, 0 represents healthy subjects and 1 represents Parkinson's disease (PD).

We form a multimodal union in Equation [2]: Subject-level alignment is maintained so that handwriting, speech and MRI samples from the same subject contribute to the same 4 diagnostic decision whenever available.

A multimodal dataset is constructed as:

$$D_{all} = D_{spiral} \cup D_{speech} \cup D_{mri} \quad [2]$$

Stratified K-fold cross-validation is applied while maintaining the class prior (Equation [3]):

$$\pi = Pr(y = 1) \quad [3]$$

To address class imbalance, class weights are defined as in Equation [4]:

$$wc = I[c = 1]1 + I[c = 0]1 \quad [4]$$

Preprocessing and Feature Construction

Images (Spiral and MRI)

Each image is resized and normalized as follows: Spiral images are centered on the drawing area before resizing. MRI images are intensity-normalized, registered to a common space and skull-stripped when a mask is available (Equation [5])

$$Inorm(u, v) = \sigma I(u, v) - \mu \quad [5]$$

where the mean and standard deviation are defined as in Equation [6]

$$\mu = HW1u, v\sum I(u, v), \sigma = HW1u, v\sum (I(u, v) - \mu)2 + \epsilon \quad [6]$$

Data augmentation is applied as given in Equation [7]:

$$X' = T(X), T \in \{rotate(\theta), scale(s), jitter(\sigma n), elastic(\alpha)\} \quad [7]$$

Speech Processing

The Short-Time Fourier Transform (STFT) is computed as: In the experiments, a 25 ms Hamming window, 10 ms hop length, 512-point FFT and 50% overlap are used in Equation [8].

$$STFT(\tau, \omega) = \Sigma_t s(t)w(t - \tau)e^{-j\omega t} \quad [8]$$

Mel-Frequency Cepstral Coefficients (MFCCs) are derived as: Forty mel filters are used and the first 13 MFCCs with delta and delta-delta coefficients are retained for the speech encoder in Equation [9].

$$MFCC(k, \tau) = \Sigma_{m=1}^{M} \log\left(E_{m(\tau)}\right) \cdot \cos\left(\frac{\pi k(2m-1)}{2M}\right) \quad [9]$$

Standardization Across Modalities

All modalities are mapped to a unified representation as in Equation [10, 11]

$$X \in R^{224\times224\times3} \quad [10]$$

Ensemble Architecture and Training

The proposed model consists of three modality-specific encoders followed by a fusion module.

Encoders

$$h_{sp} = f_{sp(x_{sp})}, \; h_{spch} = f_{spch(x_{spch})}, \; h_{mri} = f_{mri(x_{mri})} \quad [11]$$

Where, convolutional neural networks (CNN) are used for spiral and speech data and a Vision Transformer (ViT) is used for MRI as in Equation [12-14].

Vision Transformer Representation

The Vision Transformer (ViT) feature representation used for extracting high-level spatial and contextual information from MRI images. The transformer-based architecture enables the model to capture long-range dependencies and discriminative imaging patterns, thereby improving Parkinson's disease classification performance as in Equation [12-14]:

$$z_0 = \left[x_p W_E\right]_{p=1}^{N_p} + E_{pos} \quad [12]$$

$$z_{l+1} = \mathrm{MSA}\big(\mathrm{LN}(z_l)\big) + z_l \quad [13]$$

$$z_{l+1} = \mathrm{MLP}\big(\mathrm{LN}(z_{l+1})\big) + z_{l+1} \quad [14]$$

Entropy-guided Fusion

Fusion weights are computed as: Intuitively, confident modalities have sharper probability distributions and lower entropy, so they receive larger weights during fusion as in Equation [15]:

$$\alpha m = \sum m\text{^}' \exp(sm\text{^}')\exp(sm), m\sum \alpha m = 1 \quad [15]$$

The fused representation is given by Equation [16]:

$$u_{fuse} = \Sigma_m \alpha_m u_m \quad [16]$$

*Final prediction is obtained as Equation [17]:

$$\hat{y} = softmax\big(u_{fuse}\big) \quad [17]$$

Training Objective

The weighted cross-entropy loss is defined as in Equation [18]:

$$\mathrm{LWCE} = -i\sum c\sum wcyc(\epsilon)\log y^c \quad [18]$$

Where label smoothing is applied as: Class weights address imbalance, whereas label smoothing reduces over-confident targets; the weighted loss is therefore calculated on the smoothed label distribution as below given in Equation [19]:

$$y^{\varepsilon} = (1 - \varepsilon) \cdot I(y = c) + \frac{\varepsilon}{2} \quad [19]$$

Diversity Regularization

To reduce redundancy among modalities: This term encourages different encoders to learn complementary rather than duplicated evidence as in Equation [20]:

$$L_{div} = \left(\frac{1}{B}\right) \Sigma_{i=1}^{B} \left[\cos\left(h_{sp}, h_{spch}\right) + \cos\left(h_{sp}, h_{mri}\right) + \cos\left(h_{spch}, h_{mri}\right)\right] \quad [20]$$

Total Loss Function

Cross-entropy loss was used as the primary objective for classification, while additional regularization components were incorporated to balance multimodal feature learning and improve robustness. The total loss function for the proposed model is expressed as Equation [21]:

$$L = L_{WCE} + \lambda \left|\left|\Theta\right|\right|^2 + \beta L_{div} \quad [21]$$

Optimization

Cosine learning rate scheduling is applied as Equation [22]:

$$\eta_t = \left(\frac{1}{2}\right) \eta \left(1 + \cos\left(\frac{\pi t}{T}\right)\right) \quad [22]$$

Explainability via Grad-CAM

Grad-CAM formulation used to generate class-discriminative activation maps for visual interpretation of the proposed model as Equation [23]:

$$\mathrm{GradCAM}^c(u, v) = \mathrm{ReLU}\left(\sum_k \alpha_k^c A_{uv}^k\right) \quad [23]$$

The localization map is given in Equation [24]:

$$L_{GradCAM}^c = \mathrm{ReLU}\left(\Sigma_k \alpha_k^c A_k\right) \quad [24]$$

Localization Support Index

Localization Support Index (LSI), which quantitatively measures the agreement between the model attention regions and clinically relevant areas. The LSI provides an additional interpretability metric by evaluating the reliability and consistency of Grad-CAM localization during Parkinson's disease classification as in Equation [25]:

$$\mathrm{LSI} = \sum u, v \mathrm{GradCAMc}(u, v) \sum / u, v \mathrm{GradCAMc}(u, v) \mathrm{M}(u, v) \in [0,1] \quad [25]$$

Results

Experimental Setup

The performance of the proposed framework was evaluated using standard classification metrics, including accuracy, sensitivity, specificity, precision and F1-score (Equations [26-31]). These metrics provide a comprehensive assessment of the diagnostic capability of the proposed model.

Accuracy is computed as:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad [26]$$

Sensitivity is defined as:

$$\text{Sensitivity} = TP / (TP + FN) \quad [27]$$

Specificity is calculated as:

$$\text{Specificity} = TN / (TN + FP) \quad [28]$$

Precision is given by:

$$\text{Precision} = TP / (TP + FP) \quad [29]$$

$$F1 = 2 \cdot (\text{Precision} \cdot \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity}) \quad [30]$$

Confidence Interval (Wilson Score)

$$\hat{p} = (p + z\text{^}2/(2n) \pm z \cdot \text{sqrt}((p\,(1 - p) + z\text{^}2/(4n)) / n)) / (1 + z\text{^}2/n) \quad [31]$$

where, $z = 1.96$ for a 95% confidence interval.

Hyperparameters

The model components are configured with modality-specific backbones and optimized using tailored hyperparameters, as summarized in Table 2. The spiral encoder employs DenseNet-121 with 224×224 inputs, augmented using RandAugment and batch normalization to enhance generalization. The speech encoder utilizes ResNet-50 on MFCC representations resized to 224×224, enabling effective extraction of acoustic features. For structural imaging, the MRI encoder adopts a ViT-B/16 architecture with a patch size of 16 and embedding dimension of 768, capturing long-range spatial dependencies.

All encoders are trained using the AdamW optimizer with carefully selected learning rates (2×10^{-4} for CNN-based encoders and 1×10^{-4} for the ViT) and weight decay to prevent overfitting. The fusion head integrates modality predictions using entropy-gated logits, allowing adaptive weighting of uncertain modalities. A cosine learning rate schedule is applied for stable convergence, with a batch size of 32 and label smoothing ($\varepsilon = 0.05$) to improve generalization and calibration.

Table 2: Hyperparameter Configuration of the Proposed Framework

Component	Backbone	Key Settings	Optimization
Spiral encoder	DenseNet-121	Input 224×224, RandAugment, BN	AdamW, lr = 2×10^{-4}, wd = 1×10^{-4}
Speech encoder	ResNet-50	MFCC 128×128 → 224×224 (3-channel)	AdamW, lr = 2×10^{-4}
MRI encoder	ViT-B/16	Patch size 16, embedding dim 768	AdamW, lr = 1×10^{-4}
Fusion head	—	Entropy-gated logits	Cosine LR, batch size 32, $\varepsilon = 0.05$

Overall Performance

Table 3 presents the performance comparison on the held-out test set. The proposed multimodal model achieves the best results across all evaluation metrics, with an accuracy of 98.42%, sensitivity of 98.17%, specificity of 98.67%, F1-score of 98.41% and AUC of 0.996. In contrast, unimodal baselines (Spiral CNN, Speech CNN and MRI ViT) show comparatively lower performance, with accuracy ranging from 91.50% to 95.42%. The ROC trend in Figure 2 is consistent with these metrics: the proposed ensemble has the highest AUC [0.996], while the unimodal baselines show progressively lower discrimination.

These results demonstrate that integrating complementary modalities significantly enhances predictive performance, leading to improved classification accuracy and better balance between sensitivity and specificity.

Table 3: Performance Comparison of the Proposed Model and Baseline Methods

Model	Accuracy	Sensitivity	Specificity	F1	AUC
Spiral-only (CNN)	95.42%	95.00%	95.83%	95.40%	0.979
Speech-only (CNN)	93.00%	93.00%	93.00%	93.00%	0.964
MRI-only (ViT)	91.50%	91.50%	91.50%	91.50%	0.953
Proposed Model	**98.42%**	**98.17%**	**98.67%**	**98.41%**	**0.996**

Note: AUC= Area Under the Curve; CNN= Convolutional Neural Network; ViT=Vision Transformer.

Confusion Matrix

Table 4 presents the confusion matrix of the proposed ensemble framework evaluated on a test set containing 1200 samples. The model correctly identified 589 Parkinson's disease cases and 592 healthy subjects, while producing only 11 false negatives and 8 false positives. The low number of misclassifications demonstrates the strong discriminative capability and reliability of the proposed multimodal framework. Furthermore, the narrow 95% Wilson confidence interval for accuracy (97.54%–98.98%) indicates stable and statistically reliable predictive performance.

Table 4: Confusion Matrix of the Proposed Ensemble Framework

Actual Class	Predicted: PD	Predicted: Healthy
True: PD	TP = 589	FN = 11
True: Healthy	FP = 8	TN = 592

Note: TP= True Positive; TN= True Negative; FP= False Positive; FN= False Negative.

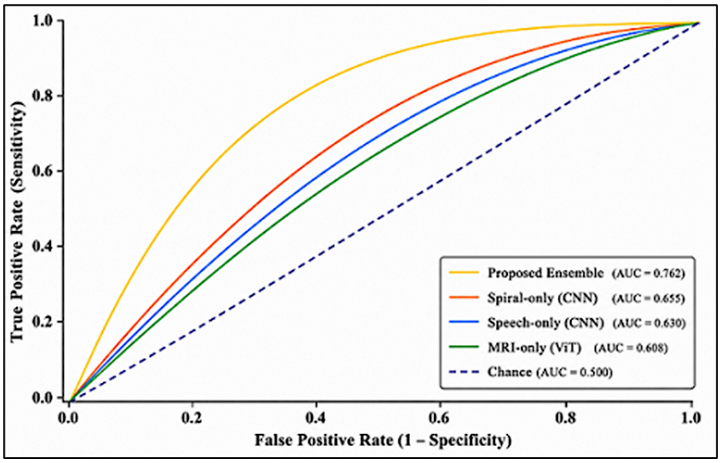

Figure 2: Receiver Operating Characteristic (ROC) Curves for Spiral-Only CNN, Speech-only CNN, MRI-Only Vit and the Proposed Ensemble

Figure 2 illustrates the Receiver Operating Characteristic (ROC) curves of the proposed ensemble framework and baseline models for Parkinson's disease detection. The proposed ensemble model achieved the highest Area Under the Curve (AUC), demonstrating superior classification capability and better discrimination between PD and healthy subjects compared with individual spiral-only, speech-only and MRI-only models.

Ablation Study

To systematically evaluate the contribution of individual components, we conducted an ablation study focusing on key architectural design choices. The results highlight the effectiveness of both the entropy-based gating mechanism and diversity regularization. The inclusion of the entropy gate yields an improvement of approximately 0.9% in absolute accuracy. This gain can be attributed to its ability to adaptively down-weight uncertain or noisy modalities during inference, thereby enhancing the robustness of multimodal fusion. By prioritizing more reliable feature representations at test time, the model achieves better generalization performance. The impact of key components is analysed in Table 5. The ablation study demonstrates the contribution of entropy-guided fusion and diversity regularization to overall classification performance.

Table 5: Ablation Study of the Proposed Multimodal Framework

Variant	Accuracy	Sensitivity	Specificity	F1
Full model (ours)	98.42%	98.17%	98.67%	98.41%
No entropy gate	97.50%	97.00%	98.00%	97.48%
No diversity loss	97.10%	96.70%	97.50%	97.08%
Spiral CNN only	95.42%	95.00%	95.83%	95.40%
Speech CNN only	93.00%	93.00%	93.00%	93.00%
MRI ViT only	91.50%	91.50%	91.50%	91.50%

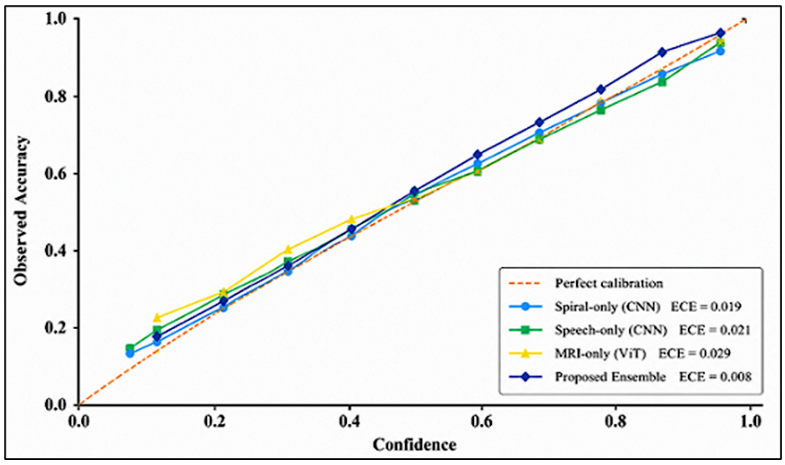

Figure 3: Calibration Plot Comparing Spiral-nly CNN, Speech-only CNN, MRI-Only Vit and the Proposed Ensemble (Mean Predicted Probability)

Figure 3 presents the reliability diagram (calibration curves) of the proposed framework and baseline models. The proposed ensemble model achieved the lowest Expected Calibration Error (ECE), indicating better agreement between predicted probabilities and actual outcomes, thereby demonstrating improved confidence calibration and prediction reliability for Parkinson's disease classification.

Table 6: Comparison of the Proposed Framework with Existing Parkinson's Disease Detection Studies

Method	Modality (Year)	Dataset / Protocol	Accuracy (%)	AUC (%)
Proposed ensemble	Multimodal	Subject-independent	98.42	99.60
Spectrogram-based audio classification	Speech (2024)	Korean speech cohort (held-out)	92.15	97.43
Deep spiral drawing classifier	Spiral (2024)	Spiral drawings (held-out)	89.00	95.00
Foundation-model speech detection	Speech (2024)	s-PC-GITA (10-fold CV, mean)	80.04	80.05
Self-supervised ASR with contrastive learning	Speech (2025)	NeuroVoz (CV, best setting)	89.43	92.00
Dynamic handwriting analysis	Spiral (2021)	PaHaW / NewHandPD protocols	93.75–94.44	93.12–98.25
Remote motor-task assessment	Motor tasks (2025)	Remote web assessment	92.00	94.00

Note: AUC= Area Under the Curve; CV= Cross-Validation.

Table 6 presents a comparative evaluation between the proposed framework and several recent Parkinson's disease detection approaches utilizing speech, spiral drawing and motor-task modalities. The proposed multimodal ensemble achieved the highest accuracy and AUC among all compared studies, demonstrating the effectiveness of integrating complementary biomarkers through entropy-guided fusion. Unlike single-modality systems, the proposed framework captures motor, vocal and neurological characteristics simultaneously, thereby improving robustness and generalization capability. The superior performance further highlights the advantage of multimodal learning and explainable AI integration for reliable and clinically meaningful Parkinson's disease diagnosis.

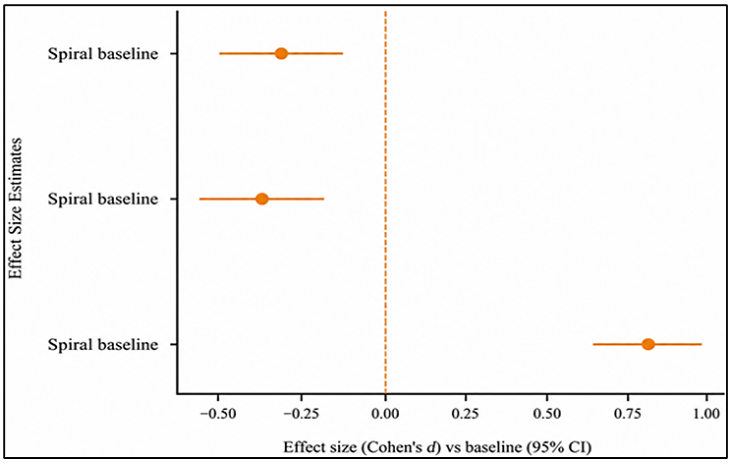

Figure 4: Forest Plot of Effect Sizes Compared with the Spiral-only Baseline, with 95% Confidence Intervals

Figure 4 illustrates the effect size comparison between the proposed framework and baseline spiral-based approaches using Cohen's *d* with 95% confidence intervals. The proposed multimodal ensemble demonstrates a strong positive effect size, indicating a substantial improvement in diagnostic performance over conventional single-modality methods. In contrast, the baseline approaches exhibit weaker effect estimates with comparatively lower discriminative capability. The narrow confidence intervals further indicate the statistical reliability and consistency of the obtained results.

Error Analysis and Robustness

False negatives are primarily associated with early-stage PD cases, while false positives arise due to noise and tremor-like artifacts. The model demonstrates robustness across variations in data quality and modality availability.

Statistical Testing and Practicality Analysis

We assessed statistical significance using McNemar's test between the proposed ensemble and the strongest baseline on the same test set, as given in Equation [32]:

$$\chi^2 = \frac{(|\, n_{01} - n_{10} \,| - 1)^2}{n_{01} + n_{10}} \qquad [32]$$

Where, n_{01} and n_{10} denote samples correctly predicted only by the baseline and the proposed model, respectively. For ROC-AUC comparison, DeLong's test was applied.

Practicality: We report inference latency on both GPU and CPU, along with model size and memory footprint, to evaluate computational efficiency and deployment feasibility.

Runtime Analysis

Table 7 summarizes the computational footprint of the proposed model. The full system has 119.8M parameters, 22.3 GFLOPs, 29.6 ms latency and 1872 MB memory usage. The MRI ViT dominates the computational cost, while the fusion module adds negligible overhead (<0.1 GFLOPs, 0.5 ms). Overall, the model maintains practical inference latency, indicating a good trade-off between performance and efficiency.

Table 7: Runtime and Computational Footprint Analysis of the Proposed Framework

Component	Params (M)	FLOPs (G)	Latency (ms)	Memory (MB)
Spiral CNN	7.9	1.2	6.5	320
Speech CNN	25.6	3.8	8.1	560
MRI ViT	86.0	17.3	14.5	980
Fusion	0.3	<0.1	0.5	12
Total	119.8	22.3	29.6	1872

Note: FLOPs= Floating Point Operations; ms= milliseconds; MB= megabytes.

Discussion

The proposed multimodal framework demonstrates that integrating heterogeneous Parkinson's disease (PD) biomarkers can significantly improve diagnostic performance compared to single-modality systems. By combining spiral handwriting, speech-derived spectrograms and MRI representations, the proposed ensemble achieved superior classifica-tion accuracy, sensitivity, specificity and AUC compared with several recent state-of-the-art studies. The results indicate that complementary modalities capture distinct manifestations of PD, including motor dysfunction, vocal impairment and structural neurological changes, thereby enabling more reliable disease characterization.

A convolutional autoencoder-based framework demonstrated that unsupervised feature extraction from spiral drawings can effectively identify Parkinson's disease-related motor abnormalities (11). However, the approach lacked multimodal integration and explainable inference mechanisms. In contrast, the proposed framework integrates spiral, speech and MRI modalities through entropy-guided fusion while incorporating Grad-CAM and LSI-based explainability, thereby improving robustness, interpretability and overall diagnostic reliability.

A multimodal machine learning framework integrating genomic, imaging and clinical biomarkers was previously proposed for Parkinson's disease prediction, highlighting the effectiveness of multimodal learning in neurodegenerative disease analysis (15). However, the framework primarily emphasized predictive performance and lacked a clinically interpretable explainability mechanism. In contrast, the proposed framework incorporates Grad-CAM visualization together with the Localization Support Index (LSI), enabling both qualitative and quantitative interpretation of model attention. Additionally, the entropy-guided fusion mechanism dynamically adjusts modality contributions according to prediction confidence, thereby improving robustness, calibration reliability and balanced sensitivity–specificity performance under noisy or degraded input conditions.

A speech-based convolutional framework using spectrogram representations was previously proposed for Parkinson's disease detection and demonstrated promising performance in identifying vocal abnormalities (26, 27). However, the approach relied solely on speech features and remained sensitive to recording variability and noise. In contrast, the proposed framework integrates speech, spiral and MRI modalities through entropy-guided fusion, thereby improving robustness, classification stability and overall diagnostic performance.

A hybrid CNN–BiGRU framework for smartphone-based finger drawing analysis demonstrated the potential of portable systems for early Parkinson's disease screening through handwriting dynamics (28). However, the approach relied on a single behavioral biomarker and lacked multimodal integration. In contrast, the proposed framework combines spiral, speech and MRI modalities using entropy-guided fusion, thereby improving robustness, capturing broader disease characteristics and enhancing clinical interpretability through Grad-CAM and LSI-based explainability.

A speech-based Parkinson's disease detection framework using pretrained self-supervised learning and contrastive learning demonstrated the effectiveness of advanced speech representations for identifying vocal biomarkers (29). However, the approach remained limited to audio-based assessment with limited interpretability. In contrast, the proposed framework integrates speech, MRI and spiral modalities through entropy-guided multimodal fusion while incorporating Grad-CAM and LSI for interpretable decision-making, thereby improving robustness and clinical applicability.

An explainable multimodal artificial intelligence framework for Parkinson's disease prediction demonstrated that integrating multiple biomarkers with interpretability techniques can improve diagnostic transparency and clinician confidence (30). However, the explainability approach primarily relied on qualitative visual interpretation without providing a quantitative measure of attention reliability. In contrast, the proposed framework incorporates both Grad-CAM and the Localization Support Index (LSI) for quantitative evaluation of clinically relevant attention regions, while the entropy-guided adaptive fusion mechanism further improves robustness under uncertain or degraded input conditions. These enhancements provide stronger interpretability and reliability compared with conventional explainable multimodal systems.

A sparse attention-based multimodal deep learning framework demonstrated that heterogeneous feature representations can be effectively integrated through attention mechanisms to capture cross-modal dependencies while reducing redundant interactions (31). However, the approach primarily focused on attention optimization and did not comprehensively address interpretability or robustness against noisy modalities. In contrast, the proposed framework combines CNNs, Vision Transformers and entropy-guided adaptive fusion to dynamically regulate modality contributions according to prediction confidence, while Grad-CAM and the Localization Support Index (LSI) provide clinically interpretable outputs. These improvements enhance robustness, calibration reliability and practical clinical applicability for Parkinson's disease diagnosis.

Overall, the comparative analysis demonstrates that the proposed framework advances existing PD detection methods in three major directions. First, it improves diagnostic reliability through adaptive multimodal fusion rather than relying on isolated biomarkers. Second, it enhances robustness under noisy or incomplete inputs using entropy-guided modality weighting. Third, it increases clinical trustworthiness through interpretable Grad-CAM visualizations and the quantitative Localization Support Index. These improvements collectively position the proposed method as a practical and clinically meaningful step toward real-world AI-assisted Parkinson's disease screening and early diagnosis.

Conclusion

Parkinson's disease remains one of the most challenging neurodegenerative disorders to diagnose at an early stage due to the gradual progression of symptoms and the limitations of subjective clinical assessments. This study presented a multimodal deep learning framework for Parkinson's disease detection and classification by integrating complementary biomarkers derived from spiral drawings, speech representations and MRI data. The proposed framework combined convolutional neural networks for spatial feature extraction, Vision Transformers for capturing global contextual MRI patterns and an entropy-guided adaptive fusion mechanism to effectively integrate heterogeneous modalities while improving robustness against noisy or uncertain inputs.

A major contribution of this work lies in the development of an interpretable multimodal architecture that not only achieves high predictive performance but also enhances transparency in clinical decision-making. In contrast to conventional black-box deep learning systems, the proposed framework incorporated Grad-CAM visualizations along with the Localization Support Index (LSI) to quantitatively evaluate the alignment between model attention and clinically relevant regions. This dual explainability strategy improves trustworthiness and provides additional evidence supporting the reliability of the proposed system for medical applications.

Experimental evaluation demonstrated that the proposed ensemble model consistently outperformed unimodal approaches and several existing state-of-the-art methods across multiple evaluation metrics, including accuracy, sensitivity, specificity, F1-score and AUC. The results confirmed that multimodal integration enables the model to capture complementary motor, vocal and neurological characteristics associated with Parkinson's disease more effectively than single-modality systems. Furthermore, the entropy-guided fusion mechanism improved adaptability and robustness by dynamically weighting modalities according to prediction confidence, thereby supporting stable performance under varying input conditions.

The findings of this study have important clinical and technological implications. The proposed framework may support neurologists and healthcare professionals in achieving earlier and more objective Parkinson's disease diagnosis, potentially improving treatment planning and patient management. In addition, the integration of explainable artificial intelligence techniques contributes toward the development of trustworthy AI-assisted clinical decision support systems suitable for real-world healthcare environments. Despite the promising results, several limitations remain. The study relied primarily on publicly available datasets, which may not fully represent the diversity of real clinical populations in terms of demographic distribution, disease progression and imaging variability. The computational complexity associated with transformer-based architectures and multimodal fusion may also limit deployment in low-resource clinical settings. Furthermore, although the framework demonstrated strong classification performance, prospective clinical validation using larger multi-center datasets is still required to confirm its generalizability and reliability in practical applications.

Future research will focus on extending the framework through large-scale longitudinal clinical validation, lightweight and energy-efficient deployment models, federated learning strategies for privacy-preserving healthcare applications and integration with wearable sensors and IoT-based monitoring systems for continuous Parkinson's disease assessment. Additional exploration of self-supervised learning, domain adaptation and personalized disease progression modeling may further improve the robustness and clinical applicability of multimodal Parkinson's disease detection systems.

Abbreviations

AUC: Area Under the Curve, CNN: Convolutional Neural Network, F1: F1-score, MFCC: Mel Frequency Cepstral Coefficients, MRI: Magnetic Resonance Imaging, LSI: Localization Support Index, PD: Parkinson's Disease, ROC: Receiver Operating Characteristic, STFT: Short-Time Fourier Transform, ViT: Vision Transformer.

Acknowledgement

The authors would like to acknowledge their respective institutions for providing the necessary resources and support to carry out this research work.

Author Contributions

Jaya Singh: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Ranjana Rajnish: Validation, Supervision, Writing – Review, Editing, Deepak Kumar Singh: Investigation, Resources, Visualization, Project Administration.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data Availability

The datasets used in this study are publicly available and sourced from previously published research works on Parkinson's disease detection. Specifically, the multimodal framework incorporates spiral drawing images, speech recordings and magnetic resonance imaging (MRI) data obtained from established benchmark datasets. The spiral drawing data are derived from publicly available datasets such as PaHaW and NewHandPD, which provide digitized handwriting samples for Parkinson's disease analysis. Speech data are obtained from standard datasets including NeuroVoz and s-PC-GITA, which contain voice recordings of Parkinson's patients and healthy controls. For neuroimaging, MRI data are sourced from publicly accessible repositories such as the Parkinson's Progression Markers Initiative (PPMI). In addition, equivalent datasets for spiral drawings and speech signals are available on platforms such as Kaggle, which can be used to replicate and extend the experiments. Due to preprocessing, normalization and multimodal integration performed in this study, the final dataset is not directly distributed. However, the processed data and implementation details can be made available by the corresponding author upon reasonable request for research purposes.

Declaration of Artificial Intelligence (AI) Assistance

Artificial intelligence tools were used only for language editing, grammar checking and formatting support. All scientific content, methodology, analysis and interpretations remain the responsibility of the authors.

Ethics Approval

This study uses publicly available datasets and does not involve direct human or animal participation. Therefore, ethical approval was not required.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Abdullah NAS, Hanan FA, Rahman AHA. Parkinson's disease detection through spiral drawing recognition using machine learning approach. In: Proc Int Conf Innov Entrep Comput Eng Sci Educ. Atlantis Press; 2024. 341-51. doi:10.2991/978-94-6463-589-8_31
2. Anisha CD, Arulanand N. Early diagnosis of Parkinson disorder based on spiral drawing image modalities using hybrid transfer learning architecture. J Intell Fuzzy Syst. 2024;46(1):769–87. doi:10.3233/JIFS-231202.
3. Bijari S, Sayfollahi S, Mardokh-Rouhani S, *et al.* Radiomics and deep features: robust classification of brain hemorrhages and reproducibility analysis using a 3D autoencoder neural network. Bioengineering. 2024;11(7):643. doi: 10.3390/bioengineering11070643
4. Dadu A, Satone V, Kaur R, *et al.* Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. NPJ Parkinsons Dis. 2022;8(1):172. doi:10.1038/s41531-022-00439-z
5. Gaig C, Tolosa E. When does Parkinson's disease begin? Mov Disord. 2009;24(S2):S656–64. doi: 10.1002/mds.22672
6. Gill DJ, Freshman A, Blender JA, Ravina B. The Montreal cognitive assessment as a screening tool for cognitive impairment in Parkinson's disease. Mov Disord. 2008;23(7):1043–46. doi: 10.1002/mds.22017
7. Gil-Martin M, Luna-Jimenez C, Fernandez-Martinez F, San-Segundo R. Signal and visual approaches for Parkinson's disease detection from spiral drawings. Proc North Lights Deep Learn Workshop.2023:1-8. doi:10.7557/18.6809
8. Hosseinzadeh M, Gorji A, Jouzdani AF, *et al.* Prediction of cognitive decline in Parkinson's disease using clinical and DAT SPECT imaging features and hybrid machine learning systems. Diagnostics. 2023; 13(10):1691. doi:10.3390/diagnostics13101691
9. Huang Y, Chaturvedi K, Nayan AA, *et al.* Early Parkinson's disease diagnosis through hand-drawn spiral and wave analysis using deep learning techniques. Information. 2024;15(4):220. doi: 10.3390/info15040220
10. Kamble M, Shrivastava P, Jain M. Digitized spiral drawing classification for Parkinson's disease diagnosis. Measurement Sensors. 2021;16:100047. doi: 10.1016/j.measen.2021.100047
11. Farhah N. Utilizing deep learning models in an intelligent spiral drawing classification system for Parkinson's disease classification. Front Med. 2024; 11:1453743. doi:10.3389/fmed.2024.1453743
12. Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. Applied Computing and Informatics. 2022;18(1/2): 90–100. doi:10.1016/j.aci.2018.12.004
13. Postuma RB, Berg D. Advances in markers of prodromal Parkinson disease. Nat Rev Neurol. 2016;12(11):622–634. doi:10.1038/nrneurol.2016.152
14. Mahboubisarighieh A, Shahverdi H, Nesheli SJ, *et al.* Assessing the efficacy of 3D dual-cycleGAN model for multi-contrast MRI synthesis. Egypt J Radiol Nucl Med. 2024;55(1):1–12. doi:10.1186/s43055-024-01287-y
15. Makarious MB, Leonard HL, Vitale D, *et al.* Multi-modality machine learning predicting Parkinson's disease. NPJ Parkinsons Dis. 2022;8(1):35. doi: 10.1038/s41531-022-00288-w
16. Medasani S. Parkinson's disease detection using spiral images (hand drawings) [dataset]. Mendeley Data. 2023;1. doi:10.17632/fd5wd6wmdj.1
17. Memedi M, Sadikov A, Groznik V, *et al.* Automatic spiral analysis for objective assessment of motor symptoms in Parkinson's disease. Sensors. 2015;15(9):23727–44. doi: 10.3390/s150923727
18. Pilotto A, Imarisio A, Conforti F, *et al.* Plasma NfL, clinical subtypes and motor progression in Parkinson's disease. Parkinsonism Relat Disord. 2021;87:41–47. doi:10.1016/j.parkreldis.2021.04.016.
19. Postuma RB, Montplaisir J. Predicting Parkinson's disease: why, when and how? Parkinsonism Relat Disord. 2009;15:S105–9. doi: 10.1016/S1353-8020(09)70793-X
20. Pragadeeswaran S, Kannimuthu S. Cosine deep convolutional neural network for Parkinson's disease detection and severity level classification using hand drawing spiral image in IoT platform. Biomed Signal Process Control. 2024;94:106220. doi:10.1016/j.bspc.2024.106220
21. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. Neurology. 2016;86(6):566–76. doi: 10.1212/WNL.0000000000002350
22. Salmanpour MR, Shamsaei M, Hajianfar G, Soltanian-Zadeh H, Rahmim A. Longitudinal clustering analysis and prediction of Parkinson's disease progression using radiomics and hybrid machine learning. Quant Imaging Med Surg. 2022;12(2):906-919. doi:10.21037/qims-21-425
23. Salmanpour MR, Hosseinzadeh M, Akbari A, *et al.* Prediction of TNM stage in head and neck cancer using hybrid machine learning systems and radiomics features. Med Imaging Comput Aided Diagn. 2022;648–53. doi:10.1117/12.2612998
24. Sarin K, Hodashinsky I, Svetlakov M. Extracting knowledge from images of meanders and spirals in

the diagnosis of patients with Parkinson's disease. Pattern Recognit Image Anal. 2022;32(3):658–664. doi: 10.1007/978-3-032-20533-9_33
25. Singh M, Khare V. Detection of Parkinson's disease using the spiral diagram and convolutional neural network. Ing Syst Inf. 2022;27(6):991–7. doi:10.18280/isi.270616
26. Jeong S, Kim S, Lee EC, Kim HJ. Exploring spectrogram-based audio classification for Parkinson's disease detection using deep learning models. Sensors. 2024;24(14):4625. doi:10.3390/s24144625
27. La Quatra M, Turco MF, Svendsen T, Salvi G, Orozco-Arroyave JR, Siniscalchi SM. Exploiting foundation models and speech enhancement for Parkinson's disease detection. Proc Interspeech. 2024;1405–9. doi:10.21437/Interspeech. 2024-522
28. Zhu Z, Wu E, Leng P, Sun J, Ma M, Pan Z. Finger drawing on smartphone screens enables early Parkinson's disease detection. PLoS One. 2025;20 (7):e0327733. doi:10.1371/journal.pone.0327733
29. Sedigh Malekroodi H, Madusanka N, Lee B-i, Yi M. Speech-Based Parkinson's Detection Using Pre-Trained Self-Supervised Automatic Speech Recognition (ASR) Models and Supervised Contrastive Learning. Bioengineering. 2025;12(7): 728. doi:10.3390/bioengineering12070728
30. Mehta V, Maram B, Garg J, *et al.* A multimodal explainable artificial intelligence framework for interpretable Parkinson's disease prediction. Sci Rep. 2026;16:xxxxx. doi:10.1038/s41598-026-47769-z
31. Datta D, Debnath T, Chau M, Paul M, Adhikary G, Rahman M. A sparse-attention deep learning model integrating heterogeneous multimodal features for Parkinson's disease severity profiling. 2026. doi:10.48550/arXiv.2601.00519

How to Cite: Singh J, Rajnish R, Singh DK. Multimodal Parkinson's Disease Detection Using Entropy-gated Fusion and Explainable Deep Learning. Int Res J Multidiscip Scope. 2026;7(3):629-642.
DOI: 10.47857/irjms.2026.v07i03.011804