RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26323

Multimodal Neuroimaging Fusion using 3D Residual Networks and Vision Transformers for Parkinson's Disease Detection

Bharath Kumar Nagaraj [a], R. Kishore Kanna [b,*], Jacintha Jayavani Jayakaran [c], S. Vijayaraj [d], M.K. Soundarya [e], N. Aravindha Babu [f]

[a] AI, EXL New York City, USA.

[b] Department of Biomedical Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai 600062, Tamil Nadu, India.

[c] Stanley Medical College, George Town, Chennai-600001, Tamil Nadu, India

[d] Department of Electrical and Electronics Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai-600117, Tamil Nadu, India.

[e] School of Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai-600117, Tamil Nadu, India.

[f] Department of Oral and Maxillofacial Pathology, Sree Balaji Dental College & Hospital, Bharath Institute of Higher Education and Research (BIHER), Chennai-600100, Tamil Nadu, India.

* Corresponding Author Email: kishorekanna007@gmail.com

DOI: https://doi.org/10.54392/irjmt26323

Received: 17-02-2026; Revised: 02-05-2026; Accepted: 19-05-2026; Published: 30-05-2026

Abstract: The problem of initial and reliable analysis of Parkinson disease (PD) is caused by the heterogeneous nature of its clinical manifestation and indistinct neurobiological alterations of prodromal and early stages of the disease. The study hypothesizes an explainable Deep Learning (DL) framework of automated analysis and detection of PD through multimodal neuroimaging. The clinical data from the Parkinson's Progression Markers Initiative (PPMI), Structural Magnetic Resonance Imaging (MRI) (T1-weighted), resting-state functional Magnetic Resonance Imaging (fMRI), and Diffusion Tensor Imaging (DTI) were jointly leveraged to capture complementary morphological, functional, and microstructural biomarkers. Structural MRI was Z-score normalized, fMRI underwent wavelet denoising, and DTI was corrected for eddy currents to reduce noise, standardize intensities, and enhance reliable multimodal feature extraction. 3D Sobel filtering was applied to MRI, fMRI, and DTI to enhance structural edges and anatomical boundaries. A multimodal fusion architecture integrating 3D Residual Networks with vision transformers (3D RN-ViT) was developed to diagnose PD. To ensure clinical transparency, Explainable Artificial Intelligence (XAI) techniques were employed to localize disease-relevant neuroanatomical regions and quantify modality-wise contributions. A Multi-Layer Perceptron (MLP) head was employed for final classification on the extracted multimodal features. Satin Bowerbird Optimization (SBO) was employed for optimal feature selection to classify the best discriminative multimodal structures while reducing redundancy and dimensionality. Experiments were carried out in the Python framework using PyTorch. The proposed 3D RN-ViT attained 99.2% accuracy, 99.2% precision, and a 99.34% sensitivity. Overall, the proposed approach provides a transparent and scalable pathway for the analysis of PD, supporting biomarker discovery and serving as a promising proof-of-concept for AI-assisted neuroimaging systems. Future external cross-cohort validation will be required prior to clinical deployment.

Keywords: Explainable AI, Multimodal Neuroimaging, Deep Learning, Parkinson's disease, Biomarkers.

1. Introduction

Parkinson's disease (PD) is one of the most common progressive neurodegenerative diseases affecting the worldwide population and having a vast socio-economic and health care burden. The basic neurobiological structure of PD is a selective loss of dopaminergic neurons in basal ganglia system. Dopamine is an important member of the catecholamine family of neurotransmitters, and it is the main biochemical messenger responsible for the precise modulation, regulation and execution of voluntary physical motion [1]. But it is very difficult to make a reliable clinical diagnosis at the first stages of this dopaminergic decline. A wide variety of neurodegenerative diseases have been grouped together and are known as Parkinsonian Syndromes (PSs), and the clinical picture is complicated by the presence of this group of syndromes. Although they share similar motor presentations, each syndrome is

caused by different underlying mechanisms of neuro-pathology; for example, idiopathic PD, Multiple System Atrophy (MSA) and Dementia with Lewy Bodies (DLB) are classified as synucleinopathies while others like Corticobasal Syndrome (CBS) are classified as tauopathies [2].

The diagnosis of PD is further complicated by the fact that there are no obvious, objective, non-invasive, systemic biomarkers which can differentiate PD from these mimicking syndromes. The diagnosis of PD is a very complex clinical task as there are no confirmed biological assays (blood test or CSF analysis) able to conclude the pathology in vivo, which often results in late and inaccurate diagnosis [3]. The use of subjective symptomatic assessment has therefore become the mainstay of clinical evaluation today, with advanced imaging techniques like striatal dopamine transporter imaging (DAT) and structural Magnetic Resonance Imaging (MRI) increasingly being used to exclude secondary pathology and to demonstrate the loss of presynaptic dopaminergic terminals [4].

The neuropathological feature of PD at a microscopic level is the extensive accumulation of intracellular protein aggregates which are insoluble and misfolded and form spherical structures that are classically described as Lewy bodies [5]. In contemporary clinical frameworks, non-motor symptoms, including cognitive impairment, profound depression, dysautonomia and sleep dysfunction are recognized as a primary contributor to disability in patients with PD and can be variably exacerbated or improved by the use of the dopamine replacement drugs commonly used to treat motor symptoms [6].

This systemic collapse is caused by a molecular defect in the processing of certain proteins of the neurons. The basic problem underlying PD is the degeneration of neurons in specific brain regions, caused by the buildup of toxic aggregates of α-synuclein protein which form the Lewy bodies mentioned above [7]. The centers of this devastating neurodegeneration are located in the densely packed midbrain area called the Substantia Nigra pars compacta (SNpc), which plays an important role in movement and motor coordination and is continuously producing and releasing dopamine to regulate the initiation, velocity and fluency of voluntary movement patterns [8].

It is this progressive and irreversible neurodegeneration of these dopaminergic neurons in the SNpc that directly causes the "canonical" motor dysfunctions that define the disease clinically. As the dopaminergic tone is lost, patients start to develop a disabling pattern of symptoms, including asymmetric resting tremor, extreme bradykinesia, loss of balance and postural reflexes, marked muscular rigidity, dyskinesia, and a gradual alteration of fine motor control including articulation of speech and writing [9]. The exact cause that triggers this chain of neurodegenerative events still remains largely unknown, although decades of intensive and multidisciplinary work in neurobiology suggest that its origin is highly complex and involves a myriad of hereditary and genetic susceptibilities in interaction with environmental neurotoxic exposures that culminate in the onset of the disease [10].

In addition, the course of PD is highly variable. The rate of PD progression and age of onset of PD varies widely among individuals and it is thus virtually impossible to predict the rate of PD development in any given individual [11]. This inter-individual variability is exacerbated by limitations in one of the biggest challenges in current neurology: diagnostics. Another critical issue about the diagnosis of PD is that this is currently done when the motor symptoms are obvious, and unfortunately at a critical threshold of the loss of the dopaminergic neurons, which cannot be recovered at that point [12].

This slow diagnostic process highlights the need for rapid change towards objective biomarker-based computational frameworks instead of symptom-based clinical diagnostic procedures. Although unimodal neuroimaging (e.g., using only T1-weighted structural MRI) can be useful to establish anatomical context, it brings only partial information regarding the complex nature of the neural changes that are seen in PD. Multimodal neuroimaging synergistically needs to be integrated in a comprehensive neurobiological assessment. Structural MRI is able to measure macro-scale structural atrophy and cortical thinning. At the same time, resting-state functional MRI (rs-fMRI) identifies the spatiotemporal modulations of the blood-oxygen-level-dependent (BOLD) functional signals, which shows that the basal ganglia and wide-spread cortices lose their functional connectivity very early. Further, Diffusion Tensor Imaging (DTI) offers exquisite and quantifiable measures of white matter microstructural integrity, and allows for detection of axonal degradation and structural disconnectivity well prior to macroscopic tissue loss as measured by standard anatomical imaging.

While the high dimensional data provided by sMRI, rs-fMRI, and DTI has a great potential for diagnostics, the extraction, synthesis and interpretation of all of this data is a great computational challenge. The highly complex, non-linear spatial and temporal interdependencies that are spread out across these disparate neuroimaging modalities are beyond the mathematical scope of modeling with traditional machine learning and conventional statistical analysis. In recent years, Deep Learning (DL) architectures have become popular and promising paradigms for medical image analysis. But, traditional CNN's are inherently constrained by their receptive fields which are local. Although they are very good at localizing textures and identifying anomalies in a specific region, traditional CNNs are not able to capture long-range anatomical

relationships and large-scale network collapses that characterize neurodegenerative diseases such as PD.

To address these deep computational restrictions, in this study, the paradigm-shifting, explainable Deep Learning framework is proposed which synergistically integrates the Three-Dimensional Residual Networks (3D RN) with the Vision Transformers (ViT). The 3D RN component is explicitly designed to serve as a powerful spatial feature extractor based on deep, volumetric, convolutional hierarchy with the use of residual skip connections to capture highly localized anomaly in the morphology and the microstructural information across the 3D multimodal inputs without the vanishing gradient degradation. Next, high-dimensional embeddings obtained from the preceding layers are used to process the Vision Transformer (ViT) architecture. The ViT overcomes the limitations of local processing, and continuously analyses global contextual relationships throughout the whole brain volume, revealing the distributed network breakdown pathognomonic of early-stage PD in a powerful way with multi-head self-attention mechanisms.

In addition, the use of such highly complex computational architectures in clinical medicine is likely to lead to "black-box" models. These models just do not work with clinical integration, where diagnostic trust, transparency and clinical interpretability are absolute requirements. By nature, this framework integrates cutting-edge Explainable Artificial Intelligence (XAI) tools to discover the underlying neuro-biomarkers and neuro-imaging features used by the model to make predictions, to ensure clinical viability. Furthermore, dealing with this huge dimensionality of multimodal brain data is challenging as it may lead to overfitting the algorithms, increase the computational burden, and result in poor generalization to other unseen patient cohorts. To solve this problem, we apply a very advanced, biologically inspired meta-heuristic algorithm called Satin Bowerbird Optimization (SBO) for dynamically performing optimal feature selection and at the same time tuning the architectural hyper parameters.

This work combines the hierarchical spatial extraction ability of 3D Residual Networks, the global self-attention mapping of Vision Transformers, the powerful dimensionality reduction of the SBO algorithm and the clinical transparency of XAI, forming a highly precise, scalable and fully transparent multimodal diagnostic pipeline. This study assumes that this multi-tiered computational approach will successfully disentangle this complex neurobiological change in prodromal and early-stage PD, serving as a robust computational proof-of-concept for AI-supported neuroimaging systems. While demonstrating high potential to assist in diagnosis and biomarker identification, rigorous external validation across diverse cohorts remains a prerequisite before such frameworks can be safely integrated into routine clinical deployment.

1.1 Research Objective and Key Contributions

The research aims to develop a 3D RN-ViT model with a multimodal DL approach to recognize and classify early signs of PD through structural MRI, resting-state fMRI and Diffusion Tensor Imaging (DTI) of the PPMI and refine these features as well as hyper parameters to increase reliability, interpretability, and biomarker discovery. The key contributions are listed below;

- An explainable 3D RN-ViT framework is presented that is used to automatically diagnose PD using multimodal neuroimaging.
- Structural MRI, resting-state fMRI, DTI, and clinical data in the PPMI are combined to achieve complementary biomarkers.
- SBO enables the most effective features and hyperparameter optimization. The XAI methodologies enhance interpretability, dependability, and clinical transparency.

The research structure contains a rationale of PD that is provided in Section 1. Section 2 focuses on relevant multimodal and DL research. Section 3 outlines preprocessing, the 3D RN-ViT architecture, SBO optimization, and incorporation of XAI. Section 4 outlines the findings of the experiment. The findings are discussed in Section 5, where the conclusions are provided.

2. Literature Review

Here we have presented the literature review in table 1which conveys the proposed research's need and requirement to overcome the limitations which has been clearly addressed in the current existing models for the Parkinson's disease diagnosis.

2.1 Research Gap

The research on the diagnosis of PD focuses mainly on the single-modality data, including drawings, T1-weighted MRI, speech signals, EEG, gait sensors, or DAT-SPECT imaging, which limits their strength and applicability. Although there is a discussion of some multimodal methods, they are often faced with the issues of small data, imbalanced classes and low interpretability. In addition, most of the DL models are opaque models that have no clinically meaningful descriptions. Methods of joint feature selection and optimization of hyperparameters are poorly investigated. A scalable, explainable, and fully optimized multimodal system to sense the existence of PD at an initial stage. The proposed architecture is based on a unified 3D RN-ViT model that uses multimodal MRI and other clinical biomarkers of the PPMI to address the current shortcomings.

Table 1. Recent Studies of Proposed Research

Ref	Modality	Main model	Task	Key findings	Main limitations
[13]	Spiral and wave drawings	Hybrid deep transfer learning (VGG19 + GoogLeNet) with LIME-based explainability	Early PD detection from spiral and wave drawings	Hybrid model outperformed other pre-trained networks; LIME improved interpretability of decisions	Relies on drawing tasks; generalization to broader clinical populations not fully established ieeexplore.ieee +1
[14]	T1-weighted MRI (large multi-center database)	Explainable CNN using Jacobian maps and saliency maps	PD vs healthy control classification from structural MRI	Achieved high accuracy and AUC; saliency maps highlighted disease-relevant brain regions and improved explainability	Single-modality MRI; need for multimodal integration and prospective clinical validation
[15]	Speech / voice recordings (early and mid-advanced PD)	ML pipelines (KNN, SVM, Naïve Bayes) vs custom CNN; CFS feature selection	Classification of early vs mid-advanced PD and OFF vs ON treatment states	ML with CFS feature selection achieved performance comparable to CNN; CFS was the most effective selector	Dataset specific to a single language and setting; generalization to wider populations uncertain
[16]	Speech-related clinical voice features	Supervised ML with SVM and ANN using filter and wrapper feature selection	Early PD diagnosis using selected voice features	ANN showed improved performance compared with other classifiers for voice-based PD diagnosis	Depends on engineered speech features; clinical deployment requires robust cross-dataset validation
[17]	Resting-state EEG	CSP + entropy and band-power features; RF, SVM, KNN classifiers	Detection of PD and discrimination of OFF vs ON medication states	Achieved outstanding detection performance using EEG spatial patterns and entropy/band-power features	EEG recording complexity and limited ability to predict longitudinal outcomes restrict real-world use
[18]	Time-series signals (e.g., gait/sensor/voice; hybrid DL)	Hybrid LSTM–GRU deep learning model for PD detection	PD detection using temporal patterns with class imbalance handling	Hybrid recurrent model achieved high accuracy; oversampling strategies improved minority class detection	Generalization limited by dataset size and synthetic balancing; risk of overfitting
[19]	Wearable ground reaction force (GRF) sensor readings	CNN-based hybrid architecture with locally weighted RF	PD detection and motor severity assessment (regression + classification)	Reported strong performance in both severity regression and PD classification from GRF data	Dependence on wearable gait sensors limits applicability in routine clinical environments

[20]	Wearable physiographic / gait data	CNN-based AI framework for gait pattern classification	Monitoring and severity estimation of PD using wearable gait data	AI framework successfully classified gait patterns and provided decision support for severity monitoring	Small sample size constrained scalability and clinical generalization
[21]	IMU-based gait features	ML with multi-tier feature selection; SVM, DT, RF baselines	PD detection using a minimal set of gait features	High accuracy with reduced feature set; however, performance weaker than SVM, DT, and RF baselines	Reliance on wearable gait data hinders large-scale multimodal validation
[22]	Clinical variables + DAT SPECT imaging (PPMI cohort)	Hybrid ML system combining baseline, clinical, and imaging features	Prediction of cognitive decline / mental failure in PD	Clinical features markedly improved prediction; imaging features added limited incremental value	Imaging features showed low generalizability; performance dataset-dependent
[23]	Multimodal biomarkers: EEG, fMRI, speech, gait, accelerometry	Integrative multimodal computational / federated learning framework	Fusion of multimodal PD biomarkers for improved diagnosis/prognosis	Demonstrated potential of multimodal fusion to enhance PD modeling across diverse biomarkers	Classification limited by relatively small, heterogeneous datasets
[24]	Multimodal PET/MR (structural and diffusion MRI, PET)	Deep learning with ResNet variants on multimodal and multi-sequence images	PD vs MSA classification using PET and MRI	Multimodal PET + diffusion approaches were more accurate than single-modality models	Generalization constrained by limited dataset size and center-specific data
[25]	Multimodal MRI + clinical data	ML structures for multinomial Parkinsonian disorder diagnosis	Multinomial classification of Parkinsonian disorders and uncertainty estimation	Adding image biomarkers to clinical data marginally improved performance and interpretability	Class imbalance affected performance; external validation needed
[26]	Multimodal striatal MRI (SWI, functional, T1, diffusion)	Multimodal MRI framework with radiomic/functional/diffusion/volumetric features	Discrimination of MSA-P vs idiopathic PD	Combining radiomic, functional, diffusion, volumetric, and clinical features improved classification accuracy	Limited sample sizes; potential center and acquisition-protocol bias
[27]	Voice recordings + clinical motor severity	Deep multimodal fusion with ensemble and attention, multitask outputs	Voice-based PD detection and motor severity prediction	Hybrid ensemble with attention showed reliable detection and severity forecasting performance	Tiny and imbalanced datasets; limited external generalizability

It combines XAI to have interpretability and uses SBO to optimize features simultaneously and optimize hyperparameters to enhance robustness and clinical reliability.

3. Methodology

The Methodology presents a hybrid multimodal 3D RN–ViT framework for the diagnosis utilizing structural MRI, resting-state fMRI, DTI, and clinical data. Following Z-score normalization, wavelet denoising, eddy current correction, features are extracted using Sobel-based augmentation, 3D RN, and ViT is used for multimodal fusion. SBO enhances feature selection and hyperparameter tuning, facilitating robust, precise, and interpretable classification using XAI, as clearly explained in Figure 1.

3.1 Data Collection

The data in this study were obtained using the Parkinson's Progression Markers Initiative (PPMI), a multicentre project which aims at finding PD biomarkers. To evaluate the proposed multimodal framework, we specifically curated a cohort of subjects who possessed temporally aligned scans across all three required imaging modalities: structural MRI (T1-weighted), resting-state fMRI, and Diffusion Tensor Imaging (DTI), alongside complete clinical metadata.

Following this intersection, the final dataset comprised 400 unique subjects divided into two classes: Class 0 (Healthy Controls, N = 207) and Class 1 (Parkinson 's disease, N = 193). For the purposes of this study, the evaluation was strictly formulated as a binary classification task. While the PPMI dataset contains further subcategorizations (e.g., tremor-dominant), all PD variants were aggregated into the single PD class to establish the baseline diagnostic efficacy of the proposed multimodal framework.

To ensure rigorous evaluation and entirely eliminate the risk of data leakage, all data partitioning was strictly performed subject-wise. This guarantees that no longitudinal or repeated scans from a single participant appeared across different subsets. The dataset of 400 subjects was partitioned into a 60% training set (240 subjects), a 20% validation set (80 subjects), and an independent 20% test set (80 subjects). The 3D RN-ViT model was optimized exclusively on the training and validation cohorts, and the reported performance metrics were derived solely from the isolated, subject-independent test set. The dataset is freely available for research purposes at https://www.ppmi-info.org/, subject to approved data usage agreements.

3.2 Data Preprocessing

Data from the PPMI, encompassing multimodal preprocessing, Z-score normalization for structural MRI, wavelet denoising for fMRI, and eddy current correction for DTI, thereby minimizing noise, standardizing intensities, and facilitating dependable extraction of structural and microstructural features.

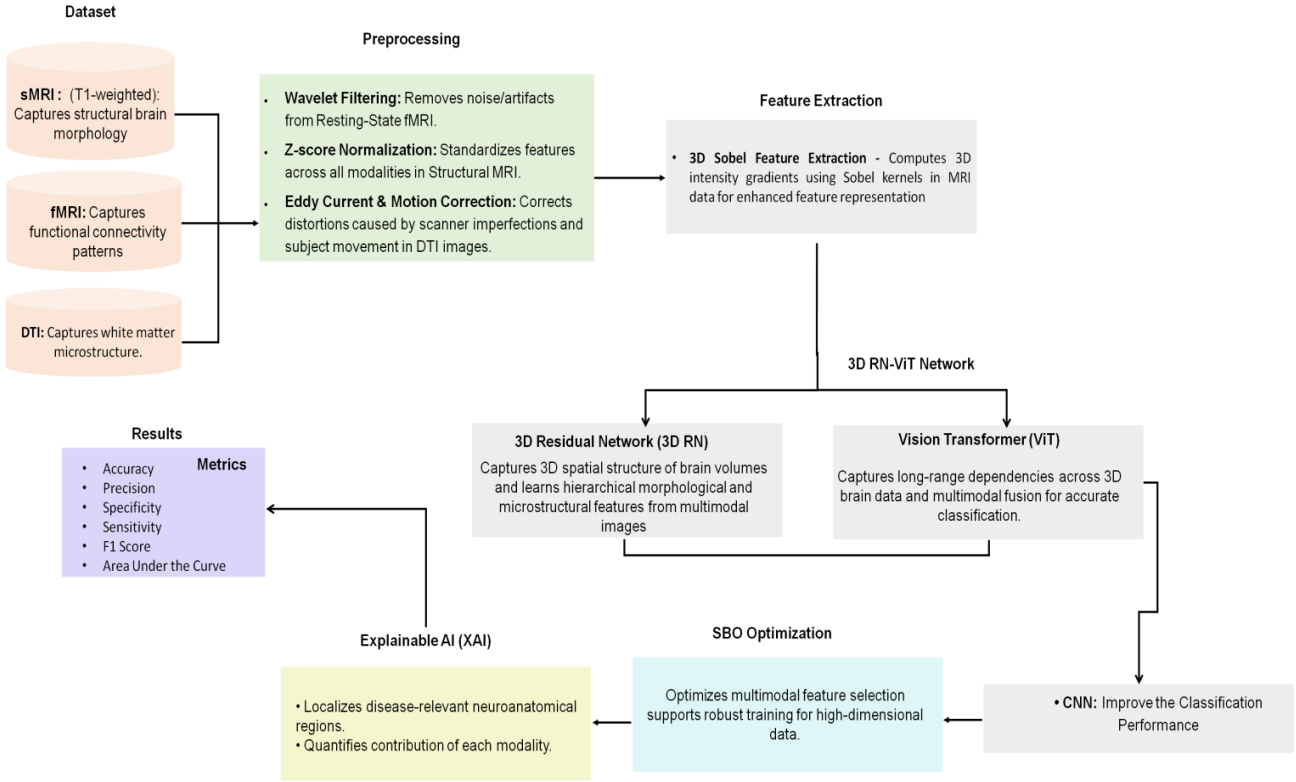

Figure 1. Methodology Flow of 3D RN-ViT Method

3.2.1 Z-Score Normalization for Structural MRI Pre-processing

It was utilized on the preprocessed Structural MRI features to normalize temporal and spatial signals across participants, so that functional connectivity parameters contributed equivalently to the multimodal 3D RN-ViT model for PD diagnosis & MRI was registered to MNI152 space. The normalization was executed as follows in Equation (1):

$$y_j' = \frac{y_j - \mu}{\sigma} \quad (1)$$

Where y_j represents the original MRI voxel intensity or connectivity measure, μ is the mean, σ is the standard deviance across subjects and y_j' is the normalized feature.

3.2.2 Noise Reduction Using Wavelet Denoising in fMRI

Wavelet denoising was utilized on resting-state fMRI signals to mitigate high-frequency noise while maintaining significant neuroimaging data. The procedure employs the Discrete Wavelet Transform (DWT) to decompose the input signal (s) into several frequency bands, where a is the scale parameter to the sentence explaining Equation 2:

$$W(b,c) = \frac{1}{\sqrt{a}} \sum_t y(s)\, \psi(bt - c) \quad (2)$$

Where$W(b,c)$is the wavelet coefficient at scale $b$and translation c,$\psi$is the mother wavelet function, and $y(s)$is the input signal (voxel intensity for MRI/fMRI or diffusion values for DTI).High-frequency coefficients representing noise were thresholded using soft thresholding:

$$\widehat{W}(b,c) = \text{sign}(W(b,c)) \cdot \max\left(|\, W(b,c)\,| - \lambda, 0\right) \quad (3)$$

Where, $\widehat{W}(b,c)$is the denoised coefficient, $\lambda$is the threshold calculated based on noise statistics, and the denoised signal $\hat{y}(s)$ was then reconstructed using the inverse DWT.

3.2.3 Eddy Current Artifact Removal in Diffusion Tensor Imaging

Preprocessing on DTI involved the correction of eddy currents and motion to rectify scanner distortions and subject movement was calculated using the Stejskal-Tanner Equation (4):

$$S(b,\mathrm{g}) = S_0\, e^{-b\, \mathrm{g}^T \mathrm{Dg}} \quad (4)$$

Where $S(b,\mathrm{g})$ is the diffusion-weighted indication, $S_0$the non-diffusion-weighted indication, $b$the diffusion weighting, gthe gradient direction, and $g^T Dg$ denotes directional diffusivity were derived in Equations (5 - 6).

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \hat{\lambda})^2 + (\lambda_2 - \hat{\lambda})^2 + (\lambda_3 - \hat{\lambda})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (5)$$

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (6)$$

Where $\lambda_1, \lambda_2, \lambda_3$ are eigenvalues of D and $\hat{\lambda}$ is their mean. These maps provided microstructural biomarkers for multimodal integration. Figure 2 illustrates the comparisons between pre-processing of three modalities of MRI. Figure 2 (a) MRI structural reveals Z-score standardization to intensity normalization. Figure 2 (b), fMRI, illustrates noise reduction on wavelets in order to better clarify functional signals.

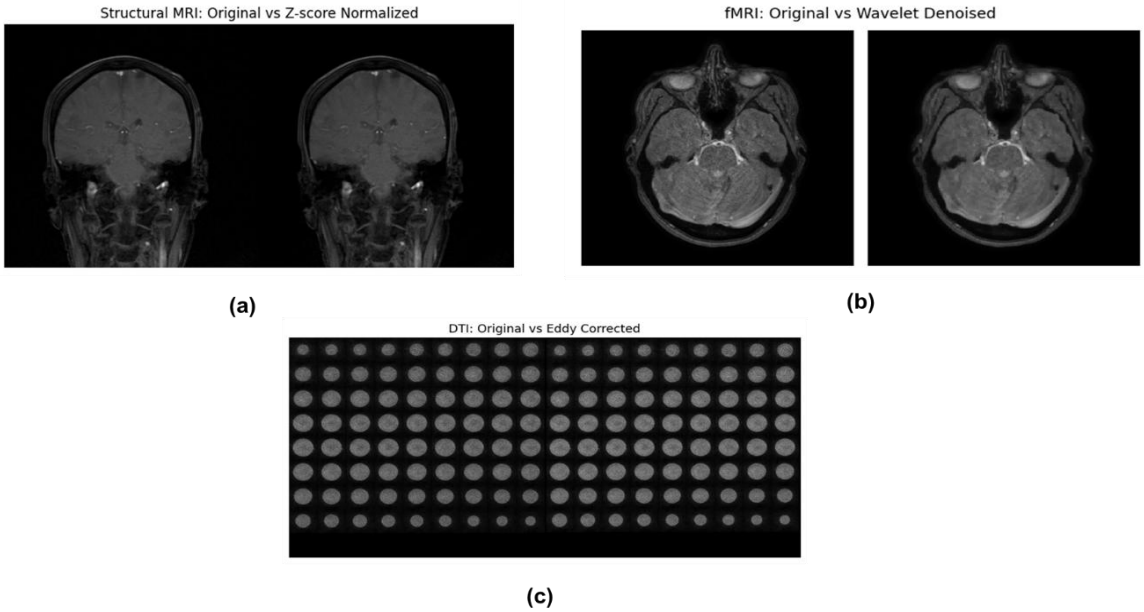

Figure 2. Comparative Image representation of original and pre-processing **(a)** Structural MRI to normalize Z scores **(b)** fMRI to use Wavelet denoising **(c)** DTI images to show effects of eddy correction.

Figure 2 (c), DTI, shows eddy current correction, which decreases distortions as well as increases consistency in the diffusion pattern to collect reliable neuroimaging studies.

3.3 3D Sobel Feature Extraction

To enhance more information to the multimodal Parkinson's disease data using Sobel operator, the T1-weighted MRI, fMRI and DTI were pre-processed and their edges were enhanced to highlight high-frequency anatomical boundaries that are interesting to the condition. Volumetric Sobel kernels were used to compute gradients in three directions as shown in Equations (7 and 8):

$$G_x = I * S_x, G_y = I * S_y, G_z = I * S_z \quad (7)$$

$$G = \sqrt{G_x^2 + G_y^2 + G_z^2} \quad (8)$$

The 3D Sobel operator computes gradients G_x, Gy, G_z of MRI I using kernels S_x, Sy, S_z; the gradient magnitude G. The additional structural features were incorporated into the 3D RN, which allows hierarchical spatial feature extraction.

Figure 3 illustrates the use of Sobel features of edges extracted on the pre-processed data. Figure 3 (a), Figure 3 (b) and Figure 3 (c) show feature extraction on structural MRI, feature extraction on fMRI and Sobel filtering of DTI images respectively. This is because the Sobel operator enhances intensity differences, which can be used to enhance structural boundaries, functional edges, as well as diffusion patterns, and therefore accentuate anatomical detail better and help conduct an accurate analysis of features.

3.3.1 Feature Construction

Additional information on the brain was retrieved by accessing multimodal characteristics. Structural MRI provided three dimensional voxel intensity maps that are reflective of morphology, fMRI provided functional connectivity matrices that are reflective of inter-regional temporal correlations and DTI provided fractional anisotropy and mean diffusivity maps which depict white matter microstructure. The features were standardized to a common 3D volume (e.g., 128x128x128) to be compatible across modalities and the deep learning model can use the anatomical, functional and microstructural patterns to diagnose.

3.4 Multimodal Fusion Architecture using 3D Residual Networks with Vision Transformers (3D RN-Vit)

The proposed approach utilizes a 3D RN for the multimodal integration of sMRI, rs-fMRI, and DTI, therefore extracting extensive spatial-volumetric information. The fused representations are subsequently analyzed by a ViT to model global relationships and facilitate precise prediction of PD.

3.4.1 3D Residual Networks (3D RN) for Spatial Multimodal Fusion

Multimodal fusion structural MRI, resting-state fMRI, and DTI data were utilized in a 3D RN for spatial and volumetric characteristics for multimodal integration. Residual blocks with skip connections were utilized to maintain gradient flow, facilitating deeper architectures and enhanced feature learning.

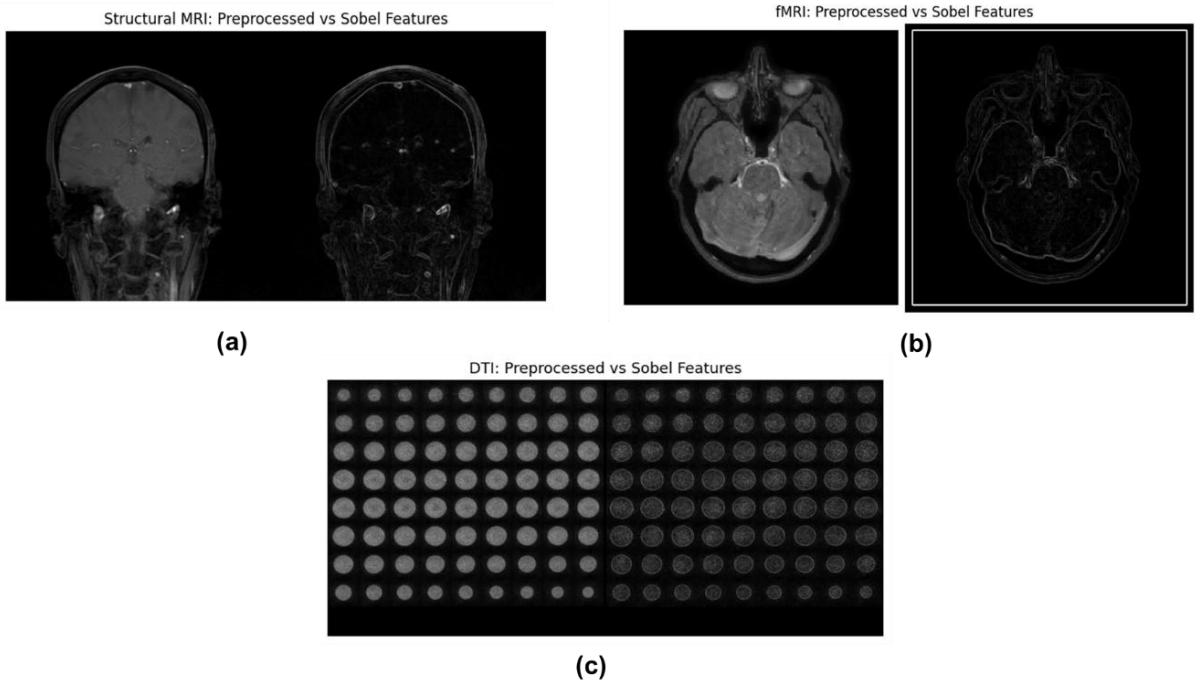

Figure 3. Comparison of processed Vs. Sobel filtering **(a)** Sobel-edge Structural MRI **(b)** Sobel-edge fMRI **(c)** Sobel-edge DTI.

Sequential 3D convolutional layers with a kernel dimension of 2 x 16 x 16effectively recorded local spatial and volumetric relationships across the three modalities. Within each residual block, the feature computation is expressed in Equation (9).

$$Q_j + 1 = Q_j + \Psi(Q_j, W_j) \quad (9)$$

Where $\Psi(\cdot)$ represents the residual mapping, and W_i are the convolution weights, and Q_j represents the feature map output of the j-th layer before adding the residual connection. This architecture enables the network to obtain rich, high-level aspects of structural MRI, fMRI, and DTI, and offers the basis of accurate diagnosis of PD & 3D RN-ViT in the multimodal mode of characterization.

3.4.2. Vision Transformer (ViT) for Multimodal Feature Modelling

Feature maps produced from 3D ResNet are fed into a ViT to identify long-range dependencies and cross-modal interactions across sMRI, fMRI, DTI, and clinical data. Patch embedding and transformer encoders model global context, facilitating precise diagnosis of PD in accordance with the research's purpose. Figure 4 depicts the suggested 3D RN-ViT multimodal framework for the diagnosis of PD. Preprocessed MRI patches are linearly transformed into embeddings and analyzed through a Transformer encoder utilizing multi-head attention and MLP blocks to capture global correlations. Retrieved depictions are integrated with clinical data, refined by SBO, and categorized using an MLP head, while XAI emphasizes disease-relevant areas and modality contributions.

3.4.3 Patch Embedding

$Y \in \mathbb{R}^{H \times W \times D \times C} P \times P \times P$ In the proposed PD multimodal framework, the 3D feature maps generated by the 3D RN from fused sMRI, rs-fMRI, and DTI data, are separated into non-overlapping 3D patches of fixed size. Each patch is flattened and linearly projected into a fixed D-dimensional embedding space to enable global contextual learning for accurate PD diagnosis.

$$y_p^j = E \cdot \text{flatten}(Y_j) + e_{\text{pos}}^j \quad (10)$$

In Equation (10) $Y_j$is the j^{th} 3D patch from fused sMRI, rs-fMRI, and DTI features. E Projects flattened patches into a unified embedding space, e_{pos}^jpreserves spatial structure, and y_p^j represents the embedded multimodal brain region for PD diagnosis.

3.4.4 Multi-Head Self-Attention (MHSA)

The core of ViT is self-attention, which models global relationships among patches and is explained in Equation (11).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_l}}\right)V \quad (11)$$

$QKVd_l$Where,, and are query, key, and value matrices derived from multimodal embeddings. Ensures stable scaling, and output captures global interactions for accurate PD classification.

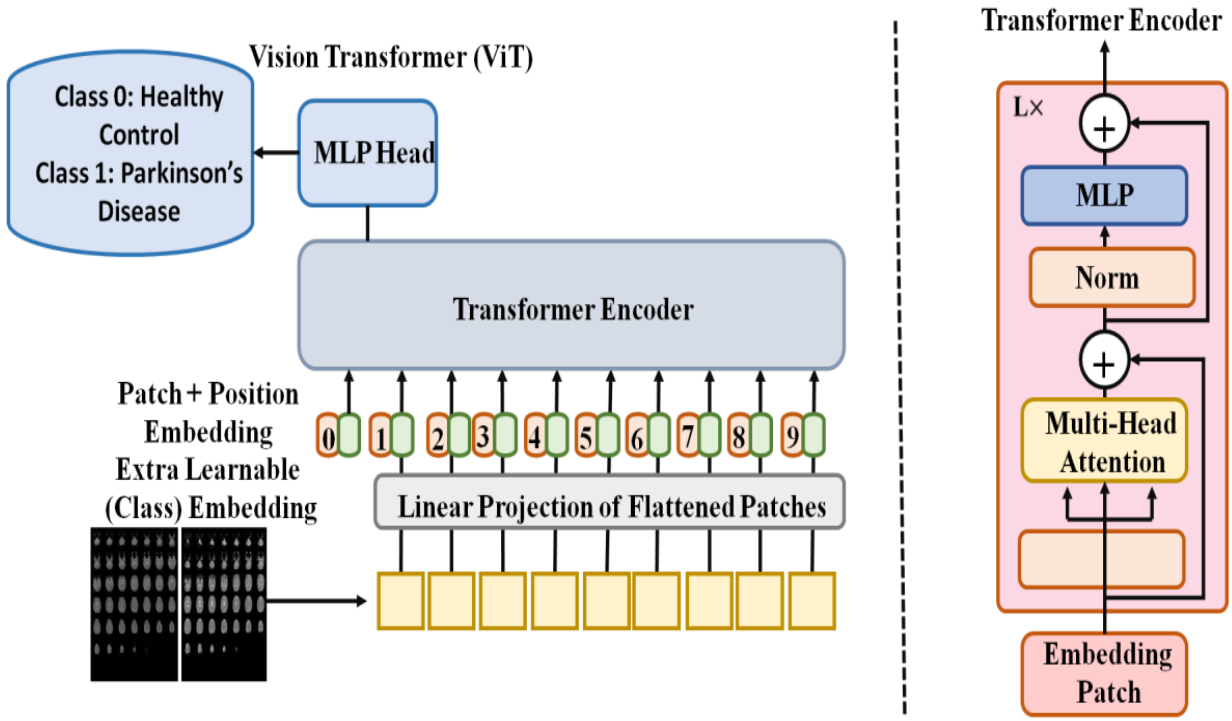

Figure 4. Vision Transformer (ViT) Architecture for PD Prediction

3.4.5 Transformer Encoder Layer

Each transformer encoder layer to enhances the embedded multimodal 3D patches via MHSA, succeeded by an MLP with residual connections and explained in Equations (12 and 13).

$$z' = z + \text{MHSA}(\text{LayerNorm}(z)) \quad (12)$$

$$z_{\text{out}} = z' + \text{MLP}(\text{LayerNorm}(z')) \quad (13)$$

Here, $z$denotes embedded multimodal patches from fused sMRI, rs-fMRI, DTI, and clinical data. LayerNorm ensures stable learning, MHSA captures global cross-modal relationships, and the MLP performs nonlinear transformation. The final output z_{out} yields globally contextualized features for accurate PD analysis.

3.5 Feature Selection Using Satin Bowerbird Optimization (SBO)

The SBO method adopts the hyperparameter optimization and multimodal feature selection to improve the classification performance and the generalizability of the framework. Inspired by satin bowerbirds' mating behavior, it balances global exploration and local exploitation, ensuring efficient search and enhanced performance in high-dimensional neuroimaging and clinical data. The proposed framework is represented as a combination of binary masks of multimodal feature selection (sMRI, rs-fMRI, and DTI) with constant hyper parameters of 3D RN-ViT architecture, including learning rate, batch size, dropout rate, and transformer depth. It is stepwise described in Equations (16 and 17).

3.5.1 Fitness Evaluation

y_jIn the proposed diagnostic paradigm for PD, each candidate solution denotes a particular amalgamation of chosen sMRI, rs-fMRI, DTI characteristics, and model hyperparameters. The fitness is intended by means of the classification error rate:

$$F(y_j) = \frac{N_{misclassified}}{N_{total}} \times 100 \quad (14)$$

Where $N_{misclassified}$ denotes the number of incorrectly predicted PD cases and N_{total} is the total evaluated samples. Lower $F(y_j)$ indicates superior multimodal diagnostic performance.

3.5.2 Position Update

To enhance feature selection and model tuning, candidate solutions are updated as:

$$y_{j,l}^{new} = y_{j,l}^{old} + \beta_l \left[\left(\frac{y_{k,l} + y_{elite,l}}{2} \right) - y_{j,l}^{old} \right] \quad (15)$$

Where $y_{j,l}^{old}$ and $y_{j,l}^{new}$ denote the previous and updated values of the l^{th} multimodal feature or hyperparameter. $y_{elite,l}$Represents the best-performing solution (minimum PD classification error), $y_{k,l}$ is a randomly selected candidate promoting exploration, and β_l controls convergence toward the optimal feature–hyperparameter combination.

The 3D RN is applied to combine the data of multimodal MRI, fMRI, and DTI and it contains hierarchical volumetric relationship between the modalities. The ViT adds to the integrated representations the long-range dependencies, while SBO enhances the discriminative power of the selected features, and a MLP Head classifies them, as illustrated in Algorithm 1.

Algorithm 1. 3D RN-ViT

```
Input: Multimodal dataset (X_{train}, Y_{train}), (X_{val}, Y_{val})
Output: Optimized Feature Mask M*, Hyperparameters H*, Trained Weights W*
def main():
sMRI, fMRI, DTI, clinical, labels = load_PPMI()
sMRI = (sMRI - mean(sMRI)) / std(sMRI)
fMRI = wavelet_denoise(fMRI)
DTI= eddy_current_correction(DTI)
FA, MD = compute_diffusion_maps(DTI)
sMRI = sobel3D(sMRI)
fMRI = sobel3D(fMRI)
FA = sobel3D(FA)
sMRI = resize(sMRI, (128,128,128))
fMRI = resize(fMRI, (128,128,128))
FA = resize(FA, (128,128,128))
fused_input = concatenate([sMRI, fMRI, FA, clinical])
x = fused_input
for i in range(num_res_blocks):
x = x + conv3D(x, weights[i])
x = batch_norm(x)
x = relu(x)
spatial_features = x
patches = split_patches(spatial_features, P)
embeddings = linear_projection(flatten(patches))
embeddings += positional_encoding()
for l in range(num_transformer_layers):
embeddings = embeddings + MHSA(layer_norm(embeddings))
embeddings = embeddings + MLP(layer_norm(embeddings))
global_features = global_pool(embeddings)
y = mlp_head(global_features)
```

```
y = relu(y)
y = fully_connected(y)
predictions = softmax(y)
population = initialize_population()
for iter in range(max_iter):
for candidate in population:
error = classification_error(candidate, labels)
update_fitness(candidate, error)
elite = select_best(population)
update_positions(population, elite)
return predictions
output = main()
print("PD Diagnosis :", output)
```

Table 2 illustrates the hyper parameter of the 3D RN-ViT framework and setting of values applied in the diagnosis of multimodal PD. These parameters are network architecture options, optimization, input resolutions and regularization to strike the balance between model performance, convergence stability and generalization.

3.6 Feature Selection and Tuning via SBO (Outer Loop) and Backpropagation (Inner Loop)

To ensure algorithmic coherence, the training pipeline is structured as a dual-loop optimization framework. The inner loop utilizes standard backpropagation (AdamW optimizer) to update the millions of network weights (W) within the 3D RN-ViT architecture based on Cross-Entropy loss. The outer loop employs Satin Bowerbird Optimization (SBO) as a meta-heuristic wrapper to optimize a binary feature selection mask (M) applied to the input modalities, alongside global hyperparameters (H, e.g., learning rate, dropout).

4. Result and Experimental Setup

Table 3 outlines the hardware and software environment used to train the 3D RN-ViT framework, including high-performance CPUs/GPUs, RAM, storage, operating systems, Python libraries, CUDA support, transformer libraries, and experiment management tools, thus providing efficient computation and reproducibility.

4.1 Evaluation Metrics

Accuracy: It quantifies the ratio of accurately identified PD and healthy individuals among all assessed cases in multimodal diagnosis, explained in Equation (16).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (16)$$

Precision: It measures the proportion of accurately acknowledged PD cases relative to the entire amount of participants diagnosed as having PD and is demonstrated in Equation (17).

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

Table 2. Hyper parameter table

Parameter	Value / Range
Backbone Type	3D ResNet-50 / 101
Transformer Layers	4–8
Embedding Dimension	256–512
Number of Attention Heads	4–8
MLP Hidden Size	512–1024
Patch Size (Temporal x Spatial)	2 x 16 x 16
Dropout Rate	0.1–0.3
Learning Rate	1e-4 – 5e-4
Optimizer	AdamW, SBO
Batch Size	8–32
Input Resolution	16–32 frames, 112x112 or 224x224
Weight Initialization	Pretrained 3D ResNet + ViT
Loss Function	Cross-Entropy / Focal Loss

Table 3. Hardware and Software environment

Category	Specification / Details
Hardware	
CPU	Intel® Core™ i9-13900K, 24 cores, 32 threads
GPU	NVIDIA RTX 4090, 24 GB GDDR6X, CUDA 12.2
RAM	128 GB DDR5, 6000 MHz
Storage	2 TB NVMe SSD (OS + datasets)
OS	Windows 64-bit
Cooling & PSU	1200W Platinum PSU, liquid cooling
Software	
Python	3.11.2
PyTorch	2.1.0+cu121
CUDA / cuDNN	CUDA 12.2 / cuDNN 8.9
Libraries	torchvision 0.16, timm 0.9.2, numpy, pandas, matplotlib
Transformer Support	huggingface/transformers 4.38
Experiment Management	Weights & Biases 0.16.7

Sensitivity: It evaluates the percentage of actual PD cases accurately recognized by the multimodal DL diagnostic model and demonstrated in Equation (18).

$$Sensitivity = \frac{TP}{TP+FN} \quad (18)$$

Specificity: It assesses the percentage of healthy balance participants accurately recognized as non-PD by the diagnostic model and illustrated in Equation (19).

$$Specificity = \frac{TN}{TN+FP} \quad (19)$$

F1 Score: It calculates the harmonic mean of precision and recall, equilibrating False Positive (FP) and False Negative (FN) predictions, as explained in Equation (20).

$$F1\ Score = 2 \times \frac{Precision.Sensitivity}{Precision+Sensitivity} \quad (20)$$

AUC: It determines the prototype's comprehensive capacity to differentiate PD from healthy individuals at all categorization thresholds and is demonstrated in Equation (21).

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (21)$$

Where TP represents accurately recognized Parkinson's cases, TN designates correctly identified healthy controls, FP refers to healthy individuals misclassified as having Parkinson's, and FN indicates missed Parkinson's cases.

4.2 Comparative Performance of the Proposed Method

The proposed 3D RN-ViT model was evaluated on a binary diagnostic task (Healthy Control vs. Parkinson's disease. The confusion matrix in Figure 5 illustrates the performance of the models which include 205 TN, 191 TP, 2 FP and 2 FN. This graphic focuses on correct and incorrect classifications, providing the understanding of the predicted precision and inaccuracies of the model.

The Precision-Recall curve has been plotted in Figure 6(a) and it shows the ability of the model to balance the precision and recall classes. Class 0 has an Average Precision of 99.53, and Class 1 has 98.66, which is an excellent measure in distinguishing between positive samples and minimal false positives. The curve of AUC as shown in Figure 6(b) indicates the equilibrium between the TP Rate and the FP rate. Classes 0 and 1 achieve an AUC of 98.6%, which shows that they have high discriminative strength and reliable category predictions.

Figure 7 (a) shows that the model achieved an approximate performance of 94% precision. Figure 7(b) also shows that the loss during preparation and substantiation reduces to approximately 0.02 and 0.05, respectively, which reflects strong convergence, effective generalization, and insignificant overfitting.

Darker areas are fluid or lower density tissues and the illuminated part represents a more dense part of the anatomy.

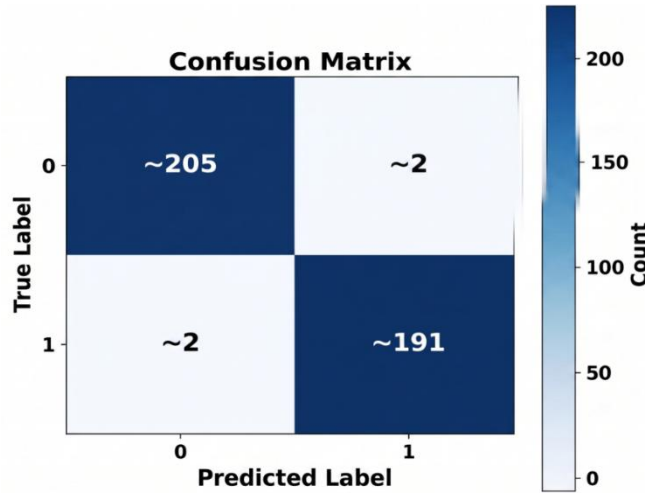

Figure 5. Confusion Matrix of Model Predictions

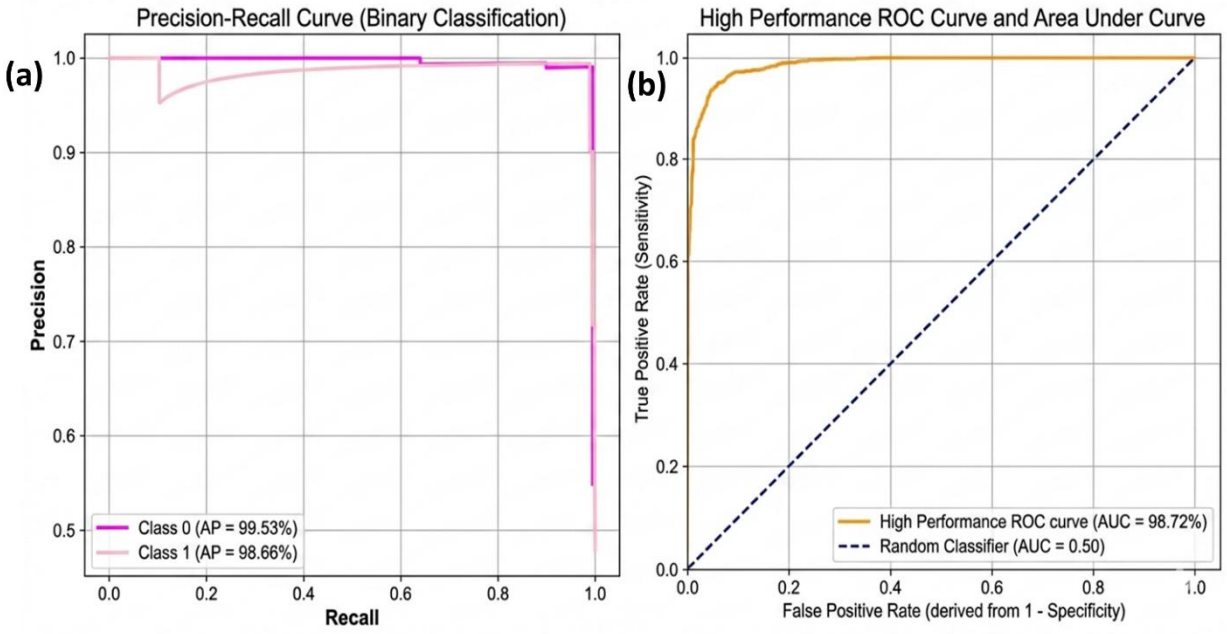

Figure 6. Comparison of PD classification **(a)** Precision-Recall curve and **(b)** AUC curve.

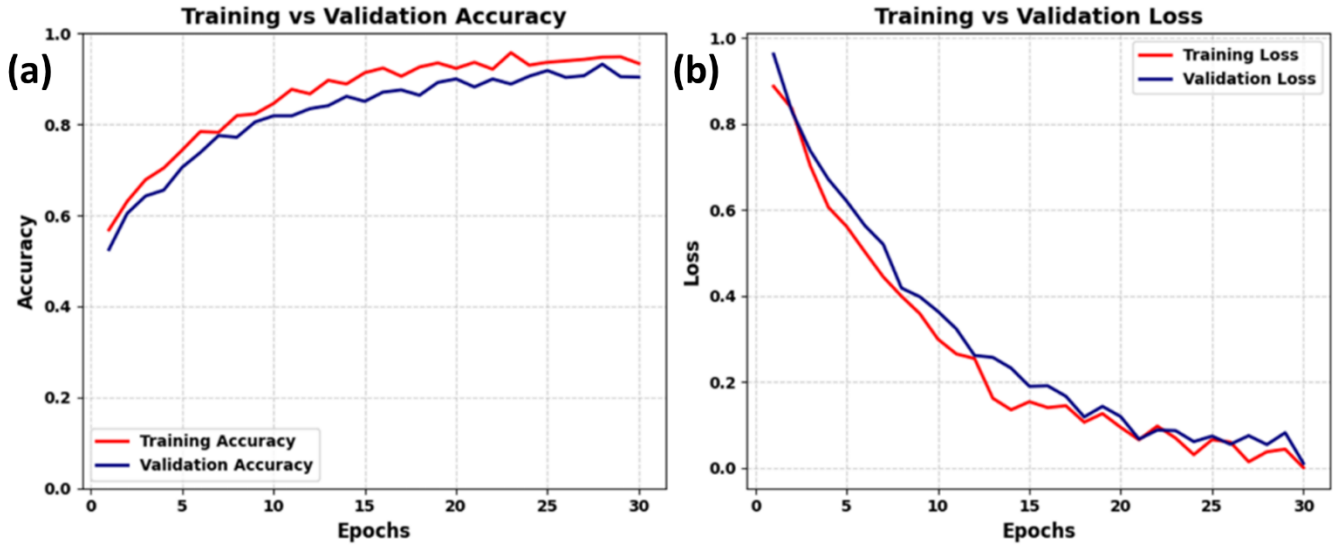

Figure 7. Performance validation on (a) Training vs. Validation Accuracy and (b) Training vs. Validation Loss

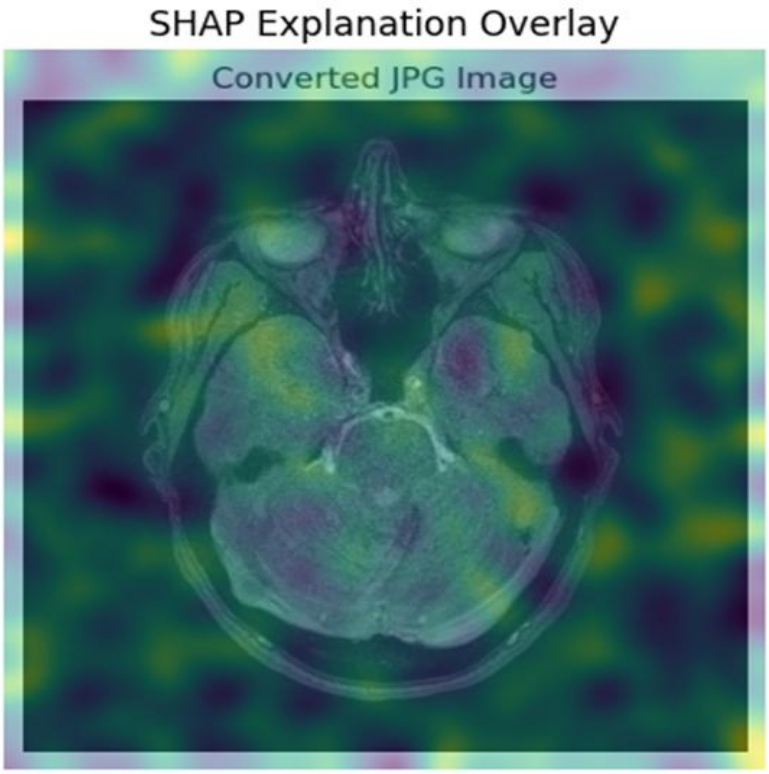

Figure 8. SHAP-based interpretation of MRI for early Parkinson's classification

Figure 8 shows the original structural MRI scan whereby the pixel intensity may be taken as black to white to reflect the characteristics of the underlying brain tissue. The purpose of this image is to furnish anatomical context for the SHAP heatmap, facilitating the viewing of disease-relevant areas that impact the model's prediction of PD.

4.3 Comparative Metrics Evaluation of the Proposed Method against Existing Methods

Fuzzy Rank Level Fusion [28] is used in PD diagnosis where multimodal integration is done with manual rules; Biomarkers-based Feature Selection (BEFS) with Adaptive Artificial Ant Colony Optimization with Hierarchical Processing (AACOAhp) and RF (BEFS+AACOAhp+RF) are used to find features but do not take into account complex spatial-temporal relationships; CNNs are used to find local structures, but they lack global 3D context (BEFS+AACOAhp+RF).

Table 4 and Figure 10 present the improved diagnostic performance of the proposed 3D RN-ViT framework in comparison with the existing multimodal methods. This explains why multimodal fusion and interpretable deep learning can effectively and early detect PD.

In addition, an ablation study was carried out to investigate the structural contribution of each component proposed. The diagnostic accuracy was found to be 95.4% without Vision Transformer module (i.e., using only the 3D RN for classification). Likewise, if the SBO feature selection was turned off (by using all the extracted features without the feature selection process), the accuracy decreased to 96.8%. All these ablation results verify the necessity of synergistically combining the 3D RN, the ViT global attention mechanism and SBO-driven dimensionality reduction to achieve the best possible accuracy of 99.2% reported in this study.

Figure 9 (a) illustrates that the 3D RN-ViT attains the maximum accuracy and sensitivity compared to the other approaches. Accuracy attains 99.2 %, while Sensitivity is 99.34 %, indicating a balanced performance. Figure 9 (b) displays that Precision and F1 Score are optimized using this method. The precision is 99.2%, and the F1 score is 99.1%, indicating dependable predictions. Figure 9 (c) depicts the highest discriminative capability in an AUC. The AUC value is 98.6%, signifying exceptional differentiation between classes. Figure 9 (d) illustrates that Specificity surpasses other methodologies. It attains 98.4%, indicating an exceptional capacity to recognize true negatives.

5. Discussion

The present investigation presents an explainable multimodal deep learning framework fusing, structural magnetic resonance imaging (sMRI), resting state functional MRI (rs-fMRI), and diffusion tensor imaging (DTI), for precise and transparent diagnosis of PD.

Table 4. The comparison of existing performance and the proposed performance of the 3D RN-ViT model

Method	Dataset	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	AUC (%)
Fuzzy Rank Level Fusion [28]	PPMI	98.45	98.84	98.84	97.67	98.84	-
BEFS+ AACOA$_{hp}$+RF [29]	PPMI	93	93	93	-	93	97
CNN [30]	PPMI	96	-	96.8	95.8	-	95
3D RN-ViT [Proposed]	PPMI	99.2	99.2	99.34	98.4	99.1	98.6

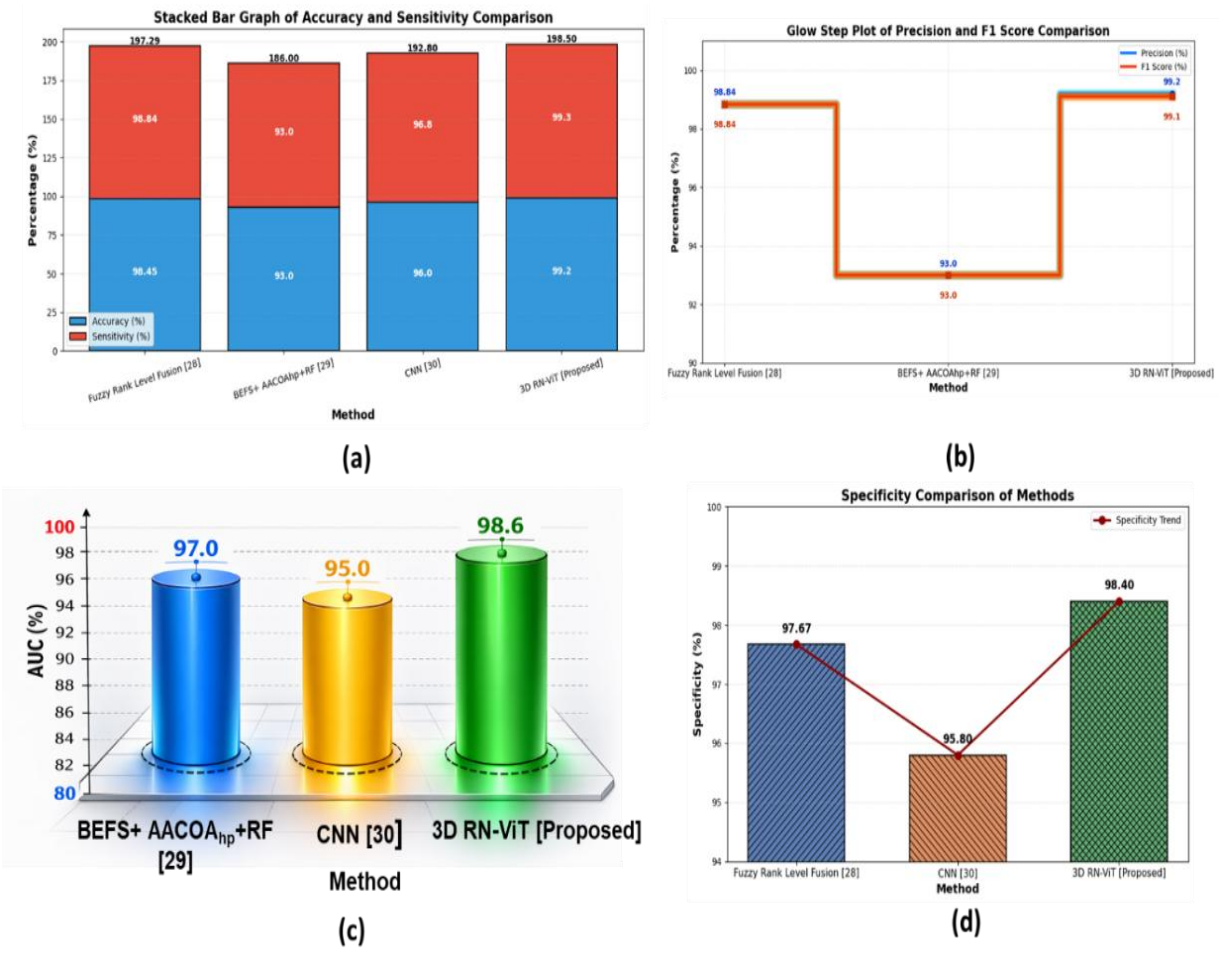

Figure 9. 3D RN-ViT performance in all metrics of evaluation **(a)** Accuracy and Sensitivity Comparison **(b)** Precision and F1 Score Comparison **(c)** AUC Comparison **(d)** Specificity Comparison.

Results show that the proposed architecture outperforms existing 3D designs by employing three-dimensional residual neural networks with transformer (3D RN-ViT) optimized on the Satin Bowerbird Optimization (SBO) algorithm associated with explainable AI (XAI) processing techniques to provide exceptional diagnostic performance to retain clinical interpretability.

A classification accuracy of (>99.0%), precision of (>98.9%) and (>98.0%) sensitivity confirms that the conjoining of complementary morphologic, functional and microstructural modalities provides an extremely discriminative representation of PD associated neuroanatomical alterations. The use of the 3D Sobel filtering in combination with modality specific pre-processing processes such as Z-score normalization for T1-weighted data, wavelet based denoising of RS-fMRI signals and eddy current correction of DTI acquisitions, presumably lead to improvements in signal-to-noise ratio and edge preservation, which can benefit robust feature extraction across the different signal modalities. The combination of 3D residual net and vision transformer utilizes locally acting convolutional receptive fields as well as long range attentional self-attention capabilities

(especially relevant to the PD pathology characterized by distributed brain network structures that involve the cortical, subcortical, white matters and not limited to an isolated area).

The comparative analysis raises a number of limitations on current methods of diagnostic of the PD. Although Fuzzy Rank Level Fusion [28] is also beneficial in terms of using multimodal data to form a single assessment, it heavily relies on present and manually developed fusion rules. This inflexible structure limits its flexibility and its capability to capture the nonlinear interactions that tend to be quite intricate between multimodal neuroimaging data and clinical variables. On the same note, approaches like BEFS and RF [29], draw a lot of attention to features selection and optimization using biomarker features. As much as these methods can be used to detect some meaningful predictive features, they do not make use of deep hierarchical representation learning and thus missing rich spatial-temporal correlations that exist in high-dimensional brain imaging data. Traditional CNN-based models [30] help to resolve some of these issues by efficiently learning local spatial structures, but they are weak in terms of long-range global correlations across volumetric images of the brain that is important in the interpretation of complex neural structures related to PD.

Despite these merits there are a number of considerations which temper interpretation of the reported performance. First, the extraordinarily high accuracy and sensitivity prompt the prospect of over-fitting on the Parkinson's Progression Markers Initiative (PPMI) cohort (despite proper pre-processing & feature selection). The PPMI dataset is a well-curated research dataset which is characterised by standardised acquisition protocols, which may not necessarily reflect the heterogeneity of different scanners, protocols and patient populations that one might encounter in routine clinical practice. While the framework supposed to guide early diagnosis, most of the evaluation in the literature addresses the diagnostic classification, with comprehensive analyses of prodromal cases, unique motor/non-motor sub-types and longitudinal disease progression providing further substantiation of the claims on disease stratification and prognosis.

Another aspect that is important is the practical feasibility of multimodal imaging on a routine basis. Simultaneous acquisition of T1-weighted MRI, resting-state fMRI, and DTI is resource-intensive and may be inaccessible in many a clinic environment, especially in low resource settings. While the modality-wise contribution analysis does provide an understanding of the relative informativeness of each sequence, in the future, there is a need to explore the reduced-modality or adaptive imaging strategies, e.g. MRI + DTI only or MRI + clinical features, in a methodical and detailed way to ensure high performance with less overhead on acquisition. Additionally, the computational burden of 3D RN-ВиT and optimisation based feature selection may pose problems in integrating them in time-critical workflows unless required hardware and pipeline optimisation support is available.

The emphasis on explainability also paved the way for further research in several ways. XAI-derived saliency and attribution maps can further be comparatively and rigorously related with known PD-related structural and functional alterations and therefore derive hypotheses regarding novel biomarkers and disease mechanisms. The proposed 3D RN-ViT framework integrates the advantages of 3D residual networks and transformer-based global attention mechanisms. While the residual modules are highly effective at extracting hierarchical spatial features, the vision transformer derives global contextual relations across the entire brain volume. This hybrid model enables better multimodal integration, better interpretability and better differentiation of PD patients versus controls, which indicates a promising development towards automated diagnosis of PD.

6. Conclusion

This research provides an explainable DL model (3D RN-ViT) that serves as a highly accurate proof-of-concept for detecting PD using multimodal neuroimaging and clinical data. The integration of 3D RN and ViT represents an advanced multimodal deep learning architecture that successfully developed and validated an advanced, explainable multimodal deep learning framework to tackle the profound diagnostic challenges associated with the early detection of PD. The intrinsic clinical heterogeneity and lack of any known systemic markers in the prodromal phase of PD support the need for a paradigm shift from symptom-based assessments to data-driven computational tools. The proposed architecture incorporates synergistically all the structural Magnetic Resonance Imaging (MRI), resting-state functional MRI (rs-fMRI), and Diffusion Tensor Imaging (DTI) data collected by the Parkinson's Progression Markers Initiative (PPMI), which collectively capture the morphological, functional, and microstructural changes pathognomonic of the neurodegenerative disorder. A careful preprocessing pipeline was adopted to extract strong and uniform features, which included Z-score normalization for structural MRI, wavelet denoising to remove high-frequency noise in rs-fMRI, and eddy current correction to correct for scanner distortion in DTI, along with 3D Sobel filtering to highlight key neuroanatomical edges. In particular, the 3D RN acts as a spatial feature extractor, deep volumetric convolutions capturing highly localized microstructural anomalies, and the subsequent ViT, with multi-head self-attention mechanisms, breaking away from localized processing, continuously assessing global contextual relationships across the entire brain volume for distributed network breakdown. To address the computational challenges

associated with high dimensional neuroimaging data, including redundancy and algorithmic overfitting, the Satin Bowerbird Optimization (SBO) algorithm was ingeniously used to select the optimal features dynamically and at the same time fine tune the architectural hyperparameters. Importantly, the framework naturally included Explainable Artificial Intelligence (XAI) methods to give absolute transparency to the clinical use of the framework by anatomically localizing the specific neuro-biomarkers driving the diagnostic predictions. Empirical evaluations showed that the proposed 3D RN-ViT model exhibited unprecedented diagnostic efficacy with an exceptional accuracy of 99.2%, a precision of 99.2%, a sensitivity of 99.34%, a specificity of 98.4%, an exceptional F1-score of 99.1%, and an Area under the Curve of 98.6%. These are much better than the traditional machine learning baselines and modern deep learning architectures, thus substantiating the hypothesis of combining local spatial hierarchy with global attention mechanism to achieve better discriminative performance. Overall, this explainable and scalable computational pipeline is a highly robust PoC for AI-assisted neuroimaging systems with a strong potential to impact early PD diagnosis, objective disease stratification and accelerate discovery of novel multimodal biomarkers, provided rigorous external cross-cohort validation.

6.1 Limitations and Future Scope

It has limitations such as dependence on one data set and too much computing requirements. Multimodal integration, cross-cohort validation, and longitudinal prediction studies further developed in future research will develop a very robust, interpretable, and relevant clinically useful model of precision PD diagnosis and biomarker discovery.

References

[1] M. Alalayah, E.M. Senan, H.F. Atlam, I.A. Ahmed, H.S.A. Shatnawi, Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals using Classification Algorithms based on recursive feature Elimination Method. Diagnostics, 13(11), (2023) 1924. https://doi.org/10.3390/diagnostics13111924

[2] E.H. Kwon, S. Tennagels, R. Gold, K. Gerwert, L. Beyer, L. Tönges, Update on CSF Biomarkers in Parkinson's Disease. Biomolecules, 12(2), (2022) 329. https://doi.org/10.3390/biom12020329

[3] M. Alrawis, S. Al-Ahmadi, F. Mohammad, Bridging Modalities: A Multimodal Machine Learning Approach for Parkinson's disease Diagnosis using EEG and MRI Data. Applied Sciences, 14(9), (2024) 3883. https://doi.org/10.3390/app14093883

[4] M. Bongianni, M. Catalan, D. Perra, E. Fontana, F. Janes, C. Bertolotti, L. Sacchetto, S. Capaldi, M. Tagliapietra, P. Polverino, V. Tommasini, Olfactory Swab Sampling Optimization for α-Synuclein Aggregate Detection in Patients with Parkinson's Disease. Translational Neurodegeneration, 11(1), (2022) 37. https://doi.org/10.1186/s40035-022-00311-3

[5] A.E. Lang, R.A. Hauser, L.V. Kalia, B. Hersh, Z. Berger, R. Llorens Arenas, C. Paisan-Ruiz, K. Fraser, D. Jennings, J.H. Kluss, S. Huntwork-Rodriguez, A.G. Henry, & J.T. Greenamyre, "LRRK2 as a potential disease-modifying target in sporadic Parkinson's disease" Movement Disorders, 41(2) (2026) 297-314. https://doi.org/10.1002/mds.70100

[6] S. Prange, H. Klinger, C. Laurencin, T. Danaila, S. Thobois, Depression in Patients with Parkinson's Disease: Current Understanding of its Neurobiology and Implications for Treatment. Drugs & Aging, 39(6), (2022) 417–439. https://doi.org/10.1007/s40266-022-00942-1

[7] J. Rutledge, B. Lehallier, P. Zarifkar, P.M. Losada, M. Shahid-Besanti, D. Western, P. Gorijala, S. Ryman, M. Yutsis, G.K. Deutsch, E. Mormino, Comprehensive Proteomics of CSF, Plasma, and Urine Identify DDC and Other Biomarkers of Early Parkinson's Disease. Acta Neuropathologica, 147(1), (2024) 52. https://doi.org/10.1007/s00401-024-02706-0

[8] M. Alissa, M.A. Lones, J. Cosgrove, J.E. Alty, S. Jamieson, S.L. Smith, M. Vallejo, Parkinson's Disease Diagnosis using Convolutional Neural Networks and Figure-Copying Tasks. Neural Computing and Applications, 34(2), (2022) 1433–1453. https://doi.org/10.1007/s00521-021-06469-7

[9] S.M. Abdullah, T. Abbas, M.H. Bashir, I.A. Khaja, M. Ahmad, N.F. Soliman, W. El-Shafai, Deep Transfer Learning based Parkinson's Disease Detection using Optimized Feature Selection. IEEE Access, IEEE, 11, (2023) 3511–3524. https://doi.org/10.3934/Neuroscience.2023017

[10] K. Chen, H. Wang, I. Ilyas, A. Mahmood, L. Hou, Microglia and Astrocytes Dysfunction and Key Neuroinflammation-Based Biomarkers in Parkinson's Disease. Brain Sciences, 13(4), (2023) 634. https://doi.org/10.3390/brainsci13040634

[11] B. Majhi, A. Kashyap, S.S. Mohanty, S. Dash, S. Mallik, A. Li, Z. Zhao, An Improved Method for Diagnosis of Parkinson's Disease using Deep Learning Models enhanced with Metaheuristic Algorithm. BMC Medical Imaging, 24(1), (2024)

156. https://doi.org/10.1186/s12880-024-01335-z

[12] L. Nechushtai, D. Frenkel, R. Pinkas-Kramarski, Autophagy in Parkinson's disease. Biomolecules, 13(10), (2023) 1435. https://doi.org/10.3390/biom13101435

[13] S. Saravanan, K. Ramkumar, K. Narasimhan, S. Vairavasundaram, K. Kotecha, A. Abraham, Explainable Artificial Intelligence (EXAI) Models for Early Prediction of Parkinson's Disease based on Spiral and Wave Drawings. IEEE Access, IEEE, 11, (2023) 68366–68378. https://doi.org/10.1109/ACCESS.2023.3291406

[14] M. Camacho, M. Wilms, P. Mouches, H. Almgren, R. Souza, R. Camicioli, Z. Ismail, O. Monchi, N.D. Forkert, Explainable Classification of Parkinson's Disease using Deep Learning Trained on a Large Multi-Center Database of T1-Weighted MRI Datasets. NeuroImage: Clinical, 38, (2023) 103405. https://doi.org/10.1016/j.nicl.2023.103405

[15] G. Costantini, V. Cesarini, P. Di Leo, F. Amato, A. Suppa, F. Asci, A. Pisani, A. Calculli, G. Saggio, Artificial Intelligence-based Voice Assessment of Patients with Parkinson's disease off and on Treatment: Machine vs. Deep-Learning Comparison. Sensors, 23(4), (2023) 2293. https://doi.org/10.3390/s23042293

[16] A. Rana, A. Dumka, R. Singh, M. Rashid, N. Ahmad, M.K. Panda, An Efficient Machine Learning Approach for Diagnosing Parkinson's Disease by Utilizing Voice features. Electronics, 11(22), (2022) 3782. https://doi.org/10.3390/electronics11223782

[17] M. Aljalal, S.A. Aldosari, K. AlSharabi, A.M. Abdurraqeeb, F.A. Alturki, Parkinson's Disease Detection from Resting-State EEG Signals using Common Spatial Pattern, Entropy, and Machine Learning Techniques. Diagnostics, 12(5), (2022) 1033. https://doi.org/10.3390/diagnostics12051033

[18] A. Rehman, T. Saba, M. Mujahid, F.S. Alamri, N. ElHakim, Parkinson's disease detection using hybrid LSTM-GRU deep learning model. Electronics, 12(13), (2023) 2856. https://doi.org/10.3390/electronics12132856

[19] T. Aşuroğlu, H. Oğul, A Deep Learning Approach for Parkinson's Disease Severity Assessment. Health and Technology, 12(5), (2022) 943–953. https://doi.org/10.1007/s12553-022-00698-z

[20] R.R. Ileșan, C.G. Cordoș, L.I. Mihăilă, R. Fleșar, A.S. Popescu, L. Perju-Dumbravă, P. Faragó, Proof of Concept in Artificial-Intelligence-based Wearable Gait Monitoring for Parkinson's Disease Management Optimization. Biosensors, 12(4), (2022) 189. https://doi.org/10.3390/bios12040189

[21] D. Trabassi, M. Serrao, T. Varrecchia, A. Ranavolo, G. Coppola, R. De Icco, C. Tassorelli, S.F. Castiglia, Machine Learning Approach to Support the Detection of Parkinson's Disease in IMU-based Gait Analysis. Sensors, 22(10), (2022) 3700. https://doi.org/10.3390/s22103700

[22] M. Hosseinzadeh, A. Gorji, A. Fathi Jouzdani, S.M. Rezaeijo, A. Rahmim, M.R. Salmanpour, Prediction of Cognitive Decline in Parkinson's Disease using Clinical and DAT SPECT Imaging Features, and Hybrid Machine Learning Systems. Diagnostics, 13(10), (2023) 1691. https://doi.org/10.3390/diagnostics13101691

[23] R. Pratihar, R. Sankar, Integrative Federated Learning Framework for Multimodal Parkinson's Disease Biomarker Fusion. Computers, 14(9), (2025) 388. https://doi.org/10.3390/computers14090388

[24] Y. Chang, J. Liu, S. Sun, T. Chen, R. Wang, Deep Learning for Parkinson's Disease Classification using Multimodal and Multi-Sequences PET/MR Images. EJNMMI Research, 15(1), (2025) 55. https://doi.org/10.1186/s13550-025-01245-3

[25] T. Meindl, A. Hapfelmeier, T. Mantel, A. Jochim, J. Deppe, S. Zwirner, J.S. Kirschke, Y. Li, B. Haslinger, Assisted Parkinsonism Diagnosis Using Multimodal MRI—The Role of Clinical Insights. Brain and Behavior, 15(1), (2025) e70274. https://doi.org/10.1002/brb3.70274

[26] H. Pang, Z. Yu, H. Yu, M. Chang, J. Cao, Y. Li, M. Guo, Y. Liu, K. Cao, G. Fan, Multimodal Striatal Neuromarkers in Distinguishing Parkinsonian Variant of Multiple System Atrophy from Idiopathic Parkinson's Disease. CNS Neuroscience & Therapeutics, 28(12), (2022) 2172–2182. https://doi.org/10.1111/cns.13959

[27] M. Khedimi, T. Zhang, C. Dehmani, X. Zhao, Y. Geng, A Unified Deep Learning Ensemble Framework for Voice-Based Parkinson's Disease Detection and Motor Severity Prediction. Bioengineering, 12(7), (2025) 699. https://doi.org/10.3390/bioengineering12070699

[28] A. Kurmi, S. Biswas, S. Sen, A. Sinitca, D. Kaplun, R. Sarkar, An Ensemble of CNN Models for Parkinson's Disease Detection using DaTscan Images. Diagnostics, 12(5), (2022) 1173. https://doi.org/10.3390/diagnostics12051173

[29] A.O. Esan, D.B. Olawade, A.A. Soladoye, B.A. Omodunbi, I.A. Adeyanju, N. Aderinto, Explainable AI for Parkinson's Disease

Prediction: a Machine Learning Approach with Interpretable Models. Current Research in Translational Medicine, 73(4), (2025) 103541. https://doi.org/10.1016/j.retram.2025.103541

[30] S.S. Hussain, X. Degang, P.M. Shah, S.U. Islam, M. Alam, I.A. Khan, F.A. Awwad, E.A. Ismail, Classification of Parkinson's Disease in Patch-based MRI of Substantia Nigra. Diagnostics, 13(17), (2023) 2827. https://doi.org/10.3390/diagnostics13172827

Authors Contribution Statement

Bharath Kumar Nagaraj: Conceptualization, Software. R. Kishore Kanna: Methodology, Supervision, Validation, Writing - Original Draft, Writing - Review & Editing. Jacintha Jayavani Jayakaran: Validation. S. Vijayaraj: Formal Analysis. M.K. Soundarya: Visualization. N. Aravindha Babu: Supervision, Validation. All the authors read and approved the final version this manuscript.

Acknowledgement

The authors gratefully acknowledge the financial support received from Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, under the Research Development Fund (RDF), Grant No. VTU/RDF/FY2026-27/009. This support was instrumental in facilitating the successful execution of this research work.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The multimodal neuroimaging and clinical data used in this study were obtained from the Parkinson's Progression Markers Initiative (PPMI) database. The dataset is publicly available to qualified researchers who comply with the data usage agreement at www.ppmi-info.org.

Has this article screened for similarity?

Yes

About the License