Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Integrating Clinical Indicators and DaTSCAN SPECT Biomarkers Using Transfer Learning Ensemble Models for Early Parkinson's Disease Diagnosis with SWEDD Cohort

Bakaraniya Parul V.[1*], Mamta C. Padole[2]

[1]Research Scholar,[2]Associate Professor, Department of Computer Science and Engineering, The M. S. University, Vadodara, India.

E-mail: [1*]parul9980@gmail.com, [2]mpadole29@rediffmail.com

Abstract

One of the most crucial issues nowadays is the initial detection of Parkinson's disease (PD) to improve patient treatment and diagnosis. The objective of this research is to classify PD patients at an early stage among the healthy controls (HC) and the Scan without Evidence of Dopaminergic Deficit (SWEDD) group, which has a combination of symptoms from the PD and HC control groups. Accurate classification is necessary and challenging to identify Parkinson's disease in these hybrid cases. To address the aforementioned issues, the current study focuses on a novel deep neural network (DNN) architecture that can distinguish between participants in SWEDD, healthy controls, and people with PD. The data consists of clinical information and DaTSCAN single-photon emission computed tomography (SPECT) scans from 589 participants, including 62 SWEDD subjects, 135 healthy controls, and 392 PD patients. The original PPMI datasets were used. The proposed framework incorporates DNN and numerous transfer learning models, including ResNet50, DenseNet121, Xception, ResNet152, InceptionV3, VGG16, and EfficientNetV2B0.We have used Random Forest (RF) and Support Vector Machine (SVM) to create ensemble models that are comparable to all of the previously mentioned transfer learning models to improve optimization. In comparison to the estimations from hybrid model practices and conventional clustering algorithms like DBSCAN and K-Means, DNN produced distinguishing scores of 91.8% accuracy, ResNet50 with SVM produced 0.94 accuracy, ResNet152 with SVM produced 0.97 accuracy, and Xception with SVM produced 0.99 accuracy. The Xception with SVM model also yielded better results, with F1 scores of 0.98 for class 0 (HC), 1.00 for class 1 (PD), and 0.93 for class 2 (SWEDD) subjects. These results show that the proposed deep learning and transfer learning ensemble model configuration is valid for identifying PD cases alongside healthy ones in the SWEDD population.

Keywords: Transfer Learning, Parkinson Disease Progression, SWEDD, Random Forest, Ensemble Model.

*** Corresponding Author**

DOI: https://doi.org/10.36548/jtcsst.2026.2.010
Received: 12.03.2026, received in revised form: 16.04.2026, accepted: 04.05.2026, published: 18.05.2026

© 2026 Inventive Research Organization. This is an open access article under the Creative Commons Attribution-NonCommercial International (CC BY-NC 4.0) License

1. Introduction

A persistent, resistant neurological disorder that gradually reduces cognitive function is Parkinson's disease. In conjunction with an array of non-motor manifestations such as sleep irregularities, emotional upheavals, and a waning sense of smell, this state is marked by a variety of motor challenges, muscular stiffness, featuring tremors, bradykinesia (a slowdown in movement pace), and postural imbalance[1,2].The gradual but advanced loss of dopamine-producing cells surrounded by the deep division of the substantia nigra, a vital part of the brain controlling motor activities, is largely related to the main pathophysiological events on which this condition is based. As a consequence of this neuronal degeneration, the levels of dopamine within the striatum and the operational efficacy of the dopamine transporter are significantly reduced [3,4]. There are also significant changes in the structural integrity of the striatum as the dopamine deprivation progresses [4]. Crucially, a significant number of these neurons are destroyed before clinical symptoms are used to diagnose the illness, demonstrating that the onset of the disease occurs long before clinical symptoms are observed. As the underlying reasons for the degeneration process of the affected neurons are unknown, there is currently no proven treatment or method for rejuvenating the affected neurons. However, pharmaceutical treatments can be effective in controlling the symptoms and allowing patients to live a relatively normal life without accelerating the deterioration of neurons [5]. Due to such vital considerations, timely and accurate diagnosis is important, especially to start neuroprotective therapies that may potentially prevent the progression of the ailment. However, the early recognition of clinical manifestations is a considerable obstacle, since the first symptoms are very subtle and consequently, the diagnosis is delayed until more pronounced signs develop. Through this approach, the first identification of PD has been significantly improved employing neuroimaging techniques, in particular, SPECT. Dopaminergic deficits can be detected even in those who are undergoing the earliest stages of PD with the use of the radioligand, 123I-Ioflupane (DaTSCAN), which has a high affinity for dopamine transporters. Consequently, this imaging technique has developed to become an important part of the diagnostic procedure [6]. Nevertheless, there are patients with clinical diagnosis of PD who have normal DaTSCAN scans. Such cases are classified under SWEDD [7,8]. Although some patients with SWEDD will later display changes related to PD, others will not and can be considered healthy controls (HC), as seen in longitudinal studies [7,8]. Since patients with early PD require specific treatment strategies, which differ from those required in other patients, sometimes the boundaries between SWEDD, early PD, and healthy controls become less clear. Semi-quantitative methods have been used in clinical settings to enhance diagnostic precision. To facilitate visual interpretation of SPECT images, these methods calculate Specific Binding Ratios (SBRs), which provide objective indices of dopaminergic activity [9] SBRs are useful but are incapable of identifying significant morphological features such as the topography of striatal uptake. This limitation can lead to misclassification, particularly in patients with mild initial PD. Thus, exclusive use of semi-quantification is not sufficient for comprehensive assessment. Computer-aided diagnosis and detection (CADD) systems have been designed to address these limitations by employing machine learning (ML) algorithms to assess multimodal data, including imaging and clinical indices. CADD systems have been well received in medical diagnostics because they can provide reliable predictions and ensure the accuracy of their predictions, making them more accurate than conventional methods [10-14]. They have shown promise in identifying PD and providing a more accurate initial diagnosis of the disease [15–18]. In order to create an automated and effective method for primary PD identification, many machine learning models have been tested, and many researchers have shown promising results [16–19]. The document is divided into several sections. Section 2

primarily focuses on a comprehensive examination of existing scholarly works. Section 3 delineates the methodological framework. Section 4 addresses the findings and performance metrics, culminating in the research conclusions shown in Section 5.

2. Literature Survey

Numerous empirical investigations have specifically amalgamated Support Vector Machine algorithms with datasets derived from the Parkinson's Progression Markers Initiative, with the objective of refining automated diagnostic processes related to PD. The application of ML methodologies has demonstrated superior efficacy in contrast to conventional diagnostic techniques, such as ocular examination and semi-quantitative imaging analysis. For example, the study conducted by Khachnaoui et al. published in the Journal of Imaging [20] illustrates the difficulties involved in the initial identification of PD and its differentiation from a group of people known as SWEDD, who fulfil the standards for PD but have usual dopamine transporter imaging. This group has common imaging features and symptoms with both PD patients and healthy controls, making early diagnosis more difficult. They used K-means, DBSCAN, and Hierarchical Clustering methods based on three ML clustering algorithms to classify the aforementioned entities as PD or HC in the SWEDD group to verify this. In terms of sensitivity, accuracy, and especially, Hierarchical Clustering performed better than DBSCAN and K-means by 78.13%, 64%, and 38.89%, respectively, with the highest precision. Specifically in relation to the SWEDD group which has diagnostically uncertain data, these findings indicate the viability of ML algorithms among PD and HC individuals. To automatically diagnose PD using DaTSCAN SPECT scans, Kurmi et al. [22] created a complex ensemble model made up of four convolutional neural networks: VGG16, ResNet50, InceptionV3, and Xception. The entire system was capable of harnessing the distinctive advantages inherent in each individual model through the amalgamation of the previously mentioned models utilizing an ensemble methodology grounded in fuzzy logic. Using the publicly available PPMI dataset for training, the suggested model showed good diagnostic performance metrics, such as an accuracy rate of 98.45% and comparatively high values for sensitivity, precision, specificity, and F1-score, all above 97%. Similarly, Prashanth 2025 [23] used DaTSCAN images from a PPMI dataset to develop a CNN-model framework for the assessment of PD in its primary stages. The technique included preprocessing steps such as image alignment and Gaussian smoothing and was centered on selecting slices according to their diagnostic significance. The proposed CNN model scored over 99% on all performance metrics, outperforming more conventional classifiers like SVM and logistic regression in terms of distinguishing early Parkinson's patients from healthy individuals. Metaheuristic optimization has been integrated into the deep learning process by Majhi et al. (2024) [24]. They applied Grey Wolf Optimization (GWO), hyperparameter optimization, and CNN models (VGG16, DenseNet, InceptionV3). With 100% classification accuracy and 99.92% AUC on DaTSCAN images, the hybrid model showed outstanding performance. It even worked almost perfectly when MRI data was collected. Hazhirkarzar et al. (2024) [25] applied a hybrid multimodal method (namely, MRI and DAT-SPECT imaging, along with radiomic characteristics of regions of the brain that are commonly affected in PD, such as the caudate, putamen, and midbrain regions) to predict the development of the ailment. To forecast disease progression, they applied an ensemble model that integrated decision trees with adaptive boosting methods. Although the model's poor performance on external datasets indicates a need for further research on generalization, it demonstrated 75% of accuracy and an AUC of 0.90 in internal validation. Recently, there has been a significant increase in the use of deep learning architectures for diagnosing Parkinson's disease (PD), with research increasingly transitioning

from single-modality to multimodal and explainability focused methodologies. As noted in [25], Dentamaro et al. examined multimodal DL for the detection of prodromal-stage PD utilizing the PPMI database, integrating 3D architectures with an innovative Excitation Network and Explainable Artificial Intelligence (XAI) methodologies. Their joint co-learning approach facilitated end-to-end training that utilized complementary evidence from imaging and clinical modalities; DenseNet enhanced with clinical data surpassed other models. Chang et al. [26] proposed a revised ResNet18 framework trained on integrated FDG-PET and CFT-PET/MR multimodal sequences to distinguish PD from MSA and healthy controls, employing a four-fold cross-validation method on 206 clinically diagnosed patients. The rationale for integrating these complementary modalities is based on CFT-PET's direct evaluation of dopaminergic integrity and FDG-PET's indication of the subsequent metabolic effects of neurodegeneration. RACF by She et al. [27] is a multimodal DL model that provides a method for integrating single-nucleotide polymorphism (SNP) data and structural MRI to diagnose Parkinson's disease, utilizing dynamic cross-modal fusion through attention and contrastive learning. The model combined both interpretability pathways with SHAP values and Grad-CAM to explain genetic epitasis and neuroimaging biomarkers simultaneously, addressing the long-standing puzzle of disentangling genetic neuroanatomical interactions. Majhi et al. [28] suggested a metaheuristic-optimized deep learning method for detecting PD from MRI and DaTSCAN data, using Grey Wolf Optimization to automatically adjust the hyperparameters of several architectures, such as VGG16, DenseNet, and InceptionV3. The hybrid GWO-DenseNet with LSTM model was one of the best setups tested on standard datasets from PPMI. Barua et al. [29] developed MultiParkNet, a multi-modal deep learning framework that combines audio spectrograms, MRI, DaTSCAN, EEG, and physiological signals for early PD detection. This was done because single-modality approaches have known limitations. The framework used GradCAM and attention surface maps to explain things at the modality level. The authors stated that federated learning and edge deployment are important ways to ensure that clinical scalability is privacy-preserving. As stated in [30], a multimodal framework integrating PET imaging, hand tremor detection through Hough Transform on line-drawing assessments, and XGBoost-based classification of non-motor symptoms was suggested for early PD diagnosis, attaining over 95% accuracy in predicting disease stage. This study showed that it is possible to combine neurobiological imaging with both motor and non-motor clinical markers in a single computational pipeline. Suo et al. [31] created the Swin Classifier, a self-supervised vision foundation model that was trained on 75,861 clinical head MRI scans with T1-weighted, T2-weighted, and FLAIR sequences. It can differentiate between PD and Parkinson-plus syndrome. The model showed strong performance using routine clinical MRI without needing special acquisition protocols. It was pre-trained on a cross-contrast context recovery task and tested on a downstream classification cohort of almost 4,000 people. As noted in [32], to overcome the complementary shortcomings of CNNs and Vision Transformers, a fusion framework of Swin-Transformer and CNN was proposed for PD classification utilizing MRI data. CNNs usually focus on local features, while ViTs use self-attention to obtain global representations. The combined framework with skip connections and cosine attention was able to capture both local and global feature information, which helped with the gradient vanishing problem that is common in standard ViT architectures. Khalil et al. [33] addressed the clinically challenging problem of differentiating PD and other forms of parkinsonism based on wearable sensors and machine learning, pointing out that such diffuse clinical symptoms as tremor, bradykinesia, and rigidity make traditional clinical methods subjective and uncertain.

Their research underscored significant methodological challenges, notably class imbalance and limited non-PD sample sizes that must be addressed for reliable deployment. Kaur [34] and her team looked at ML and DL to help diagnose Parkinson's disease. They examined one hundred and thirty-three studies from the years two thousand twenty-one to two thousand twenty-four. The team categorized the information into groups such as movement, acoustic, medical imaging (like MRI and DaTSCAN), biomarkers, and multimodal types. They found that movement and voice data are used extensively. Using various types of data together can enhance the accuracy of results. The team tried different models, including SVM, Random Forest, and k-Nearest Neighbours. They also experimented with DL models, such as CNNs and Recurrent Neural Networks. Combining both ML and DL can yield even better results. Most studies utilize a method called cross-validation to evaluate effectiveness. They also consider metrics like accuracy, precision, recall, and F1-score, along with ROC-AUC to assess performance. The results they obtained are generally quite good, with accuracy typically ranging from 85% to 99%, and F1-scores often above 0.80. However, there are still some challenges with machine learning models. One issue is that we lack sufficient data to work with; some data is more critical than others. The machine learning models tend to perform well only on the data they are trained on, and they do not generalize well to new data. Machine learning models struggle with new data. The authors believe that researchers should collaborate across data types and ensure their tests are robust so that doctors can trust ML and DL models to detect Parkinson's disease early. The researchers aim to strengthen the machine learning models so that doctors can utilize them for detecting Parkinson's disease. Ling et al. (2025) [35] proposed a computer model called a GNN framework that uses a Regional Radiomics Similarity Network with Transformer-based attention mechanisms to help doctors differentiate between parkinsonian syndromes. They did this because many individuals diagnosed with Parkinson's disease are later found to have a different condition. The study involved a group of 1,495 participants: 220 people and 1,275 individuals with Parkinson's disease, MSA, or PSP. All of these participants underwent special ^{18}F-FDG-PET imaging tests. The ML and DL models are crucial for detecting Parkinson's disease, and the researchers want to ensure they are functioning effectively. Brain regions were modeled as graph nodes, with edge weights derived from radiomic similarity matrices. The GNNExplainer assisted in understanding the decisions by identifying subgraph structures that were significant for diagnosis. The framework performed well, achieving F1-scores of 96.3% for IPD, 92.5% for MSA, and 86.7% for PSP. It outperformed SVM and GCN baselines by 5.5% to 8.3%. The attention mechanism made it 41% less computationally intensive. It also exhibited disease signs, such as putaminal hypometabolism in IPD and midbrain-prefrontal disconnection in PSP, which are linked to the conditions. The results are based on IPD, MSA, and PSP.

3. Proposed Model

3.1 Contributions

Motivated by contemporary developments in the utilization of ML methodologies to distinguish between individuals diagnosed with PD and healthy control participants, this study seeks to address a more complex objective of differentiating between PD patients and HCs within the diagnostically ambiguous category of SWEDD in the Dopaminergic Deficit cohort. The traditional supervised learning approach may sometimes result in ambiguous results because the clinical presentation and imaging characteristics of PD, SWEDD, and HC groups are quite similar. To overcome this limitation, this paper proposes a hybrid classification

approach that seamlessly combines the basic concepts of machine learning and transfer learning approaches, such as ResNet50, DenseNet121, Xception, Resnet152, InceptionV3, VGG16, and EfficientNetV2B0, for enhanced early PD diagnosis in the SWEDD group. Additionally, we have integrated traditional machine learning algorithms including all transfer learning models, with RF and SVM. When the evidence is weak but crucial, the model acts as the clinician's thinking companion by providing a stratified signal to help make decisions.

Key Contributions and Research Goals:

1. New Algorithmic Design for Diagnostic Support in a SWEDD compartment: We propose an architectural implant consisting of different learning paradigms that have been cross-validated to produce pathognomonic symptoms that differentiate the emergence of Parkinson's disease. The SWEDD entity is especially pertinent in the current scenario; the long-term observation is the justification for the clinical urgency, demonstrating a partition progressing to typical PD within the typical neuro-radiological delay windows.

2. Using Wide Data for Robust Generalization: The widely accessible Parkinson's Progression Markers Ingenuity database, which contains a globally diverse population, forms the basis for training, testing, and model development. In our work, we have categorized 589 subjects into 62 identified SWEDD, 135 identified healthy controls, and 392 labeled PD. The input features are nine clinical and imaging variables per subject, which further divide the autonomic, motor, and cognitive functions into the three specified groups. The variables consist of five clinical variables and four DaTSCAN SPECT imaging variables.

3. The application of a three-stage class hierarchy. To classify samples into the three target classes of PD, HC, and SWEDD, the first layer of phase one integrated all clinical and imaging variables into a hybrid network consisting of convolutional and fully linked layers, resulting in a final output softmax. The secondary layer established criteria to differentiate between individuals diagnosed with PD and those deemed healthy by using traditional transfer learning model classifiers that utilize clinical variables alongside DaTSCAN SPECT imaging metrics. In the third layer, we developed an ensemble model that offers the best possible outcome for all three classes by combining transfer learning models with conventional ML algorithms such as RF and SVM.

3.2 Dataset Overview

The PPMI [35], a collaborative research effort that seeks to harmonize clinical, imaging, and biomarker data collection for PD research, was the source of the data set used in this study. We found a total of clinical data from 548 HC, 1,955 PD, and 662 SWEDD patients in the complete data set. However, a total of 589 individuals with both clinical and DaTSCAN SPECT imaging information (392 PD, 135 HC, and 62 SWEDD patients) were selected for this study. Initially, we found a total of 662 SWEDD patients with only clinical data, and then we identified 62 SWEDD subjects whose DaTSCAN SPECT images were also available. Nine features—five clinical and four imaging-related—were included in each subject's profile and were used as inputs to the classification models.

As mentioned by researchers in [20]. Four DaTSCAN SPECT Imaging Attributes and five standardized clinical measures were used to measure both the motor and non-motor aspects of PD:

1. The Unified PD Rating Scale, which gauges the degree of motor impairment.
2. Montreal Cognitive Assessment: This test assesses cognitive function.
3. The University of Pennsylvania Smell Identification Test assesses the ability to smell in patients.
4. State-Trait Anxiety Inventory: This tool separates symptoms of anxiety from those of depression.
5. Geriatric Depression Scale: This measure looks for signs of depression in senior citizens.
6. SBR (Right Caudate)
7. Left Caudate (SBR)
8. Right Putamen, or SBR
9. Left Putamen (SBR)

Objective distinction between HC and SWEDD groups is made possible by these imaging findings, which offer quantifiable indicators of dopamine transporter density, which is decreased in PD.

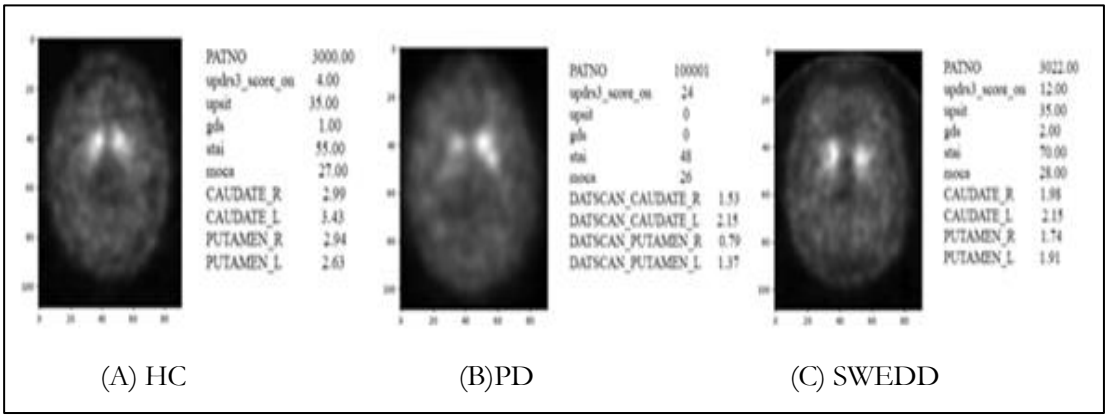

Figure 1. DaTSCAN SPECT Imaging and Clinical Data of the Dopaminergic System for HC, PD and SWEDD Subjects [21]

3.3 Methodology

The next section details the methods the authors employed to discriminate PD patients within the SWEDD group. The first step in the analysis was to extract and interpret clinical information from a number of source documents, including 662 SWEDD patients, 1,955 PD patients, and 548 HC. From this massive amount of information, a group of 589 people was chosen who had clinical information as well as DaTSCAN SPECT scans. The most recent information was split into three groups: 62 SWEDD patients, 135 healthy individuals, and 392 PD patients. Every individual had nine attributes, four of which were from the neuroimaging

information and five of which were from the clinical assessment. The proposed framework initially utilized an 80:20 proportion to partition the dataset into training and testing subsets, which is also denoted as Model 1. By integrating the clinical and SPECT imaging components for the same subjects, the above approach was extended. A tailored DNN developed specifically to classify subjects into PD, HC, or SWEDD groups was then employed to analyze the integrated set of features. The objective of this multi-modal methodology was to augment diagnostic efficacy by leveraging complementary information from imaging biomarkers and clinical assessments. The suggested DNN method's structure diagram is shown in Figure 2.

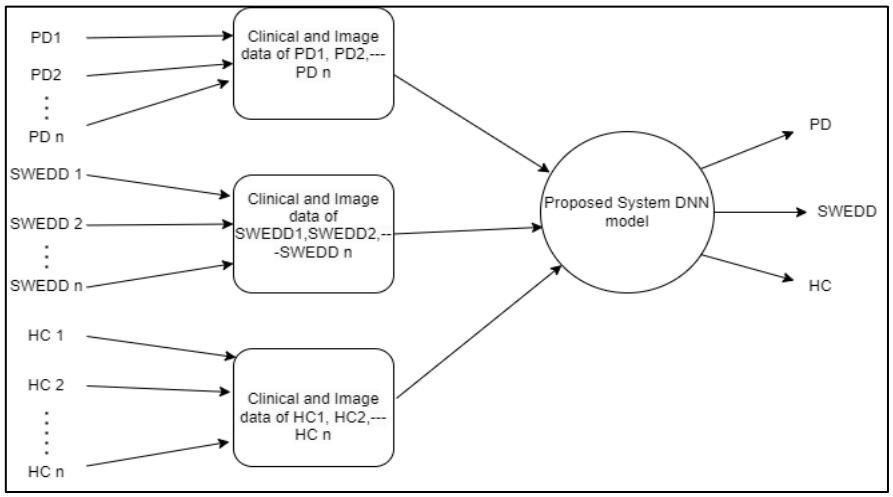

Figure 2. Diagram of the DNN Model (Model 1) with Hybrid Data for Same Patient [36]

Clinical characteristics and DaTSCAN SPECT imaging of the same participants were harmoniously combined in Model 1, and this wealth of information was then fed into our novel deep neural network model to categorize the subjects into three groups: Parkinson's Disease, Healthy Control, and SWEDD. Figure 1 above shows the initial clinical and SPECT imaging data for each of the selected patients. The PPMI is the source of this data [35]. For the purpose of cohort identification based on both neuroimaging information and related clinical variables, a dual-input deep learning model was developed in this study using the TensorFlow Keras API. The model consists of two parallel branches; one for processing organized clinical data and the other for processing grayscale 2D medical images. Tensors representing grayscale image slices, such as those from MRI or DaTSCAN SPECT imaging, of size 91 × 128 × 1 are taken by the image input branch. We obtain the SPECT DICOM images from PPMI [21] with a 91 X 128 image size, so we have kept the same resolution for image input. At the same time, a nine-element vector containing clinical and demographic information is taken by the clinical input branch. The starting step in the image processing pipeline involves rescaling the intensity of each pixel to a normalized scale between 0 and 1. This is followed by two operations: pooling and convolution. The first convolutional layer, activated by the ReLU function, applies 128 filters of size 3 × 3. A max pooling operation with a 2 × 2 window is then performed. The next convolutional layer uses 64 filters, also 3 × 3, and is activated by ReLU, followed by another max pooling strategy. To facilitate integration with the clinical feature stream, the output is transformed into a one-dimensional vector. At the same time, the structured data is processed by the clinical feature branch through a completely connected layer with 64 neurons, activated by ReLU. The outputs of both branches are combined to generate a multimodal feature vector, which is passed through a dense layer with three output units. Softmax activation is applied to classify the input as PD, SWEDD, or healthy control. Optimization is performed using sparse categorical cross-entropy loss in conjuction with the Adam optimizer. Precision is used as the initial evaluation metric. The team trained the model for 10 epochs with a batch size of 16 on

the resampled training data. When tested on an imbalanced dataset, an impressive accuracy of 0.91 was achieved, effectively distinguishing PD, HC, and SWEDD patients. The dataset, comprising 589 SPECT images, was divided into the training set (469 images) and testing set (118 images) with an 80:20 ratio. The training set had the following distribution: HC = 108, PD = 312, and SWEDD = 49. The outcomes are presented in Table 1.

Table 1. Comparision of Existing and Propose Model 1 Performance Parameters

	Existing System [21]			**Proposed DNN model (Model 1)**
Measure	**Kmeans**	**DBSCAN**	**Hierarchical Clustering**	**Hybrid (SPECT with Clinical) Dataset**
Accuracy	61.29	60.98	64.00	91.8
Sensitivity	59.38	76.92	78.13	86.0
Specificity	38.89	33.33	38.89	82.0
F1_Score	61.29	71.43	73.53	82.3

Model 1 gives the highest accuracy of 91.8 % compared to the existing algorithms accuracy of 64%. Then, we have implemented different transfer learnig models on the same hybrid dataset. The system flow shown below in Figure 3 is proposed by us. Here different transfer learning models are used, including ResNet50, DenseNet121, Xception, ResNet152, InceptionV3, VGG16, and EfficientNetV2B0.

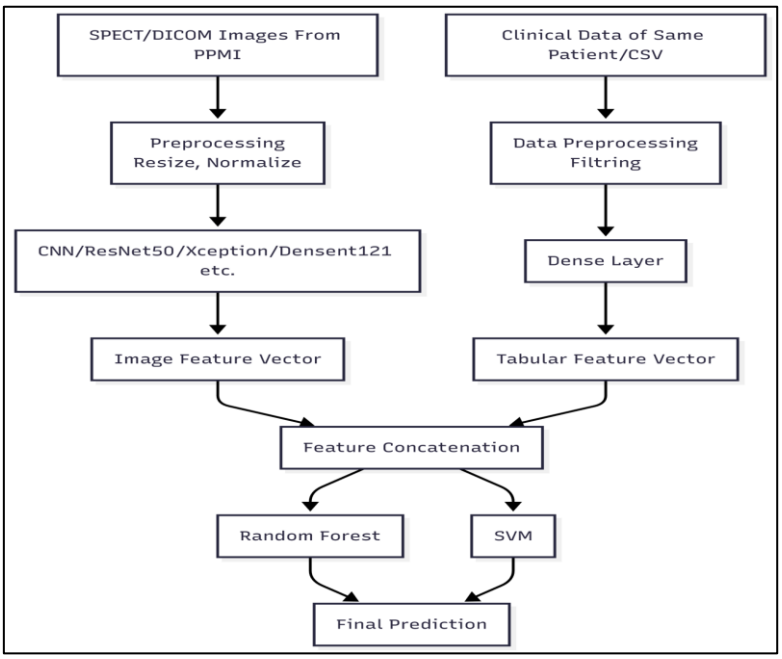

Figure 3. Proposed Framwork for Hybrid (SPECT with Clinical) Dataset

4. Results and Discussion

For accurate disease classification, the proposed approach provides a multimodal learning framework that integrates clinical and neuroimaging information. The model consists of two parallel processing pipelines, one for SPECT/DICOM image analysis and the other for patient-specific clinical information, as depicted in the flow diagram. Feature-level fusion and classification follow. To ensure uniform input representation, the first pipeline uses standard image preprocessing steps, including scaling and intensity normalization, for SPECT/DICOM images retrieved from the PPMI database. Deep CNN models such as CNN, ResNet50,

DenseNet121, Xception, ResNet152, InceptionV3, VGG16, and EfficientNetV2B0 are used as image inputs. These models act as automatic feature extractors, detecting high-level textural and spatial patterns indicative of neurodegenerative changes. A concise image feature vector encapsulating imaging biomarkers is the output of this stage. Simultaneously, tabular (CSV) clinical information for the same subjects is retrieved. To address missing values, noise, and irrelevant attributes, this information is preprocessed and filtered. A tabular feature vector that holds patient-specific clinical information is obtained by passing the refined clinical features through fully connected dense layers to obtain nonlinear representations of the features. The effective feature-level fusion of multiple data modalities is thus enabled by combining the image feature vector and the tabular feature vector using feature concatenation. By integrating the complementary data from the clinical and imaging domains, the combined representation enhances the discriminative capability of the model. To obtain the final forecast, the concatenated feature vector is then passed to two traditional machine learning classifiers: RF and SVM. Strong decision boundaries can thus be learned from the combined feature space using these classifiers. The final diagnostic prediction is obtained by the framework, demonstrating the efficacy of multimodal feature integration for improved classification performance. When seven transfer learning approaches (Model2) are evaluated, there are large variations in classification performance, with overall truth values varying between 0.91 and 0.94. This is depicted in the following Table 2.

Table 2. Performance Comparison of Transfer Learning Models for Multiclass Classification (Model 2)

Sr.No.	Model Name	Class (0-HC,1-PD,2-SWEDD)	Precision	Recall	F1-score	Test Accuracy
1	DNN	0	0.92	0.74	0.82	0.9180
		1	1.00	0.99	0.99	
		2	0.54	0.85	0.66	
2	ResNet50	0	0.77	0.95	0.85	0.93
		1	1.00	0.99	0.99	
		2	0.78	0.54	0.64	
3	DenseNet121	0	0.80	0.76	0.78	0.56
		1	0.93	0.48	0.63	
		2	0.15	0.62	0.24	
4	Xception	0	0.85	0.82	0.84	0.90
		1	0.96	0.98	0.97	
		2	0.38	0.38	0.38	
5	ResNet152	0	0.80	0.95	0.87	0.92
		1	1.00	0.96	0.98	
		2	0.67	0.62	0.64	
6	InceptionV3	0	0.90	0.70	0.79	0.89
		1	0.99	0.97	0.98	
		2	0.52	0.79	0.63	
7	VGG16	0	0.82	1.00	0.90	0.94
		1	0.99	1.00	0.99	
		2	1.00	0.50	0.67	
8	EfficientNetV2B0	0	0.84	0.96	0.90	0.94
		1	0.99	1.00	0.99	
		2	0.89	0.57	0.70	

Table 2 delivers a thorough performance analysis of the different DL models considered for evaluation in the multiclass classification task of neurological disorders, specifically HC – Class 0, PD – Class 1, and SWEDD – Class 2. The models are assessed using F1-score, precision, recall, and overall test accuracy. VGG16 and EfficientNetV2B0 outperformed the other models evaluated in terms of generalization capability, with an overall test accuracy of

94%. With high F1-scores of 0.90 and 0.99, respectively, VGG16 has perfect recall values for both HC and PD, indicating excellent discrimination between PD patients and healthy individuals. EfficientNetV2B0 performed better than most models in SWEDD classification and had a balanced performance for all classes, with good precision and recall values for both HC and PD. Additionally, with test accuracies of 93% and 92%, respectively, ResNet50 and ResNet152 demonstrated competitive performance. With almost perfect precision and recall values for the PD class, both models demonstrated their efficacy in classifying Parkinson's disease. Their performance in the SWEDD class was somewhat worse, nevertheless, suggesting that this subgroup's classification is comparatively more intricate. With test accuracies of 90% and 89%, respectively, the Xception and InceptionV3 models demonstrated consistent performance for both HC and PD classes. Both models, however, performed poorly in recall and precision values for the SWEDD class, suggesting that it is difficult to differentiate SWEDD from PD and HC because of their similar traits. The lowest test accuracy of 56% was mostly due to the low recall and F1-score values for the PD and SWEDD classes in the DenseNet121 model. This indicates that the DenseNet121 model would perform less well in extracting the discriminative features pertinent to this multiclass classification task in the current experimental setup. With a test accuracy of 91.8%, the baseline DNN model performed outstandingly well for the PD class but relatively poorly for the SWEDD class in terms of precision, thereby emphasizing the natural complexity of SWEDD classification. In conclusion, the results clearly indicate that the classification of SWEDD remains complex, even when most deep learning models correctly classify PD and healthy controls with high accuracy. Architectures with strong multimodal feature fusion and hybrid classification model capabilities, such as a balance between accuracy and class performance, include models such as EfficientNetV2B0 and VGG16. Then we applied an ensemble learning approach for better optimization with a hybrid dataset. We derived the feature set utilizing transfer learning models. Features extracted for training data have the shape: (469, 15424). Subsequently, we employed that feature vector within conventional ML algorithms such as RF and SVM. For RF, we tried different combinations of hyperparameters as shown here: No. of Estimators: [100,200,500,1000], Max Depth: [None, 10,20,30], Min Sample Split: [2,5,10], Min Samples Leaf: [1,2,4]. From the various combinations of hyperparameters evaluated, we found that for the best results, the following hyperparameters were considered. Best Parameters for Random Forest: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}. The comparative findings are presented below in Table 3.

Table 3. Comparative Analysis of Ensemble Approach Using RF for Hybrid Dataset (Model 3.a)

Sr. no.	Model Name	Class (0-HC, 1-PD, 2-SWEDD)	Precision	Recall	F1-Score	Test Accuracy
1	DNN	0 1 2	0.92 1.00 0.54	0.74 0.99 0.85	0.82 0.99 0.66	0.9180
2	ResNet50 with RF	0 1 2	1.00 1.00 0.87	1.00 0.98 1.00	1.00 0.99 0.93	0.98
3	DenseNet121 with RF	0 1 2	0.78 0.99 0.70	0.86 1.00 0.54	0.82 0.99 0.61	0.92
4	Xception with RF	0 1 2	0.87 1.00 0.67	0.96 0.99 0.50	0.92 0.99 0.57	0.95
5	ResNet152 with RF	0 1 2	1.00 1.00 0.93	1.00 0.99 1.00	1.00 0.99 0.96	0.99

6	InceptionV3 with RF	0 1 2	0.82 0.95 0.56	0.85 1.00 0.36	0.84 0.97 0.43	0.88
7	VGG16 with RF	0 1 2	1.00 1.00 1.00	1.00 1.00 1.00	1.00 1.00 1.00	1.00
8	EfficientNetV2B0 with RF	0 1 2	0.81 0.95 0.50	0.93 0.97 0.29	0.86 0.96 0.36	0.88

The class-wise and cumulative performance of the hybrid DL models in combination with a RF classifier for multiclass classification of neurological disorders HC – Class 0, PD – Class 1, and SWEDD – Class 2 is presented in Table 3. Precision, F1-score, and recall for each class, as well as overall test accuracy, are used to evaluate the models. The total test accuracy of the baseline DNN model was 91.80%. The model achieved an F1-score of 0.82, a recall of 0.74, and a precision of 0.92 for HC (Class 0). The model performed almost perfectly for PD (Class 1), with recall = 0.99, precision = 1.00, and F1-score = 0.99. However, for the SWEDD class (Class 2), with recall = 0.85, precision = 0.54, and F1-score = 0.66, performance was quite poor, thereby emphasizing the difficulty of distinguishing SWEDD based on solo deep features. With a test accuracy of 98%, the ResNet50 with RF model demonstrated a significant improvement. With recall = 1.00, precision = 1.00, and F1-score = 1.00, the model performed flawlessly for HC. The model also showed perfect performance for PD, with recall = 0.98, precision = 1.00, and F1-score = 0.99. With recall = 1.00, accuracy = 0.87, and F1-score = 0.93, the SWEDD class demonstrated the effectiveness of RF in modeling non-linear decision boundaries in the fused feature space. ResNet152 with RF also achieved 99% accuracy with near-perfect performance for PD (recall = 0.99, precision = 1.00, F1-score = 0.99) and perfect precision and recall for HC (1.00 / 1.00 / 1.00). When combined with RF classifiers, the SWEDD class showed the durability of deeper residual networks, with precision = 0.93, recall = 1.00, and F1-score = 0.96.

The overall accuracy of the DenseNet121 with RF model was 92%. The reported F1-score, precision, and recall for HC were 0.78, 0.86, and 0.82, respectively. While the SWEDD classification performance was unsatisfactory with recall = 0.54, accuracy = 0.70, and F1-score = 0.61, the PD class performed well with recall = 1.00, precision = 0.99, and F1-score = 0.99. The Xception with RF model achieved a 95% test accuracy. It provided HC with an F1-score of 0.92, recall of 0.96, and precision of 0.87. The PD classification performed exceptionally well (recall = 0.99, precision = 1.00, and F1-score = 0.99), whereas the SWEDD class performed rather poorly (recall = 0.50, accuracy = 0.67, F1-score = 0.57). This implies that there is still some misunderstanding regarding the differences between the PD and SWEDD classes. The overall accuracy of the InceptionV3 with RF model was 88%. HC had recall = 0.85, precision = 0.82, and F1-score = 0.84, despite the fact that the PD classification performance was outstanding (precision = 0.95, recall = 1.00, F1-score = 0.97). Nevertheless, SWEDD classification was subpar (precision = 0.56, recall = 0.36, and F1-score = 0.43). The test accuracy of the VGG16 with RF model was 100%. It performed the best all around. Its precision, recall, and F1-score were 1.00 for HC, PD, and SWEDD, respectively. This shows that all three classes have excellent categorization performance. This is an impressive illustration of the separability of the learned features when paired with RF-based ensemble learning. The model achieved 100% accuracy on the test set. This result is likely due to the high separability of the dataset and the effectiveness of the selected features for the small test size dataset. Additionally, consistent performance across validation suggests that the model generalizes well. However, given the unusually high accuracy, further evaluation on larger and

more diverse datasets is recommended using cross validation techniques. The EfficientNetV2B0 with RF model had an 88% test accuracy rate. For HC, its F1-score was 0.86, precision was 0.81, and recall was 0.93. While the SWEDD classification performance was comparatively bad (precision = 0.50, recall = 0.29, F1-score = 0.36), the PD classification performance was outstanding (recall = 0.97, precision = 0.95, F1-score = 0.96). For all models, PD (Class 1) consistently had the highest precision and recall because of the different imaging and clinical markers. On the other hand, since SWEDD (Class 2) has commonalities with both PD and HC, it performed poorly, making it a difficult clinical subgroup. By effectively using feature-level heterogeneity as well as non-linear interactions, the hybridization of the models with Random Forest classifiers significantly improved classification accuracy compared to standalone deep learning models. Although VGG16 with RF enabled perfect classification by combining ensemble decision-making with dense hierarchical feature extraction, deeper models such as ResNet152 benefited from residual learning. In conclusion, the empirical findings elucidate that hybrid deep learning architectures incorporating the baseline models are markedly outperformed by Random Forest classifiers in the context of multiclass classification pertaining to neurological disorders. With 100% accuracy and perfect class-wise performance, VGG16 with RF outperformed the rest, followed by ResNet152 with RF (99%) and ResNet50 with RF (98%). The findings confirm the efficiency of the proposed hybrid model for clinical decision support systems and also support that ensemble decision-making with deep feature representations improves classification accuracy, especially for difficult classes such as SWEDD. Next, we applied an ensemble method by combining SVM with a transfer learning model.For SVM, we tried different combinations of hyperparameters as shown here, such as ‘C’ : [0.1, 1, 10, 100], ‘gamma’: [0.001, 0.01, 0.1, 1], ‘kernel’: [rbf, poly, sigmoid] From the different combinations of hyperparameters in evaluation, we found that for the best results, the following hyperparameters were considered. The best Parameters for SVM are: (C=0.1, gamma=0.001, kernel='poly').

Table 4. Comparative Analysis of Ensemble Approach Using SVM for Hybrid Dataset (Model 3.b)

Sr no.	Model Name	Class (0-HC, 1-PD, 2-SWEDD)	Precision	Recall	F1-Score	Test Accuracy
1	DNN	0 1 2	0.92 1.00 0.54	0.74 0.99 0.85	0.82 0.99 0.66	0.9180
2	ResNet50 with SVM	0 1 2	0.88 1.00 0.78	1.00 0.95 0.77	0.93 0.98 0.74	0.94
3	DenseNet121 with SVM	0 1 2	0.69 0.99 0.71	0.95 0.96 0.38	0.80 0.98 0.50	0.90
4	Xception with SVM	0 1 2	0.97 1.00 1.00	1.00 1.00 0.88	0.98 1.00 0.93	0.99
5	ResNet152 with SVM	0 1 2	0.90 1.00 0.85	0.90 1.00 0.85	0.90 1.00 0.85	0.97
6	InceptionV3 with SVM	0 1 2	0.96 1.00 1.00	1.00 1.00 0.93	0.98 1.00 0.96	0.99
7	VGG16 with SVM	0 1 2	0.92 1.00 0.80	0.89 1.00 0.86	0.91 1.00 0.83	0.96

8	EfficientNetV2B0 with SVM	0 1 2	0.93 0.99 0.85	0.93 1.00 0.79	0.93 0.99 0.81	0.96

Table 4 summarizes the presentation of transfer learning models integrated with a SVM classifier for multiclass classification of neurological disorders: HC – Class 0, PD – Class 1, and SWEDD Class 2. Model assessment is conducted using recall, class-wise precision, and F1-score, along with overall test accuracy. Table 5 below presents a comparative analysis of all models.

Table 5. The Comparative Analysis of all the Models

Sr.no	Model name	Class 0-HC, 1-PD, 2-SWEDD	Precision	Recall	F1_Score	Accuracy
1	DNN	0 1 2	0.92 1.00 0.54	0.74 0.99 0.85	0.82 0.99 0.66	0.9180
2	ResNet50	0 1 2	0.77 1.00 0.78	0.95 0.99 0.54	0.85 0.99 0.64	0.93
3	ResNet50 with RF	0 1 2	1.00 1.00 0.87	1.00 0.98 1.00	1.00 0.99 0.93	0.98
4	ResNet50 with SVM	0 1 2	0.88 1.00 0.78	1.00 0.95 0.77	0.93 0.98 0.74	0.94
5	DenseNet121	0 1 2	0.80 0.93 0.15	0.76 0.48 0.62	0.78 0.63 0.24	0.56
6	DenseNet121 with RF	0 1 2	0.78 0.99 0.70	0.86 1.00 0.54	0.82 0.99 0.61	0.92
7	DenseNet121 with SVM	0 1 2	0.69 0.99 0.71	0.95 0.96 0.38	0.80 0.98 0.50	0.90
8	Xception	0 1 2	0.85 0.96 0.38	0.82 0.98 0.38	0.84 0.97 0.38	0.90
9	Xception with RF	0 1 2	0.87 1.00 0.67	0.96 0.99 0.50	0.92 0.99 0.57	0.95
10	Xception with SVM	0 1 2	0.97 1.00 1.00	1.00 1.00 0.88	0.98 1.00 0.93	0.99
11	Resnet152	0 1 2	0.80 1.00 0.67	0.95 0.96 0.62	0.87 0.98 0.64	0.92
12	Resnet152 with RF	0 1 2	1.00 1.00 0.93	1.00 0.99 1.00	1.00 0.99 0.96	0.99
13	Resnet152 with SVM	0 1 2	0.90 1.00 0.85	0.90 1.00 0.85	0.90 1.00 0.85	0.97
14	InceptionV3	0 1 2	0.90 0.99 0.52	0.70 0.97 0.79	0.79 0.98 0.63	0.89
15	InceptionV3 with RF	0 1	0.82 0.95	0.85 1.00	0.84 0.97	0.88

		2	0.56	0.36	0.43	
16	InceptionV 3 with SVM	0 1 2	0.96 1.00 1.00	1.00 1.00 0.93	0.98 1.00 0.96	0.99
17	VGG16	0 1 2	0.82 0.99 1.00	1.00 1.00 0.50	0.90 0.99 0.67	0.94
18	VGG16 with RF	0 1 2	1.00 1.00 1.00	1.00 1.00 1.00	1.00 1.00 1.00	1.00
19	VGG16 with SVM	0 1 2	0.92 1.00 0.80	0.89 1.00 0.86	0.91 1.00 0.83	0.96
20	EfficientNe tV2B0	0 1 2	0.84 0.99 0.89	0.96 1.00 0.57	0.90 0.99 0.70	0.94
21	EfficientNe tV2B0 with RF	0 1 2	0.81 0.95 0.50	0.93 0.97 0.29	0.86 0.96 0.36	0.88
22	EfficientNe tV2B0 with SVM	0 1 2	0.93 0.99 0.85	0.93 1.00 0.79	0.93 0.99 0.81	0.96

Table 5 presents a thorough assessment of 22 DL and hybrid models for three-class classification (HC, PD, and SWEDD). In general, hybrid models that combine deep CNN feature extractors with traditional machine learning classifiers (RF or SVM) routinely outperform standalone deep learning models. ResNet50 with RF, Xception with SVM, InceptionV3 with SVM, and ResNet152 with RF/SVM achieve nearly perfect F1-scores and very high accuracies (≈0.97-0.99), which clearly indicates the success of ensemble and hybrid learning approaches in discovering highly discriminative patterns from multimodal or imaging data. Conversely, standalone models such as DenseNet121 and InceptionV3 perform relatively poorly, particularly for class 2 (SWEDD), where suboptimal recall and F1-scores indicate class confusion. SWEDD subjects may or may not be converted into PD patients. They have hybrid symptoms; thus some samples are misclassified between PD and SWEDD. This means that for successful multi-class classification in clinically overlapping conditions like PD and SWEDD, it may be necessary to look beyond deep features. In general, the study has found that hybrid CNN-ML models are better and more reliable, and hence more suitable for trustworthy clinical decision support systems in the classification of neurological disorders. The confusion matrices pertaining to each model are presented in the subsequent section, accompanied by a comprehensive explanation, analysis, and justification. Figure 4 illustrates the confusion matrix corresponding to Xception&Resnet152.

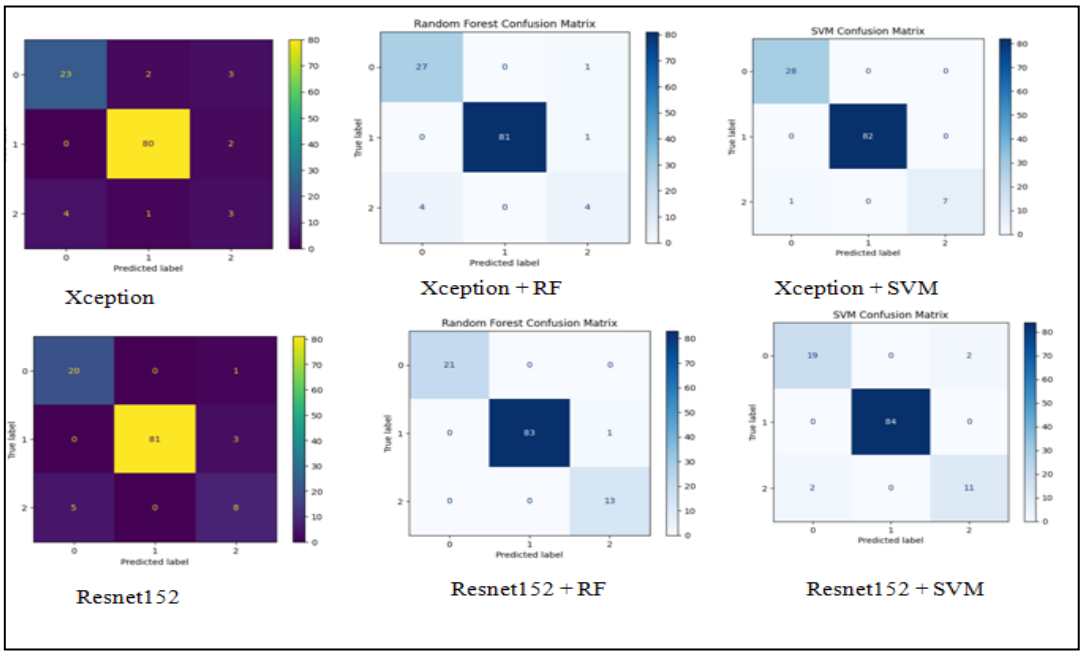

Figure 4. Confusion Matrix for Xception and Resnet152

For a three-class classification task (Class 0, Class 1, and Class 2), confusion matrices for the Xception, Xception with RF, Xception with SVM, ResNet-152 with RF, ResNet-152, and ResNet-152 with SVM models are depicted in Figure 5 and 6. These confusion matrices provide an in-depth class-wise analysis of the effectiveness, error distribution, and prediction accuracy of the hybrid deep learning models. With 80 correctly classified samples, the standalone Xception model demonstrates a strong discriminative ability for Class 1, thereby validating the depthwise separable convolutional operation as a superior technique for extracting high-level semantic features. However, when the CNN classifier is employed independently, errors are primarily observed between Class 0 and Class 2, thereby establishing the similarity in feature representations. The focus of our research is the detection of PD class, not the SWEDD class. SWEDD subjects have hybrid symptoms so it is difficult to obtain good results for that. Similarly, the Xception with SVM model shows good accuracy for Class 0 and Class 1, with 28 and 82 correctly identified samples, respectively. The lower misclassification errors suggest that the margin-based learning paradigm of SVM is a good fit for high-dimensional deep feature fields. Although Class 2 is still a source of confusion, the hybridization technique is validated because overall performance is better than that of the standalone Xception model. The ResNet-152 baseline model also performs well for Class 1 (81 true positives) because of its deeper design and residual connections that enable effective feature propagation. The hybrid model of ResNet-152 with RF shows notable improvements in performance, with high true positive rates for all classes (21 for Class 0, 83 for Class 1, and 13 for Class 2). The Random Forest classifier efficiently utilizes the deep hierarchical features of ResNet-152, improving class separability and lowering misclassifications, especially for the minority class. Lastly, out of all the investigated methods, the ResNet-152 with SVM model produces the most balanced confusion matrix with high diagonal values and minimal errors. The effectiveness of SVM in identifying optimal decision boundaries using deep residual features is highlighted by the correct classification of 84 instances for Class 1 and the improved detection of Class 2 (11 true positives). Hybrid deep learning approaches that leverage classical machine learning classifiers (RF and SVM) clearly outperform the CNN-based architectures alone, as evidenced by the confusion matrix analysis. More discriminative features are obtained by the use of deeper networks such as ResNet-152 and Xception, but their performance is

further enhanced by the combination of RF or SVM, particularly for the minority and overlapping classes. The most balanced and reliable class-wise performance among all the tested models is demonstrated by ResNet-152 with SVM and Xception with RF, which highlight their suitability for robust multi-class classification in real-world biomedical and clinical applications. After that we have drawn the AUC-ROC curve for the DNN and Xception transfer learning ensemble models, which is shown below in Figures 5 and 6.

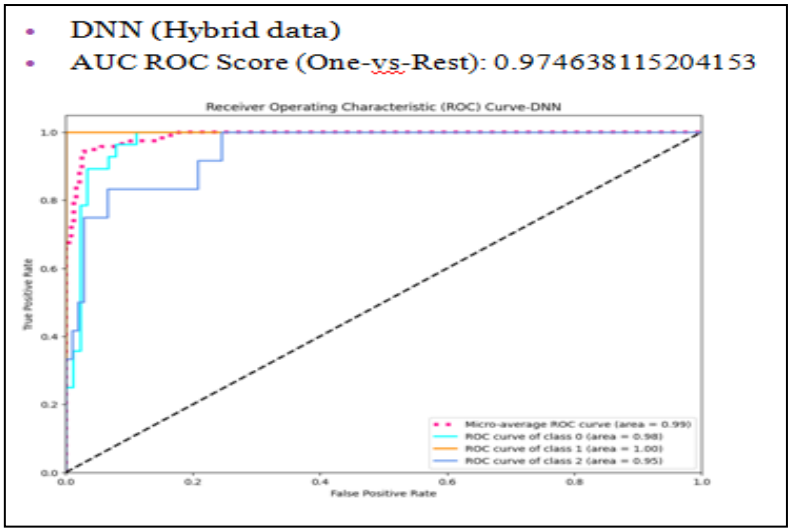

Figure 5. AUC- ROC Curve for DNN

As shown by the ROC curves in Figure 5, the Deep Neural Network (DNN) trained on hybrid data is evaluated using the one-vs-rest (OvR) approach for a three-class classification problem. Plotting the TPR versus the FPR at different decision thresholds allows the ROC curves to deliver a comprehensive assessment of the model's separability abilities. Random categorization is denoted by the diagonal dashed line, which provides a reference point. With an extremely high overall AUC-ROC value of 0.9746, the DNN model clearly exhibits high separability capability among the three classes and performs outstandingly well in the classification task. The AUC values remain persistently high in the class-wise ROC analysis; Class 1 obtained an AUC of 1.000, which is nearly perfect. Class 0 and Class 2 also perform exceptionally well, with AUC values of approximately 0.983 and 0.953, respectively. The ROC curves of all classes approach the top left corner of the plot, signifying increased sensitivity and low false positive rates. The well-rounded performance of the model is further validated by the almost identical macro-averaged ROC curve, which closely matches the individual class ROC curves. Despite the possibility of class imbalance in the data, the slightly different curves for each class reflect the small differences, which make the DNN generalized and not overly weighted toward any specific class. With hybrid data, which most likely contains complementary information from more than one modality or feature, the high AUC-ROC value indicates the high discriminative ability of the DNN model. The results are quite reliable for multi-class clinical or biomedical decision support applications, as the one-vs-rest classification assessment strategy guarantees separate comparisons of each class against all other classes. Although the slightly lower but still praiseworthy AUC for Class 2 indicates the presence of inherent inter-class overlap or increased variability in the feature space, the near-perfect AUC for Class 1 specifies that the DNN model is adept at successfully learning the dominant and highly discriminative patterns of this specific class. Nevertheless, the appropriateness of the DNN model for high-risk classification applications, where both high sensitivity and high specificity are required, is justified by the consistently high AUC values

for all classes. The suitability of the proposed DNN model as a strong baseline and a competitive component within the proposed hybrid DL framework is supported by the ROC analysis, which generally confirms that it provides excellent class separability, high robustness, and reliable generalization capabilities.

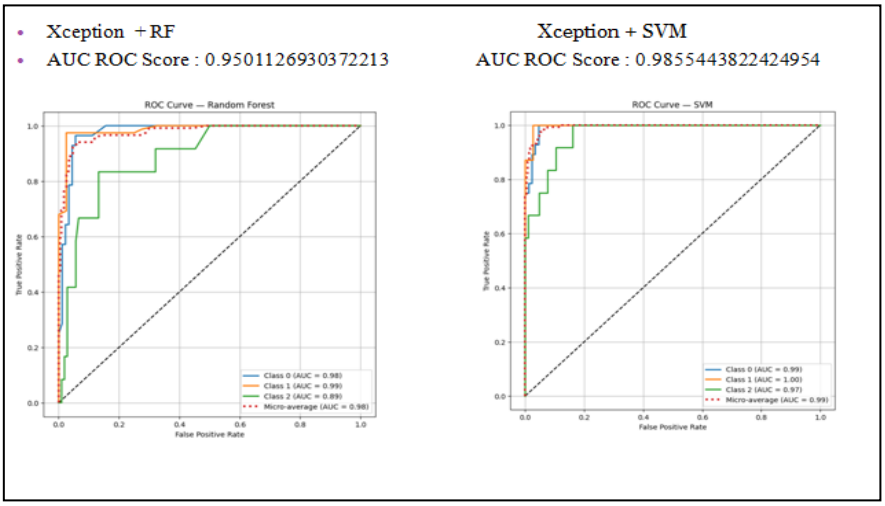

Figure 6. AUC- ROC Curve for Xception with RF and Xception with SVM

Using RF and SVM classifiers, Figure 6 shows the ROC curves for Xception feature extraction on three classes. The overall AUC-ROC for the Xception with Random Forest combination is 0.9501 (95.01%), with the class-wise AUCs being as follows: Class 0, AUC = 0.969; Class 1, AUC = 0.961; Class 2, AUC = 0.920. The micro-average AUC is 0.98. Although Class 2 shows relatively less performance with a less steep slope, all curves show excellent discriminative power, evident from the initial steep rises. The Xception with SVM combination results in an overall AUC-ROC of 0.9855 (98.55%), showing an improvement of 3.54 percentage points over the RF combination. The class-wise AUCs are as follows: Class 0, AUC = 0.993; Class 1, AUC = 0.982; Class 2, AUC = 0.982, with a micro-average AUC of 0.99. The ROC curves show initial vertical steep rises and compact clustering around the upper-left corner, signifying increased sensitivity and specificity. For all classes combined, SVM outperformed RF, with the maximum improvement of 6.2 percentage points in Class 2 (from 0.920 to 0.982). Improvements for Classes 0 and 1 were 2.4 and 2.1 percentage points, respectively. This observation indicates that SVM handles complex decision boundaries better, especially for difficult classes. Both models have AUCs above 0.95, ensuring the effectiveness of Xception-based feature extraction. However, the Xception with SVM combination offers the best and most balanced multi-class discrimination, with the least variations among classes.

Then finally we have calculated no. of parameter, Model training time Time and FLOPS for each model and it's comparative analysis is shown below in table 6.

Table 6. Comparative of All Models with it's Parameters, Time and FLOPs

Sr no.	Model	Accuracy	Parameters	FLOPS	Time (Sec)
1	DNN model	0.9180	4959942	0.1055 GFLOPs	0.00728
2	ResNet50 with RF	0.98	23662275	0.3069 GFLOPs	176.86
3	DenseNet121 with RF	0.92	7112067	1.2372 GFLOPs	360.35
4	Xception with SVM	0.99	20936043	0.0992 GFLOPs	187.46
5	ResNet152 with RF	0.99	58445507	0.4854 GFLOPs	470.01
6	InceptionV3 with SVM	0.99	21877347	0.2265 GFLOPs	214.20
7	VGG16 with SVM	0.96	14789251	5.7065 GFLOPs	147.00
8	EfficientNetV2B0 with SVM	0.96	5993875	1.9823 GFLOPs	220.53

Table 6 offers a thorough comparison of eight ensemble models assessed for classification task accuracy, computational complexity (parameters and FLOPs), and inference time.

1. DNN Model (Baseline): This model had the quickest inference time (7.28 ms), smallest computational complexity (4,959,942 parameters, 0.1055 GFLOPs), and medium accuracy (0.9180). Although it had the quickest inference time, the limited depth restricts the capability of feature extraction, leading to lower accuracy compared to other models with greater depths.

2. ResNet50 with RF: It had high accuracy (0.98) with a medium inference time (176.86 ms) and relatively low computational complexity (23,662,275 parameters, 0.3069 GFLOPs). It is a balanced model since the residual learning approach allows for efficient deep learning while maintaining computational simplicity.

3. DenseNet121 with RF: It was as accurate as the baseline model (0.92) despite having a high computational complexity (7,112,067 parameters, 1.2372 GFLOPs, 360.35 ms). The current feature and classifier combination is suboptimal because, despite having a theoretically efficient architecture, the dense connection pattern did not result in high accuracy with the RF classifier.

4. Xception with SVM: Its low computational complexity (20,936,043 parameters, 0.0992 GFLOPs—the lowest among high-performing models) and medium inference time (187.46 ms) were accompanied by high accuracy (0.99). When paired with the depthwise separable convolution technique, SVM's excellent classification accuracy offers the best possible balance between accuracy and efficiency.

5. ResNet152 with RF: This model had the longest inference time (470.01 ms) and the highest computational complexity (58,445,507 parameters, 0.4854 GFLOPs), but it also had the highest accuracy (0.99). Although the accuracy was the highest, the 2.47× increase in parameters compared to ResNet50 only resulted in a 1 percentage point improvement, suggesting that too much depth causes diminishing returns.

6. The combination of InceptionV3 and a support vector machine (SVM) produced exceptional accuracy (0.99) with a reasonable inference time (214.20 ms) and comparatively balanced computing requirements (21,877,347 parameters; 0.2265 GFLOPs). Multi-scale feature extraction using inception layers and SVM is added to increase discriminative capability without compromising computational viability.

7. VGG16 with SVM: This combination has a moderate inference time (147.00 ms) and very high computing requirements (5.7065 GFLOPs—highest among the comparisons), while having comparatively fewer parameters (14,789,251) and moderate-high accuracy (0.96). Due to the absence of contemporary architectural improvements (dense connections and residual learning), computational viability is not always enhanced by a lower parameter count.

8. EfficientNetV2B0 with SVM: This combination offers moderate-to-high accuracy (0.96), but at the expense of a higher number of parameters (5,993,875) and

comparatively high computational complexity (1.9823 GFLOPs, 220.53 ms). Its accuracy is lower than that of Xception and InceptionV3, despite its efficiency-focused design, suggesting that the compound scaling strategy might not be the most appropriate for this particular issue.

We have concluded some important observations as follows.

With an accuracy of 0.99, the models with the best performance levels were Xception with SVM, ResNet152 with RF, and InceptionV3 with SVM. With an accuracy of 0.98, ResNet50 with RF also demonstrated strong performance. The EfficientNetV2B0 with SVM and VGG16 with SVM models, on the other hand, performed reasonably well, achieving accuracies of 0.96.

In terms of computational efficiency, the Xception with SVM has the best accuracy-to-FLOPs ratio (0.99 accuracy at 0.0992 GFLOPs), followed by the InceptionV3 with SVM (0.99 at 0.2265 GFLOPs). The dense neural network (DNN) and DenseNet121 with RF are baseline models (0.92). Although it has high accuracy, the ResNet152 with RF consumes 4.89 times more FLOPs than the Xception to achieve the same level of performance. The DNN model has the fastest inference time (7.28 ms) but the lowest accuracy (0.9180) when the tradeoff between inference time and accuracy is taken into account. The ResNet152 with RF model has a high latency of 470.01 ms, while the Xception with SVM (187.46 ms) and InceptionV3 with SVM (214.20 ms) provide the best tradeoff among the high-accuracy models. The better discriminative ability of SVM for deep CNN-based features is demonstrated by the consistent outperformance of RF-based models with the same architectures in the Classifier Impact evaluation.

5. Conclusion

This research aims to improve the discernment between PD patients and HC in the diagnostically uncertain SWEDD population by combining clinical information with the characteristic DaTSCAN SPECT imaging features. Compared to previous approaches, such as Khachnaoui et al., who evaluated 548 patients using hierarchical clustering, our model provides significantly better results. In the prior study, hierarchical clustering produced an accuracy of 64%, specificity of 38.89%, sensitivity of 78.13%, and F1 score of 73.53% using clustering methods such as DBSCAN and K-means. However, we utilized a somewhat larger collection of 589 patients (392 PD, 135 HC, and 62 SWEDD) with both clinical and imaging data. With an accuracy of 91.8%, sensitivity of 86%, specificity of 82%, and F1 score of 82.3%, the DNN model performed noticeably better when evaluated on the unbalanced set. We then built an ensemble for improved optimization using various transfer learning models that included RF and SVM. The top model was found to be Xception with SVM configuration, which achieved high accuracy (0.99), the lowest computational cost (0.0992 GFLOPs) among the top models, and a reasonably acceptable inference time (187.46 milliseconds). ResNet152 with RF demonstrated considerable computational complexity despite achieving great accuracy. The most practical approach is the Xception with SVM setup, which offers the right balance between computational complexity and accuracy. The aforementioned results show how well the integrated model can detect patients with PD, HC, and SWEDD, particularly in the challenging, data-constrained diagnostic task environment. Additional data modalities, such as motor function assessments, genetic screenings, more clinical and imaging data or overall clinical outcomes, could be added to the model to further improve it. The framework is scalable

for larger clinical imaging datasets. Future research using PET scans or MRI can be a better choice, as investigating alternative imaging modalities is much more sensitive and specific. Additionally, this method might be scalable and dependable in real-world clinical settings when used on larger datasets, thus leading to more accurate SWEDD patient discrimination.

Funding

No outside funding was obtained

Data Availability Statement

Sharing data is not authorized by the authors.

Acknowledgments

The PPMI constitutes a collaborative endeavor between the public and private sectors, smoothed by the Michael J. Fox Foundation for Parkinson's Research alongside an array of pharmaceutical & biotechnology firms. The participating corporations encompass Abbott Laboratories, Avid Radiopharmaceuticals, GE Healthcare, Covance, Elan, Genentech, Bristol-Myers Squibb, EliLilly, GlaxoSmithKline and Company, Hoffmann-La Roche, Merck & Co., Pfizer, Meso Scale Discovery,and UCB.

Conflicts of Interest

No conflicts of interest or personal affiliations could have impacted the research conducted for this investigation.

References

[1] Tolosa, Eduardo, Gregor Wenning, and Werner Poewe. "The Diagnosis of Parkinson's Disease." The Lancet Neurology 5, no. 1 (2006): 75-86.

[2] Jankovic, Joseph. "Parkinson's Disease: Clinical Features and Diagnosis." Journal of neurology, neurosurgery & psychiatry 79, no. 4 (2008): 368-376.

[3] Schrag, A1, C. D. Good, K. Miszkiel, H. R. Morris, C. J. Mathias, A. J. Lees, and N. P. Quinn. "Differentiation of Atypical Parkinsonian Syndromes with Routine MRI." Neurology 54, no. 3 (2000): 697-697.

[4] Hayes, Michael T. "Parkinson's Disease and Parkinsonism." The American journal of medicine 132, no. 7 (2019): 802-807.

[5] Das, Samarjit, Laura Trutoiu, Akihiko Murai, Dunbar Alcindor, Michael Oh, Fernando De la Torre, and Jessica Hodgins. "Quantitative Measurement of Motor Symptoms in Parkinson's Disease: A Study with Full-Body Motion Capture Data." In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, 6789-6792.

[6] Thobois, Stéphane, Stéphane Prange, Christian Scheiber, and Emmanuel Broussolle. "What a Neurologist Should Know About PET and SPECT Functional Imaging for

Parkinsonism: A Practical Perspective." Parkinsonism & related disorders 59 (2019): 93-100.

[7] Marek, Kenneth, John Seibyl, Shirley Eberly, David Oakes, Ira Shoulson, Anthony E. Lang, Chris Hyson, Danna Jennings, and Parkinson Study Group PRECEPT Investigators. "Longitudinal Follow-Up of SWEDD Subjects in the PRECEPT Study." Neurology 82, no. 20 (2014): 1791-1797.

[8] De Rosa, Anna, Claudia Carducci, Carla Carducci, Silvio Peluso, Maria Lieto, Andrea Mazzella, Francesco Saccà et al. "Screening for Dopa-Responsive Dystonia in Patients with Scans Without Evidence of Dopaminergic Deficiency (SWEDD)." Journal of neurology 261, no. 11 (2014): 2204-2208.

[9] Taylor, Jonathan Christopher, and John Wesley Fenner. "Comparison of Machine Learning and Semi-Quantification Algorithms For (I123) FP-CIT Classification: The Beginning of the End for Semi-Quantification?." EJNMMI physics 4, no. 1 (2017): 29.

[10] Jalalian, Afsaneh, Syamsiah BT Mashohor, Hajjah Rozi Mahmud, M. Iqbal B. Saripan, Abdul Rahman B. Ramli, and Babak Karasfi. "Computer-Aided Detection/Diagnosis of Breast Cancer in Mammography and Ultrasound: A Review." Clinical imaging 37, no. 3 (2013): 420-426.

[11] Roth, Holger R., Le Lu, Jiamin Liu, Jianhua Yao, Ari Seff, Kevin Cherry, Lauren Kim, and Ronald M. Summers. "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation." IEEE transactions on medical imaging 35, no. 5 (2015): 1170-1181.

[12] Ayech, Marouane Ben Haj, and Hamid Amiri. "Texture Description Using Statistical Feature Extraction." In 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE, 2016, 223-227.

[13] Firmino, Macedo, Giovani Angelo, Higor Morais, Marcel R. Dantas, and Ricardo Valentim. "Computer-Aided Detection (CADe) and Diagnosis (CADx) System for Lung Cancer with Likelihood of Malignancy." Biomedical engineering online 15, no. 1 (2016): 2.

[14] Aboudi, Noura, Ramzi Guetari, and Nawres Khlifa. "Multi-Objectives Optimisation of Features Selection for the Classification of Thyroid Nodules in Ultrasound Images." IET Image Processing 14, no. 9 (2020): 1901-1908.

[15] Mastouri, Rekka, Nawres Khlifa, Henda Neji, and Saoussen Hantous-Zannad. "A Bilinear Convolutional Neural Network for Lung Nodules Classification on CT Images." International Journal of Computer Assisted Radiology and Surgery 16, no. 1 (2021): 91-101.

[16] Prashanth, R॥, Sumantra Dutta Roy, Pravat K. Mandal, and Shantanu Ghosh. "Automatic Classification and Prediction Models for Early Parkinson's Disease Diagnosis from SPECT Imaging." Expert Systems with Applications 41, no. 7 (2014): 3333-3342.

[17] Mabrouk, Rostom, Belkacem Chikhaoui, and Layachi Bentabet. "Machine Learning Based Classification Using Clinical and Datscan SPECT Imaging Features: A Study on

Parkinson’s Disease and SWEDD." IEEE Transactions on Radiation and Plasma Medical Sciences 3, no. 2 (2018): 170-177.

[18] Segovia, F., J. M. Górriz, J. Ramírez, J. Levin, M. Schuberth, M. Brendel, A. Rominger, G. Garraux, and C. Phillips. "Analysis of 18F-DMFP PET Data Using Multikernel Classification in Order to Assist the Diagnosis of Parkinsonism." In 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE, 2015, 1-4.

[19] Segovia, Fermín, Juan Manuel Górriz, J. Ramırez, and Diego Salas-Gonzalez. "Multiclass Classification of 18 F-DMFP-PET Data to Assist the Diagnosis of Parkinsonism." In 2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI), IEEE, 2016, 1-4.

[20] Khachnaoui, Hajer, Nawres Khlifa, and Rostom Mabrouk. "Machine Learning for Early Parkinson’s Disease Identification Within SWEDD Group Using Clinical and Datscan SPECT Imaging Features." Journal of Imaging 8, no. 4 (2022): 97.

[21] Parkinson's Progression Markers Initiative. (n.d.). PPMI – Parkinson's Progression Markers Initiative. Laboratory of Neuro Imaging, USC. https://ida.loni.usc.edu/home/projectPage.jsp?project=PPMI

[22] Kurmi, Ankit, Shreya Biswas, Shibaprasad Sen, Aleksandr Sinitca, Dmitrii Kaplun, and Ram Sarkar. "An Ensemble of CNN Models for Parkinson’s Disease Detection Using DaTscan Images." Diagnostics 12, no. 5 (2022): 1173.

[23] Prashant, R. "Accurate Early Detection of Parkinson’s Disease from Single Photon Emission Computed Tomography Imaging Through Convolutional Neural Networks." Artificial Intelligence in Health 2, no. 4 (2025): 22.

[24] Majhi, Babita, Aarti Kashyap, Siddhartha Suprasad Mohanty, Sujata Dash, Saurav Mallik, Aimin Li, and Zhongming Zhao. "An Improved Method for Diagnosis of Parkinson’s Disease Using Deep Learning Models Enhanced with Metaheuristic Algorithm." BMC medical imaging 24, no. 1 (2024): 156.

[25] Dentamaro, Vincenzo, Donato Impedovo, Luca Musti, Giuseppe Pirlo, and Paolo Taurisano. "Enhancing Early Parkinson’s Disease Detection Through Multimodal Deep Learning and Explainable AI: Insights from the PPMI Database." Scientific reports 14, no. 1 (2024): 20941.

[26] Chang, Yan, Jiajin Liu, Shuwei Sun, Tong Chen, and Ruimin Wang. "Deep Learning for Parkinson’s Disease Classification Using Multimodal and Multi-Sequences PET/MR Images." EJNMMI research 15, no. 1 (2025): 55.

[27] Cao, Jiangbo, and Xiaojing Long. "Racf: A Multimodal Deep Learning Framework for Parkinson’s Disease Diagnosis Using Snp and MRI Data." Applied Sciences 15, no. 8 (2025): 4513.

[28] Majhi, Babita, Aarti Kashyap, Siddhartha Suprasad Mohanty, Sujata Dash, Saurav Mallik, Aimin Li, and Zhongming Zhao. "An Improved Method for Diagnosis of Parkinson’s Disease Using Deep Learning Models Enhanced with Metaheuristic Algorithm." BMC medical imaging 24, no. 1 (2024): 156.

[29] Sar, Ayan, Pranav Singh Puri, Huma Naz, Sumit Aich, Tanupriya Choudhury, and Lubna Abdelkhreim Gabralla. "Multi-Modal Deep Learning Framework for Early Detection of Parkinson's Disease Using Neurological and Physiological Data for High-Fidelity Diagnosis." Scientific Reports 15, no. 1 (2025): 34835.

[30] Chowdhury, Nishu, Utpol Kanti Das, and Amit Chowdhury. "Multimodal Approach for Early Diagnosis of Parkinson's Disease Using PET Imaging, Tremor Detection, and Machine Learning." Psychiatry Research: Neuroimaging (2025): 112063.

[31] Suo, Xueling, Mengyao Chen, Li Chen, Chunyan Luo, Graham J. Kemp, Su Lui, and Huaiqiang Sun. "Automatic Identification of Parkinsonism Using Clinical Multi-Contrast Brain MRI: A Large Self-Supervised Vision Foundation Model Strategy." EBioMedicine 116 (2025).

[32] Welton, Thomas, Septian Hartono, Weiling Lee, Peik Yen Teh, Wenlu Hou, Robert Chun Chen, Celeste Chen et al. "Classification of Parkinson's Disease by Deep Learning on Midbrain MRI." Frontiers in Aging Neuroscience 16 (2024): 1425095.

[33] Khalil, Rana M., Lisa M. Shulman, Ann L. Gruber-Baldini, Stephen G. Reich, Joseph M. Savitt, Jeffrey M. Hausdorff, Rainer von Coelln, and Michael P. Cummings. "Applying Wearable Sensors and Machine Learning to the Diagnostic Challenge of Distinguishing Parkinson's Disease from Other Forms of Parkinsonism." Biomedicines 13, no. 3 (2025): 572.

[34] Shokrpour, Sahar, AmirMehdi MoghadamFarid, Sepideh Bazzaz Abkenar, Mostafa Haghi Kashani, Mohammad Akbari, and Mostafa Sarvizadeh. "Machine Learning for Parkinson's Disease: A Comprehensive Review of Datasets, Algorithms, And Challenges." npj Parkinson's Disease 11, no. 1 (2025): 187.

[35] Ling, Ronghua, Xingxing Cen, Shaoyou Wu, Min Wang, Ying Zhang, Juanjuan Jiang, Jiaying Lu et al. "Explainable Graph Neural Network Based on Metabolic Brain Imaging for Differential Diagnosis of Parkinsonism." Frontiers in Aging Neuroscience 17 (2025): 1580910.

[36] Bakaraniya Parul V., Dr. Mamta C. Padole. (2025). Multimodal Deep Learning Framework Integrating Clinical and DaTSCAN SPECT Features for Early Parkinson's Disease Detection in SWEDD Cohort. Journal of Applied Bioanalysis, 11(S2), 269-284. https://doi.org/10.53555/jab.v11si2.539