ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

# Accurate early detection of Parkinson's disease from single photon emission computed tomography imaging through convolutional neural networks

**R. Prashanth***

Independent Researcher, Bengaluru, Karnataka, India

## Abstract

Early and accurate detection of Parkinson's disease (PD) remains a crucial diagnostic challenge with substantial clinical implications, particularly for ensuring effective treatment and patient management. For instance, a group of subjects with scans without evidence of dopaminergic deficit (SWEDD) who are initially diagnosed as PD but exhibit normal single photon emission computed tomography (SPECT) scans. Over time, follow-up assessments often lead to a revised diagnosis of non-PD. In the meantime, these subjects may receive PD-specific medications that can cause more harm than benefit. In this paper, a case study is presented in which machine learning models are developed and trained on SPECT images to distinguish early PD from healthy controls, as well as to differentiate SWEDD cases from early PD. The case study utilizes a well-known, publicly available dataset and explores several machine learning classifiers, including support vector machines, logistic regression, feed forward neural networks, and convolutional neural networks (CNNs). The CNN model gave the best performance in differentiating PD from healthy subjects. All these models demonstrated strong potential for early differentiation of SWEDD cases from PD. These results suggest that the proposed approach could support clinicians in making more accurate and timely diagnostic decisions.

***Keywords:*** Computer-aided diagnosis; Machine learning; Deep learning; Parkinson's disease; Medical imaging

***Corresponding author:**
R. Prashanth
(prashanth.r.iitd@gmail.com)

**Citation:** Prashanth R. Accurate early detection of Parkinson's disease from single photon emission computed tomography imaging through convolutional neural networks. *Artif Intell Health*. 2025;2(4):22-32.
doi: 10.36922/AIH025040005

**Received:** January 21, 2025

**1st revised:** May 13, 2025

**2nd revised:** May 22, 2025

**Accepted:** May 30, 2025

**Published online:** June 17, 2025



## 1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting millions of people worldwide and is characterized by the loss of dopaminergic neurons in the substantia nigra.[1,2] Its prevalence increases with age, impacting approximately 1% of individuals over 60 years.[3] The clinical diagnosis of PD is challenging as there are no definitive diagnostic tests and the diagnosis is based on the presence of cardinal symptoms, such as tremor at rest, rigidity, and bradykinesia, along with a subject's response to PD medications.[1] However, these symptoms appear in the later stages of the disease and by the time the patient manifests these symptoms, the patient might have already crossed the early stage of the disease.[4] Early detection of PD is important because appropriate targeted therapies could be initiated before any major deterioration occur.[5]

It can also help develop treatments and identify patients eligible for therapeutic clinical trials.[5]

Single photon emission computed tomography (SPECT) imaging using $^{123}$I-Ioflupane (DaTSCAN or [123I]FP-CIT) has been shown to increase the diagnostic accuracy of PD, mainly in the earlier stages of the disease, by showing the functional deterioration or dopaminergic deficit in the striatal region of the brain (which is one of the primary regions affected in PD).[6-9] The accuracy of diagnosis of PD at an early phase is the poorest based on clinical indices as early symptoms are mild/moderate, unlike in advanced stages of the disease.[4,5] Furthermore, these symptoms are common in other neurodegenerative disorders like essential tremor and multiple system atrophy, which often leads to misdiagnosis.[10-12] The effects of misdiagnosis are severe as it may lead to unnecessary medical examinations and therapies, and associated side-effects. Recent studies have shown that around 3.6 – 19.6% of clinically diagnosed PD subjects show no dopaminergic deficit, and these subjects are classified as scans without evidence of dopaminergic deficit (SWEDD).[10-12] Subsequent follow-up on these subjects have shown that they neither deteriorate nor respond to levodopa (a primary medication in PD) and that their SPECT scans remain normal in the follow-up imaging. Thus, these subjects were considered highly unlikely of having PD and that the initial diagnosis of PD was incorrect.[13-15] These studies evidently point out that dopaminergic imaging is highly useful and that an abnormal imaging, at least in cases of diagnostic uncertainty, is strongly supportive of a diagnosis of neurodegenerative Parkinsonism (PS), such as PD.

In clinical practice, SPECT images are usually analyzed by visual inspection and/or by region of interest (ROI) analysis.[16] The visual analysis relies on the judgment of the observer that heavily depends on his expertise, experience, and knowledge.[17] ROI techniques involve outlining or positioning the ROI over the striatum (target region) and the occipital cortex (reference region), and a quantitative measure termed the background subtracted striatal uptake ratio is computed.[6] Despite the odds, the latter method or the quantitative method is the most acceptable one, since, according to a trial study, it provides an excellent intra- and inter-observer agreement.[18] However, the ROI-based approach relies on manual intervention for placing the ROIs.

There have been many studies that make use of machine learning techniques to develop predictive models from SPECT imaging features for the early detection of PD.[11,19-31] Segovia *et al.*[28] extracted voxels corresponding to the striatum and performed data decomposition using partial least squares followed by classification into controls and Parkinsonism by means of a support vector machine (SVM) classifier. Illan *et al.*[29] also used voxels corresponding to the striatum to train an SVM classifier with a linear kernel to classify controls and PS. Rojas *et al.*[30] used voxels corresponding to the striatum and then carried out feature reduction through principal component analysis followed by classification using SVM. Towey *et al.*[31] performed feature extraction on all voxels through singular value decomposition followed by classification into PS or non-PS. Huertas-Fernández *et al.*[32] calculated the bilateral caudate and putamen uptake and asymmetry indices from SPECT images and developed predictive models using logistic regression, SVM, and LDA to classify PD from vascular Parkinsonism. Kim *et al.*[20] used image augmentation to increase the size of data and a classifier based on the Inception v3 model that can classify normal from abnormal SPECT scans.

There are also many studies using the SPECT data from the Parkinson's progression marker initiative (PPMI), which is among the most popular, widely used, and largest database for PD,[11,19,21-27,33,34,45-48] and the same data were used in the present study. Choi *et al.*[11] trained a convolutional neural network (CNN), which they called PD net, using SPECT images to classify PD from normal and non-Parkinsonism tremor. They also used the model to classify SWEDD subjects. In their analysis, they used the complete volume data, rather than considering a selected range of slices, due to which the CNN network became complex with many layers. Martínez-Murcia *et al.*[22] also used a CNN to differentiate PD from others (healthy normal and SWEDD). They used a threshold-based approach to select sub-volumes from the volume which they later input to the CNN. They observed that due to this sub-volume selection, the complexity of the CNN became small with just two convolutional layers.[22] Martínez-Murcia *et al.*[21] used the features extracted from SPECT images through independent component analysis to train an SVM classifier to distinguish PD from normal. They observed much better performance than their previous work using the voxel-as-features approach. Hirschauer *et al.*[19] used data from different clinical examinations and SPECT imaging, and trained an enhanced probabilistic neural network model to differentiate PD from SWEDD. Oliveira and Castelo-Branco[23] used voxels as features that were extracted based on volumes of interest defined (which required manual intervention), and an SVM classifier was used to classify PD from normal. The standard binding potential features along with other features related to the volume and length of the striatal region from SPECT images was utilized to train an SVM classifier that could classify PD from healthy normal.[33] Ortiz *et al.*[24] extracted features from isosurfaces

computed from the ROI and trained a CNN-based model to classify PD from healthy normal. Prashanth *et al.*[27] computed shape- and surface-fitting-based features and used machine learning methods to develop classification models to differentiate scans with deficit, as in PD, from scans without deficit, as in normal and SWEDD. Prashanth *et al.*[26] also used data from multiple modalities including clinical examinations, laboratory examinations, and dopaminergic imaging, and developed classification models to distinguish early PD from normal. The same researcher group had used the striatal binding ratios to develop classification and prognostic models for PD.[25] Zhang *et al.*[34] employed multimodal data which included SPECT imaging data to identify different PD subtypes through the long-short term memory (LSTM) networks and dynamic time warping. Shiiba *et al.*[45] extracted radiomics features including intensity- and texture-based features in the caudate, putamen, and pallidum volumes of interest from the SPECT images and used machine learning methods to classify PD from normal. Tufail *et al.*[46] developed a 3D CNN model (consisting of 14 layers including 5 convolution, 5 max pooling, and 3 fully connected layers) that is capable of performing multiclass classification of Alzheimer's and PDs using positron emission tomography and SPECT neuroimaging modalities. Majhi *et al.*[47] used magnetic resonance imaging and SPECT imaging data to train many deep learning models including VGG16, DenseNet, DenseNet-LSTM, and InceptionV3 that are optimized through gray wolf optimization. Khachnaoui *et al.*[48] trained deep learning models based on EfficientNet-B0, Mobilenet-V2, and a custom CNN with 10 layers (4 convolutional, 4 max pooling) using SPECT images.

However, several limitations that persist in prior work are as follows:

- Use of entire image volumes, increasing model complexity and the risk of overfitting;
- Dependence on explicit feature extraction pipelines followed by machine learning classifiers;
- Manual intervention for ROI placement, reducing reproducibility and scalability;
- Training on small subject cohorts, limiting generalizability;
- Focus solely on binary classification (*e.g.*, PD vs. healthy), with limited attention to diagnostically challenging cases such as SWEDD.

The present study addresses the above limitations by developing a compact CNN-based model optimized through Bayesian hyperparameter optimization to distinguish early PD from healthy controls, as well as to differentiate diagnostically challenging SWEDD cases from those with early PD. Unlike prior approaches that relied on full volumes or manual feature engineering, this work uses only the most relevant slice(s) from SPECT volumes thereby reducing model complexity and the risk of overfitting. The model was trained and evaluated on data from the PPMI, one of the most extensive and standardized PD imaging databases available. By incorporating SWEDD into the classification task, this work contributes toward differential diagnosis within the Parkinsonian spectrum. The proposed method combines the strengths of automated feature learning with informed slice selection, enabling improved diagnostic accuracy and practical utility for early-stage PD and SWEDD detection.

## 2. Materials and methods

### 2.1. Dataset details

The data used in the study were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (http://www.ppmi-info.org/data). For up-to-date information, please visit http://www.ppmi-info.org. The PPMI is a landmark, large-scale, comprehensive, observational, international, multi-center study that recruits *de novo* (early-untreated) PD patients, and age- and gender-matched healthy subjects to identify PD progression biomarkers.[4,35]

In this work, SPECT imaging data from the screening visits of 209 healthy normals, 443 early PD, and, 80 SWEDD were used. All the subjects in the three groups are age- and gender-matched to minimize demographic bias. Table 1 shows the age, gender, and Hoehn and Yahr (HY) stage distribution for the three groups. All PD patients were in their early stage (HY stage 1 or 2 with mean ± standard deviation as $1.50 \pm 0.50$[36]) and all SWEDD subjects exhibited early-stage PD symptoms (HY stage as $1.46 \pm 0.53$).

### 2.2. Image pre-processing by PPMI

All the SPECT scans taken at different PPMI sites undergo a standard pre-processing procedure before they are publicly shared through the database.[35] This pre-processing was carried out so that all scans were in the same anatomical alignment (spatially normalized). The process includes reconstruction from raw projection data and attenuation correction, followed by application of a standard Gaussian 3D 6.0 mm filter and then normalization of these images to standard Montreal Neurologic Institute space. These pre-processed scans, which were then shared for public access, were used for this analysis. The analysis pipeline is shown in Figure 1.

### 2.3. Slice selection

Each SPECT scan consists of 91 transaxial slices (from bottom to top of the head) each of size $109 \times 91$, which

**Table 1. Details of the subjects in terms of age, gender, and the HY stage**

| Gender | Normal | | Early PD | | | SWEDD | | |
|---|---|---|---|---|---|---|---|---|
| | Count | Age (mean) | Count | Age (mean) | HY stage | Count | Age (mean) | HY stage |
| Female | 73 | 59.32 | 157 | 60.91 | 1.46±0.50 | 30 | 58.16 | 1.4±0.50 |
| Male | 136 | 61.65 | 286 | 62.13 | 1.53±0.50 | 50 | 61.80 | 1.5±0.54 |
| All | 209 | 60.79 | 443 | 61.7 | 1.51±0.50 | 80 | 60.43 | 1.46±0.53 |

Note: HY stands for Hoehn and Yahr stage.
Abbreviations: PD: Parkinson's disease; SWEDD: Scans without evidence of dopaminergic deficit.

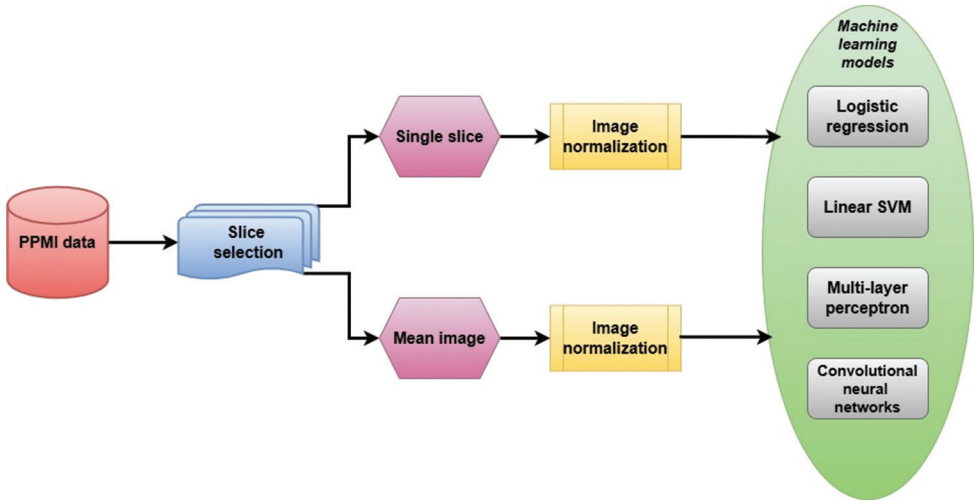


**Figure 1.** Flowchart of the analysis
Abbreviations: PPMI: Parkinson's Progression Markers Initiative; SVM: Support vector machine.

means each scan is of 3D type with size 91 × 109 × 91. In published literature,[27] the areas of striatal activity from SPECT images were analyzed and it was observed that the most relevant striatal activity came from slices 35 to 48, with the highest activity occurring in slice number 41.

In this work, two types of images were used for the analysis.

- Single slice: It is the 41$^{st}$ slice extracted from the SPECT volume, as this is the slice with maximum striatal uptake, making it very relevant for PD detection.
- Mean image: It is the average of slices from 35 to 48 extracted from the SPECT volume as these are the slices that show striatal activity.

Figure 2 shows both the single slice and mean image for the three groups: Normal control, early PD, and SWEDD. Normal scans are characterized by intense, uniform, and symmetric high-intensity regions (corresponding to the caudate and striatum) on both hemispheres that appear as two comma-shaped regions, as evident in Figure 2A and 2C. In PD, dopaminergic neuron deterioration leads to a reduction in the comma-shaped region, which becomes smaller and more circular in shape, as observed in Figure 2B.

## 2.4. Image normalization

The intensities in the original SPECT image ranged from 0 to $2^{15}$-1. To standardize the data, the selected images (both single slice as well as the mean image) were normalized by dividing the intensity values by $2^{15}$-1, so that the normalized intensity is in the range [0 – 1].

## 2.5. Data partitioning

Data were divided into two parts, namely, Partition 1 and Partition 2, in the ratio of about 80:20. Partition 1 was used for model training and evaluation using an approach based on cross-validation (10-fold). That is, Partition 1 data were split into 10 folds and then one of the folds became the evaluation set, and the remaining nine folds were used for training the model, with the whole process repeated nine times such that every fold became a test set and the remaining nine folds became the training data. Partition 2 was exclusively used for hyperparameter tuning of the machine learning methods and was not involved in model training or evaluation. All reported performance measures were solely based on the cross-validation output from Partition 1 data. An illustration of the data partitioning is shown in Figure 3.

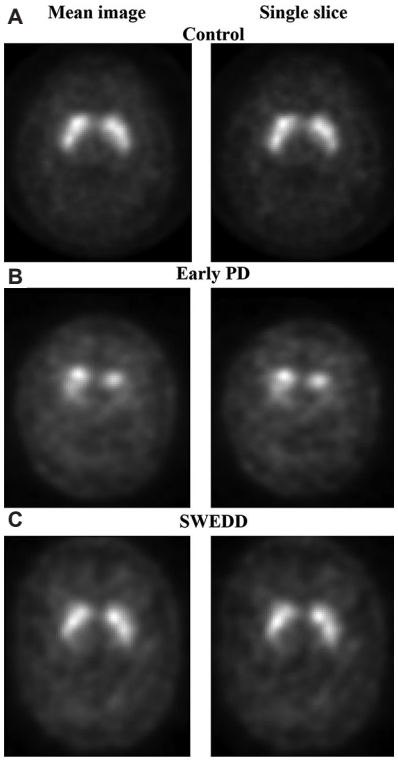


**Figure 2.** Mean image and single slice image for the three groups: normal control, early PD, and SWEDD. Mean image was created by taking the average of slices from 35 to 48 (from the total 91 slices), and the 41st slice represents the single slice used in the study.
Abbreviations: PD: Parkinson's disease; SWEDD: Scans without evidence of dopaminergic deficit.

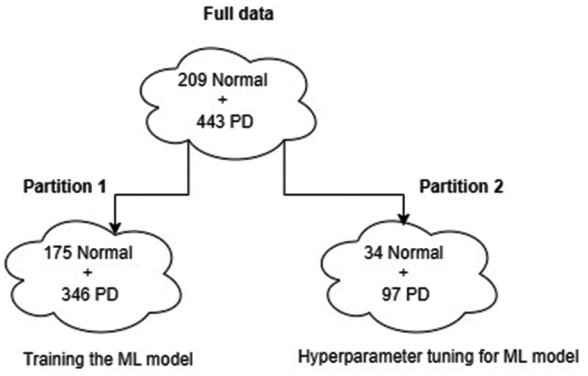


**Figure 3.** An illustration of data partitioning
Abbreviations: ML: Machine learning; PD: Parkinson's disease.

## 2.6. Machine learning techniques

The techniques utilized in the study include SVM using the linear kernel,[38] logistic regression,[39] CNN,[37] and multilayer perceptron (MLP).[37] MLP is a feed-forward neural network with one or more hidden layers. For both logistic regression and SVM, L1 regularization was employed as it has the inherent ability of feature selection and thereby enhancing numerical stability.[39] For neural network based methods, dropout technique is used that involves randomly dropping out a fraction of neurons during the training process, thereby can help in preventing overfitting.[41]

The hyperparameters of these methods are as follows:

- SVM: Regularization of parameter *C*.
- Logistic regression: Regularization of parameter *C*.
- CNN: Number of convolutional layers, number of filters in each layer, filter sizes, dropout rate in each layer, number of neurons in the fully connected layer, dropout in the fully connected layer, batch size, and number of epochs.
- MLP: Number of hidden layers, number of neurons in each layer, and dropout rate.

These hyperparameters were estimated using a hold-out set (Partition 2, as explained in the above section), while the models were trained and evaluated using the normalized images from Partition 1 through a 10-fold cross-validation approach.

### *2.6.1. CNNs for predictive modeling*

In this work, a CNN model was created and trained to classify early PD subjects from healthy normal controls. Unlike traditional approaches that rely on handcrafted features such as textures or shapes, CNNs automatically learn hierarchical feature representations directly from raw image data. When trained effectively, CNNs can extract both low-level features (*e.g.*, edges, textures) and high-level abstractions (*e.g.*, disease-relevant patterns), eliminating the need for manual feature engineering.[37] A CNN typically consists of a convolutional layer(s), a transformation layer(s), a pooling layer(s), and a fully connected layer(s). In the convolutional layer, filters are applied to regions (tiles) of the input feature map to generate new features. The hyperparameters in this layer are the size of the filter and the number of filters. During training, the CNN learns the optimal filter matrices that enable it to extract meaningful patterns from the data.

After the convolution operation, a transformation function, typically the rectified linear unit, is applied to the convolved feature. This will introduce non-linearity into the model. This is followed by a pooling step, where max pooling is typically carried out, downsampling feature map through the reduction of its spatial dimensions while preserving the most important information. In max pooling, tiles are extracted and the maximum value is taken to generate a new feature map. The max pooling filter size is $2 \times 2$, and the stride was kept as 2 in the study. At last, there is a fully connected layer(s) that performs classification based on the features from the pooling layer.

The hyperparameters in a CNN model need to be fine-tuned for optimal performance and to prevent overfitting. For instance, increasing the number of filters in the convolutional layers can help capture more diverse features, but also leads to higher computational costs and training time. Moreover, beyond a certain point, additional filters may contribute to only minimal improvements in performance.

#### *2.6.2. Fine-tuning: Hyperparameter optimization*

Model fine-tuning is important for achieving optimal performance. In this study, all models were subjected to hyperparameter optimization. For SVM and logistic regression, a grid search combined with cross-validation was employed to identify the best configuration. For the neural network-based models – MLP and CNN – a more sophisticated approach was adopted.

Research has shown that Bayesian hyperparameter optimization is more efficient than manual, random, or grid search-based methods, particularly for neural networks, both in terms of performance and the computation time required to identify optimal hyperparameters.[40] In Bayesian optimization, unlike in random search, it keeps track of past evaluation scores which is used to form a probabilistic model mapping hyperparameters to a probability of a score on the objective function $p(y|x)$. Now this probabilistic model is much easier to optimize than the original objective function, thereby helping in finding the next best set of hyperparameters to evaluate. In this paper, the Tree-structured Parzen Estimator Approach was used to estimate the probabilistic model.[40] The optimal architecture for CNN and MLP was estimated based on this optimization and is presented in the Results section.

## 3. Results and discussion

The hyperparameters of classification algorithms were estimated separately for both cases, which are single slice image and mean image, using a hold-out set (Partition 2) which was not used in either training or evaluation of the models. Table 2 shows the estimated hyperparameters.

The hyperparameters estimated vary between the single slice and mean image cases. The regularization parameter C for SVM and logistic regression increased slightly for the mean image case. This indicates that the models benefited from less regularization on the averaged data, and this is because averaging could have led to a reduction in noise and better overall representation of the pattern needed for detection. The CNN configurations estimated in the study are much more compact and efficient as compared to a related work[24] where five convolutional layers and three fully connected layers were used.

Table 3 shows the 10-fold cross-validation performance metrics for all methods applied to both single slice and mean image cases. All models demonstrated excellent classification performance, with CNN achieving the highest accuracy. The performance measures obtained for the mean image consistently gave better results as compared to a single slice, except for the MLP model. This is because the mean image likely provides a more comprehensive representation of the striatal region by integrating information across multiple slices, thereby smoothing out noise and capturing more consistent patterns relevant to early PD detection. This richer representation can help these models generalize better and improve classification accuracy.

The metrics obtained here significantly improve the results obtained elsewhere[27] and other closely related works.[11,19-31,33,34] For instance, in Prashanth *et al.*'s work,[27] a classification model was developed for the detection of early PD from normal controls using features extracted from SPECT images, achieving an accuracy of 97.29% and an area under the region operating characteristic curve (AUC) of 99.26%. In contrast, our approach achieved an accuracy of 99.08% and an AUC of 99.93% using single slice images – surpassing previous benchmarks. Notably, our method does not rely on any explicit feature extraction. Instead, CNN automatically learns discriminative features directly from the input data through its convolution and pooling operations, highlighting its capacity for effective end-to-end learning.

### 3.1. Error analysis

Among all the attempted methods, CNN gave the best performance metrics. However, as observed in Table 3, very few records were misclassified. Figure 4 shows examples of images that were misclassified, with Figure 4A showing a SPECT image from a normal subject but classified as PD, and Figure 4B showing a SPECT image from a PD subject but classified as normal. It should be noted that a normal scan is characterized by intense, uniform, and symmetric high-intensity regions on both hemispheres that appear as two comma-shaped regions (as observed in Figure 2A). In the case of Figure 4A, it is observed that the tail or the bottom of the comma-shaped region is less intense as compared to the upper region. This might be an interesting case of misdetection from the CNN model as the model is actually detecting the non-uniformity in the comma-shaped region in the image. Training the network with more images like these can help alleviate these errors. In fact, such errors may even assist clinicians by flagging potentially ambiguous or borderline cases. Similarly, Figure 4B is a case of misdetection where an early PD case is detected as normal. Here as well, it is an interesting

**Table 2. The hyperparameters estimated for machine learning models**

| Method | Single slice | Mean slice |
|---|---|---|
| SVM (linear kernel) | C=0.5 | C=1 |
| Logistic regression | C=3 | C=5 |
| MLP | Two-layer neural network with the configuration below:<br>• Hidden layer 1: Dense layer with 64 neurons, activation=ReLU<br>• Output layer: Dense with 2 neurons, activation=Sigmoid<br>• Batch size: 8<br>• Number of epochs: 30<br>• Dropout: None | Two-layer neural network with the configuration below:<br>• Hidden layer 1: Dense layer with 128 neurons, activation=ReLU<br>• Output layer: Dense with 2 neurons, activation=Sigmoid<br>• Batch size: 8<br>• Number of epochs: 50<br>• Dropout: 0.2 |
| CNN | • Layer 1: Conv2D with 64 filters, kernel size (3×3), activation=ReLU<br>• Layer 2: MaxPooling2D with pool size (2×2)<br>• Dropout: 0.1<br>• Fully connected layer: Dense with 64 neurons, activation=ReLU<br>• Dropout: 0.3<br>• utput layer: Dense with 2 neurons, activation=Sigmoid | • Layer 1: Conv2D with 32 filters, kernel size (5×5), activation=ReLU<br>• Layer 2: MaxPooling2D with pool size (2×2)<br>• Layer 3: Conv2D with 32 filters, kernel size (5×5), activation=ReLU<br>• Layer 4: MaxPooling2D with pool size (2×2)<br>• Dropout: 0.1<br>• Fully connected layer: Dense with 64 neurons, activation=ReLU<br>• Dropout: 0.1<br>• Output layer: Dense with 2 neurons, activation=Sigmoid |

Abbreviations: CNN: Convolutional neural network; MLP: Multilayer perceptron; ReLU: Rectified linear unit; SVM: Support vector machine.

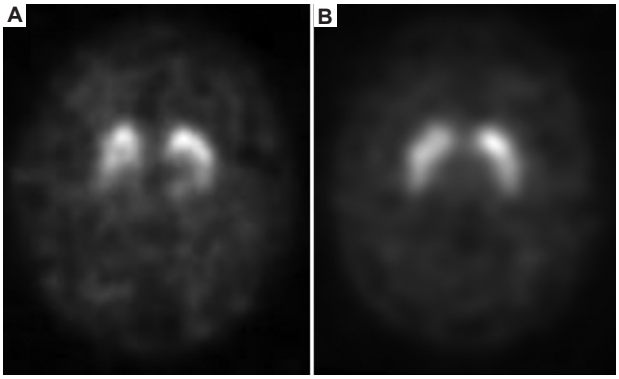


**Figure 4.** An illustration of misclassifications from the CNN model. (A) Normal detected as early PD. (B) Early PD detected as normal.
Abbreviations: CNN: Convolutional neural network; PD: Parkinson's disease.

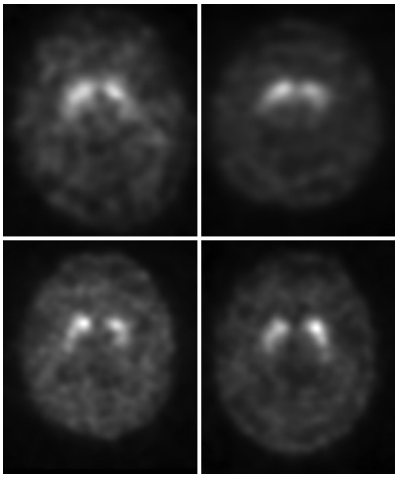

**Figure 5.** SWEDD images that were misclassified as early PD by the CNN model
Abbreviations: CNN: Convolutional neural network; PD: Parkinson's disease; SWEDD: Scans without evidence of dopaminergic deficit.

observation that the characteristics of the image appear similar to the patterns in a normal image which might have caused the misdetection.

### 3.2. Performance on SWEDD data

The SWEDD data consist of 80 subjects and were input to the machine-learned models. The performance of these methods is given in Table 4. CNN gave the best detection with an accuracy of 95% (76 out of 80). Figure 5 shows the cases of misclassification from the CNN model. It is interesting to observe that all these misclassified images show unexpected pattern of dull and uneven comma-shaped regions, which deviates from the bright and even comma-shaped regions seen in normal images. Using the PPMI data for analysis, Choi *et al.*[11] observed that a few SWEDD cases showing unusual image pattern were classified as abnormal (or PD), and the diagnosis of the majority of these cases was later changed to clinical PD based on a 2-year follow-up. This finding underscores the potential of machine learning techniques, particularly

**Table 3. Performance metrics obtained for different methods for using single slice image and mean image**

| Method | Confusion matrix | Accuracy | AUC | APR | Precision | Recall | Specificity |
|---|---|---|---|---|---|---|---|
| **A. Single slice image** | | | | | | | |
| SVM | $\begin{bmatrix} 335 & 11 \\ 13 & 162 \end{bmatrix}$ | 95.39 | 98.70 | 95.87 | 96.26 | 96.82 | 92.57 |
| Log Reg | $\begin{bmatrix} 336 & 10 \\ 13 & 162 \end{bmatrix}$ | 95.58 | 98.62 | 96.25 | 96.28 | 97.11 | 92.57 |
| MLP | $\begin{bmatrix} 338 & 8 \\ 4 & 171 \end{bmatrix}$ | 97.69 | 99.57 | 99.11 | 98.83 | 97.69 | 97.71 |
| CNN | $\begin{bmatrix} 340 & 6 \\ 3 & 172 \end{bmatrix}$ | 98.27 | 99.78 | 99.45 | 99.13 | 98.27 | 98.29 |
| **B. Mean image** | | | | | | | |
| SVM | $\begin{bmatrix} 338 & 8 \\ 9 & 166 \end{bmatrix}$ | 96.73 | 98.78 | 96.30 | 97.41 | 97.69 | 94.86 |
| Log Reg | $\begin{bmatrix} 338 & 8 \\ 12 & 163 \end{bmatrix}$ | 96.16 | 98.74 | 96.95 | 96.57 | 97.69 | 93.14 |
| MLP | $\begin{bmatrix} 338 & 8 \\ 5 & 170 \end{bmatrix}$ | 97.50 | 99.12 | 96.56 | 98.54 | 97.69 | 97.14 |
| CNN | $\begin{bmatrix} 340 & 6 \\ 2 & 173 \end{bmatrix}$ | 98.46 | 99.91 | 99.80 | 99.41 | 98.27 | 98.86 |

Notes: The confusion matrix is represented as $\begin{bmatrix} \text{True positive} & \text{False negative} \\ \text{False positive} & \text{True negative} \end{bmatrix}$. Single slice is the 41st slice image and mean image is the image obtained after taking the mean of all slices numbered from 35 to 48 in the 3D scan. All values are expressed in percentage (%).
Abbreviations: AUC: Area under the region operating characteristic curve; CNN: Convolutional neural network; LogReg: Logistic regression; MLP: Multilayer perceptron; SVM: Support vector machine.

**Table 4. Classification of the SWEDD data using different methods**

| Method | Mean image | | Single slice image | |
|---|---|---|---|---|
| | True negative | False positive | True negative | False positive |
| CNN | 75 | 5 | 76 | 4 |
| LogReg | 73 | 7 | 73 | 7 |
| LinearSVM | 73 | 7 | 73 | 7 |
| MLP | 74 | 6 | 73 | 7 |

Abbreviations: CNN: Convolutional neural network; LogReg: Logistic regression; MLP: Multilayer perceptron; SVM: Support vector machine; SWEDD: Scans without evidence of dopaminergic deficit.

CNNs, in this domain, as these models are capable of learning subtle and complex patterns from training data and making inferences that may even precede clinical judgment.

## 4. Limitations and future work

Recent research shows that deep learning techniques such as the CNN could benefit from the latest advances such as data augmentation, which represents a technique used to increase the training data using information from the available training data.[42] Traditional transformations which include a combination of various affine transformations and using generative adversarial networks[43] are effective ways to augment the data. Label smoothing is another advancement that has shown to improve the performance of deep learning models.[44] In label smoothing, the hard class labels are converted to soft labels. Both data augmentation and label smoothing are methods for regularizing the neural network models, which can help in preventing overfitting and also help networks in converging faster.

Despite the strong performance of the proposed models, it is not without any limitations. First, the model was trained and evaluated solely on the PPMI dataset, and its generalizability to external datasets remains to be validated. Second, the slice selection and averaging strategies used in this study are based on fixed indices that are applicable in the PPMI data, which may not optimally capture relevant features in all subjects in other datasets. Third, while CNNs perform well, they are inherently black-box models, making it difficult to interpret specific feature-driven decisions. Finally, although promising results were achieved in this controlled research setup, further validation is necessary before deployment in clinical environments.

## 5. Conclusion

Accurate and early detection of PD is a challenging clinical problem. The numerous common symptoms shared by this class of Parkinsonism disorders represent a source of misdiagnosis. Accurate identification of degenerative Parkinsonism from other non-degenerative ones is crucial for effective patient treatment and management. In this work, machine learning models that could classify early PD from healthy normal and also SWEDD from PD using SPECT images were developed. These machine learning models exhibit very high performance, of which the CNN model especially achieves the best performance. These predictive models carry enormous potential to be used in a clinical setting and can act as an aid the clinical diagnostic process.

## Acknowledgments

None.

## Funding

PPMI, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and other funding partners include AbbVie, Allergan, Amathus Therapeutics, Avid Radiopharmaceuticals, Biogen Idec, BioLegend, Bristol-Myers Squibb, Celgene, Denali Therapeutics, GE Healthcare, Genentech, GlaxoSmithKline, Eli Lilly and Company, Lundbeck, Merck & Co., Meso Scale Discovery, Pfizer, Piramal, Prevail Therapeutics, Hoffmann-La Roche, Sanofi Genzyme, Servier, Takeda Pharmaceutical Company, Teva, Verily Life Sciences, Voyager Therapeutics, and UCB (Union ChimiqueBelge).

## Conflict of interest

The author declares no competing interests.

## Author contributions

This is a single-authored article.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data

Not applicable.

## Further disclosure

This work was carried out independently by the author. The author is currently employed at Siemens Healthineers, Bangalore, India, however, the views expressed and the work presented here are solely those of the author and do not reflect the views of the company.

## References

1. Jankovic J. Parkinson's disease: Clinical features and diagnosis. *J Neurol Neurosurg Psychiatry*. 2008;79(4):368-376.

   doi: 10.1136/jnnp.2007.131045

2. Fahn S. Description of Parkinson's disease as a clinical syndrome. *Ann N Y Acad Sci*. 2003;991:1-14.

   doi: 10.1111/j.1749-6632.2003.tb07458.x

3. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. *J Neural Transm (Vienna)*. 2017;124(8):901-905.

   doi: 10.1007/s00702-017-1686-y

4. Marek K, Jennings D, Lasch S, *et al*. The Parkinson progression marker initiative (PPMI). *Prog Neurobiol*. 2011;95(4):629-635.

   doi: 10.1016/j.pneurobio.2011.09.005

5. Groveman BR, Orrù CD, Hughson AG, *et al*. Rapid and ultra-sensitive quantitation of disease-associated α-synuclein seeds in brain and cerebrospinal fluid by αSyn RT-QuIC. *Acta Neuropathol Commun*. 2018;6:7.

   doi: 10.1186/s40478-018-0508-2

6. Booij J, Tissingh G, Boer GJ, *et al*. [123I]FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labelling in early and advanced Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 1997;62(2):133-140.

   doi: 10.1136/jnnp.62.2.133

7. Booth TC, Nathan M, Waldman AD, Quigley AM, Schapira AH, Buscombe J. The role of functional dopamine-transporter SPECT imaging in Parkinsonian syndromes, Part 2. *AJNR Am J Neuroradiol*. 2014;36(2):236-244.

   doi: 10.3174/ajnr.a3971

8. Cummings JL, Henchcliffe C, Schaier S, Simuni T, Waxman A, Kemp P. The role of dopaminergic imaging in patients with symptoms of dopaminergic system neurodegeneration. *Brain*. 2011;134(11):3146-3166.

doi: 10.1093/brain/awr177

9. Seibyl J, Jennings D, Grachev I, Coffey C, Marek K. 123-I Ioflupane SPECT measures of Parkinson disease progression in the Parkinson Progression Marker Initiative (PPMI) trial. *J Nucl Med.* 2013;54(suppl 2):190.

10. Beach TG, Adler CH. Importance of low diagnostic accuracy for early Parkinson's disease. *Mov Disord.* 2018;33(10):1551-1554.

doi: 10.1002/mds.27485

11. Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *Neuroimage Clin.* 2017;16:586-594.

doi: 10.1016/j.nicl.2017.09.010

12. Coarelli G, Garcin B, Roze E, Vidailhet M, Degos B. Invalidation of Parkinson's disease diagnosis after years of follow-up based on clinical, radiological and neurophysiological examination. *J Neurol Sci.* 2019;406:116454.

doi: 10.1016/j.jns.2019.116454

13. Catafau AM, Tolosa E, DaTscan Clinically Uncertain Parkinsonian Syndromes Study Group. Impact of dopamine transporter SPECT using 123I-Ioflupane on diagnosis and management of patients with clinically uncertain Parkinsonian syndromes. *Mov Disord*. 2004;19(10):1175-1182.

doi: 10.1002/mds.20112

14. Kupsch AR, Bajaj N, Weiland F, *et al*. Impact of DaTscan SPECT imaging on clinical management, diagnosis, confidence of diagnosis, quality of life, health resource use and safety in patients with clinically uncertain parkinsonian syndromes: A prospective 1-year follow-up of an open-label controlled study. *J Neurol Neurosurg Psychiatry.* 2012;83(6):620-628.

doi: 10.1136/jnnp-2011-301695

15. Marek K, Jennings D, Seibyl JP. Long-term follow-up of patients with scans without evidence of dopaminergic deficit (SWEDD) in the ELLDOPA study. *Neurology.* 2005;64:A274.

16. Benamer TS, Patterson J, Grosset DG, *et al*. Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: The [123I]-FP-CIT study group. *Mov Disord*. 2000;15(3):503-510.

17. Scherfler C, Nocker M. Dopamine transporter SPECT: How to remove subjectivity? *Mov Disord.* 2009;24(Suppl 2):S721-S724.

doi: 10.1002/mds.22590

18. Staff RT, Ahearn TS, Wilson K, *et al*. Shape analysis of 123I-N-omega-fluoropropyl-2-beta-carbomethoxy-3beta-(4-iodophenyl) nortropane single-photon emission computed tomography images in the assessment of patients with parkinsonian syndromes. *Nucl Med Commun.* 2009;30(3):194-201.

doi: 10.1097/MNM.0b013e328314b863

19. Hirschauer TJ, Adeli H, Buford JA. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. *J Med Syst.* 2015;39(11):179.

doi: 10.1007/s10916-015-0353-9

20. Kim DH, Wit H, Thurston M. Artificial intelligence in the diagnosis of Parkinson's disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning. *Nucl Med Commun.* 2018;39(10):887-893.

doi: 10.1097/MNM.0000000000000890

21. Martínez-Murcia FJ, Górriz JM, Ramírez J, Illán IA, Ortiz A, Parkinson's Progression Markers Initiative. Automatic detection of Parkinsonism using significance measures and component analysis in DaTSCAN imaging. *Neurocomputing.* 2014;126:58-70.

doi: 10.1016/j.neucom.2013.01.054

22. Martinez-Murcia FJ, Ortiz A, Górriz JM, *et al*. A 3D Convolutional Neural Network Approach for the Diagnosis of Parkinson's Disease. *In: Natural and Artificial Computation for Biomedicine and Neuroscience (Lecture Notes in Computer Science)*. Springer International Publishing; 2017:324-333.

doi: 10.1007/978-3-319-59740-9_32

23. Oliveira FPM, Castelo-Branco M. Computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines. *J Neural Eng.* 2015;12(2):026008.

doi:10.1088/1741-2560/12/2/026008

24. Ortiz A, Munilla J, Martínez-Ibañez M, Górriz JM, Ramírez J, Salas-Gonzalez D. Parkinson's disease detection using isosurfaces-based features and convolutional neural networks. *Front Neuroinform*. 2019;13:48.

doi: 10.3389/fninf.2019.00048

25. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. *Expert Syst Appl.* 2014;41(8):3333-3342.

doi: 10.1016/j.eswa.2013.11.031

26. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int J Med Inform*. 2016;90:13-21.

doi: 10.1016/j.ijmedinf.2016.03.001

27. Prashanth R, Roy SD, Mandal PK, Ghosh S. High-accuracy classification of Parkinson's disease through shape analysis and surface fitting in 123I-Ioflupane SPECT imaging. *IEEE J Biomed Health Inform*. 2017;21(2):794-802.

doi: 10.1109/JBHI.2016.2547901

28. Segovia F, Gorriz JM, Ramirez J, Alvarez I, Jimenez-Hoyuela JM, Ortega SJ. Improved parkinsonism diagnosis using a partial least squares based approach. *Med Phys*. 2012;39(7):4395-4403.

doi: 10.1118/1.4730289

29. Illan IA, Gorrz JM, Ramirez J, Segovia F, Jimenez-Hoyuela JM, Ortega Lozano SJ. Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging. *Med Phys*. 2012;39(10):5971-5980.

doi: 10.1118/1.4742055

30. Rojas A, Górriz JM, Ramírez J, *et al*. Application of empirical mode decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson disease. *Expert Syst Appl*. 2013;40(7):2756-2766.

doi: 10.1016/j.eswa.2012.11.017

31. Towey DJ, Bain PG, Nijran KS. Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images. *Nucl Med Commun*. 2011;32(8):699-707.

doi: 10.1097/MNM.0b013e328347cd09

32. Huertas-Fernandez I, Garcia-Gomez F, Garcia-Solis D, *et al*. Machine learning models for the differential diagnosis of vascular parkinsonism and Parkinson's disease using [(123) I] FP-CIT SPECT. *Eur J Nucl Med Mol Imaging*. 2015;42(1):112-119.

doi: 10.1007/s00259-014-2882-8

33. Oliveira FP, Faria DB, Costa DC, Castelo-Branco M, Tavares JMR. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [$^{123}$I] FP-CIT SPECT images. *Eur J Nucl Med Mol Imaging*. 2018;45(6):1052-1062.

doi: 10.1007/s00259-017-3918-7

34. Zhang X, Chou J, Liang J, *et al*. Data-driven subtyping of Parkinson's disease using longitudinal clinical records: A cohort study. *Sci Rep*. 2019;9(1):797.

doi: 10.1038/s41598-018-37545-z

35. Marek K, Chowdhury S, Siderowf A, *et al*. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol*. 2018; 5(12):1460-1477.

doi: 10.1002/acn3.644

36. Hoehn MM, Yahr MD. Parkinsonism: Onset, progression and mortality. *Neurology*. 1967;17(5):427-442.

doi: 10.1212/wnl.17.5.427

37. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.

doi: 10.1038/nature14539

38. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297.

doi: 10.1007/BF00994018

39. Lee SI, Lee H, Abbeel P, Ng AY. Efficient l-1 regularized logistic regression. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. Association for the Advancement of Artificial Intelligence; 2006:401-408.

40. Bergstra JS, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In: *Advances in Neural Information Processing Systems 24*. 2011:2546-2554.

41. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929-1958.

42. Perez L, Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv*. Preprint posted online 2017.

doi: 10.48550/arXiv.1712.04621

43. Goodfellow I, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. In: *Advances in Neural Information Processing Systems 27*. 2014:2672-2680.

44. Müller R, Kornblith S, Hinton GE. When does label smoothing help? In: *Advances in Neural Information Processing Systems 32*. 2019:4696-4705.

45. Shiiba T, Takano K, Takaki A, Suwazono S. Dopamine transporter single-photon emission computed tomography-derived radiomics signature for detecting Parkinson's disease. *EJNMMI Res*. 2022;12(1):39.

doi: 10.1186/s13550-022-00910-1

46. Tufail AB, Ma YK, Zhang QN, *et al*. 3D convolutional neural networks-based multiclass classification of Alzheimer's and Parkinson's diseases using PET and SPECT neuroimaging modalities. *Brain Inform*. 2021;8(1):23.

doi: 10.1186/s40708-021-00144-2

47. Majhi B, Kashyap A, Mohanty SS, *et al*. An improved method for diagnosis of Parkinson's disease using deep learning models enhanced with metaheuristic algorithm. *BMC Med Imaging*. 2024;24(1):156.

doi: 10.1186/s12880-024-01335-z

48. Khachnaoui H, Chikhaoui B, Khlifa N, Mabrouk R. Enhanced Parkinson's disease diagnosis through convolutional neural network models applied to spect datscan images. *IEEE Access*. 2023;11:91157-91172.

doi: 10.1109/ACCESS.2023.3308075