3D Convolutional Neural Network for the Diagnosis of Parkinson's Disease with DaTscan SPECT Imaging

Govind Ramesh and Patrick Emedom-Nnamdi

Abstract. Parkinson's Disease (PD) is a progressive neurodegenerative disease caused by dopaminergic deterioration in the striatum. DaTscan SPECT imaging enables the assessment of dopamine transporters in the striatum and is widely used in the clinical diagnosis of PD. In this study, we design and implement a 3D Convolutional Neural Network for the diagnosis of PD using DaTscan SPECT images. Our system achieves high performance while avoiding the influence of observer variability seen with the clinical standard of visual interpretation and other computer aided diagnosis methods. Results of 97.2% accuracy and 97.5% sensitivity for differentiating between PD patients and healthy subjects show greater performance than both visual interpretation and similar computer aided systems. We also show improvement over current results in the classification of a subtype of PD known as scans without evidence of dopaminergic deficit (SWEDD). The performance of our system indicates the possibility of clinical use for the identification of dopaminergic loss in DaTscan SPECT imaging and the diagnosis of PD, as well as validation of SWEDD classification.

1 Introduction

Parkinson's Disease (PD) is a chronic, progressive neurodegenerative disease with motor and nonmotor symptoms, the most common being resting tremor, bradykinesia, and muscular rigidity [1]. As of 2017, PD was estimated to affect 8.5 million people globally, a number that is expected to increase by over 1 million each year [2]. The onset of PD typically develops between the ages of $55-65$, and is estimated to occur in $1\%-2\%$ of people 60 years of age and older. The cause of PD is unknown, but believed to be a combination of environmental and genetic factors. The symptoms of PD are attributed to the progressive loss of dopamine transporters (DaT) in the substantia nigra pars compacta and nigrostriatal pathway, resulting in decreased levels of dopaminergic function in the striatum. There is currently no cure for PD, limiting treatment to relieving symptoms and slowing the progression of the disease. Symptoms generally present themselves after a $50\%-80\%$ DaT deficit [3], highlighting the importance of accurate diagnosis in early stages.

Nuclear imaging is an important step in the diagnosis of PD. Common procedures include ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET), magnetic resonance imaging (MRI), and functional magnetic resonance imaging (fMRI). However, the most popular procedure specific to PD is ^{123}I-ioflupane (DaTscan), a neuro-imaging radiopharmaceutical tracer that binds to dopamine transporters in the striatum and can be detected with Single Photon Emission Computed Tomography (SPECT). DaTscan SPECT imaging reveals dopaminergic deficit in the striatum characteristic of PD, making it useful for diagnosing PD in a clinical

setting. Visual interpretation is the approved method for PD diagnosis with DaTscan SPECT imaging, but even with guidelines [4] it has shown to be subject to intra- and inter-observer variability [5, 6, 7, 8].

A subtype of PD patients known as SWEDD (scans without evidence of dopaminergic deficit) do not show signs of loss of dopamine transporters or general abnormalities with DaTscan SPECT imaging. Approximately 10% – 15% of PD patients are classified as SWEDD [9]. There is some controversy over the accuracy of the term SWEDD with respect to being a subtype of PD, since multiple studies show that most SWEDD patients are clinically misdiagnosed while only a few show evidence of being PD without abnormal DaTscan SPECT [10]. The effects of the uncertainty of SWEDD classification on this study are discussed in Section 4. There have been multiple attempts to diagnose or classify SWEDD patients using various methods, however, none have been successful thus far, either as a result of the similarities between SWEDD and healthy DaTscan SPECT images, SWEDD patients displaying symptoms of other diseases, or misdiagnoses in the SWEDD cohorts [10, 11, 12, 13, 14].

Computer aided diagnosis has become increasingly common in the medical field, and multiple attempts have been made in the context of PD with DaTscan SPECT. The most popular approach involves feature extraction by image segmentation and a Support Vector Machine (SVM) classifier. Bhalchandra et al. [15] was able to obtain high accuracy with this method, but also implemented striatal binding ratios as features, which introduces observer variability as described in Section 2.4. Augimeri et al. [16] also achieved high accuracy with an SVM with a dataset of 43 subjects, including 31 late-stage PD patients. Illán et al. [17] implemented an end-to-end network with fully automated feature extraction and classification by SVM, however there was a significant drop in accuracy compared to others. Choi et al. [18] and Martinez-Murcia et al. [12] used methods similar to the one to be proposed in this study, but with significantly less data. We will use the results of these two studies as the baseline for our model, i.e., 96% accuracy reported by [18] and 95.5% accuracy reported by [12].

Convolutional Neural Networks (CNNs; Sec. 2.3) have become increasingly popular in the medical field due to their effectiveness in image analysis. They have been used throughout literature for tasks such as image segmentation, disease classification, abnormality detection, and computer aided diagnosis [19, 20]. For our purposes, we will use the TensorFlow backend to adapt the standard CNN to fit three-dimensional data, specifically DaTscan SPECT images. With this, we hope to provide an end-to-end method for diagnosis without the need for feature extraction or visual interpretation in order to avoid observer variability.

In this study, we will implement and evaluate a 3D Convolutional Neural Network for the diagnosis of PD with DaTscan SPECT images. In Section 2, we present the data and preprocessing, the 3D Convolutional Neural Network framework, and the study design. In Section 3, we report the results of the experiments and an evaluation of our model. In Section 4, we discuss our findings and their implications, and in Section 5 we present conclusions from the study and propose future experimentation.

2 Materials and Methods

2.1 Dataset

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. Standard acquisition protocol for the PPMI database images subjects 4 ± 0.5 hours after the injection of 111–185 MBq of DaTscan. Raw SPECT data are acquired onto a 128 × 128 matrix stepping each 3 degrees for a total of 120 (or 4 degrees for a total of 90) projections into two 20% symmetric photopeak windows centered on 159 KeV and 122 KeV with a total scan duration of approximately 30–45 minutes [21], then reconstructed to final dimensions of 91 × 109 × 91.

In this study, a total of 871 three-dimensional DaTscan SPECT images were collected from the database, including 602 patients with PD, 193 normal controls (NCs), and 76 patients with SWEDD. The subset of PD patients consists of 365 males and 237 females, with an average age of 62.0 ± 10.0. The subset of NCs consists of 128 males and 65 females, with an average age of 60.9 ± 11.2, and the subset of SWEDD patients consists of 47 males and 29 females, with an average age of 60.4 ± 10.6. Further demographic information is reported in Table 1. The age (to account for DaT degeneration over time [21, 22]) and sex of each subject are included as additional features in the training and evaluation of the model. We also experiment with the implementation of striatal dopamine transporter binding ratios (SBRs) as added features, specifically at the left and right putamen and caudate. Therefore, the 3D DaTscan SPECT image, age, sex, and SBRs of each subject from the PPMI database are the features to be utilized in the model presented in Section 2.3.

	Count			Age			Striatal Binding Ratios				Race				Categorical Hoehn & Yahr	MDS-UPDRS Part III Score
Class	Total	Male	Female	Total	Male	Female	Right Caudate	Left Caudate	Right Putamen	Left Putamen	White	Black	Asian	Other		
PD	602	365	237	62.0 ± 10.0	62.2 ± 10.0	61.7 ± 9.9	1.91 ± 0.64	1.91 ± 0.61	0.82 ± 0.41	0.80 ± 0.38	554	9	12	27	1.6 ± 0.5	21.0 ± 8.8
NC	193	128	65	60.9 ± 11.2	62.0 ± 10.7	58.9 ± 12.0	2.94 ± 0.46	2.97 ± 0.61	2.13 ± 0.57	2.13 ± 0.56	179	9	1	4	–	–
SWEDD	76	47	29	60.4 ± 10.6	61.2 ± 9.9	58.7 ± 11.6	2.82 ± 0.59	2.83 ± 0.57	2.06 ± 0.52	2.03 ± 0.50	72	1	1	2	1.4 ± 0.5	14.5 ± 9.6

Table 1. Demographic and clinical data of subjects.

2.2 Preprocessing

2.2.1 Volume Reduction

DaTscan binds to dopamine transporters in the striatum, revealing the potential deficiency of dopamine-producing neurons seen in PD. Therefore, the full 3D volume of the DaTscan SPECT image is not relevant for diagnosis of PD. We perform this volume reduction through

thresholding, resulting in a selection of relevant features and a reduction of noise created by the inclusion of unimportant features.

The thresholding process used in this study is one that has previously been used in similar contexts [12, 25], in which a binary mask is created from the average of each 3D image in the dataset by an intensity threshold. The intensity threshold is a percentage of the maximum value of the average 3D image such that any value above the threshold will compose the binary contour mask. From the binary contour mask, the maximum and minimum indices of the mask for each dimension were used to create a rectangular 3D mask, which was then applied to every subject's DaTscan SPECT image. Fig. 1 demonstrates the process with an intensity threshold of 0.35 and shows the final output for slice $z = 39$ of a DaTscan SPECT image. Performing the thresholding with an intensity threshold of 0.35 resulted in a reduction of each DaTscan SPECT from dimensions $(91, 109, 91, 1)$ to $(39, 23, 50, 1)$, a reduction factor of ≈ 20.

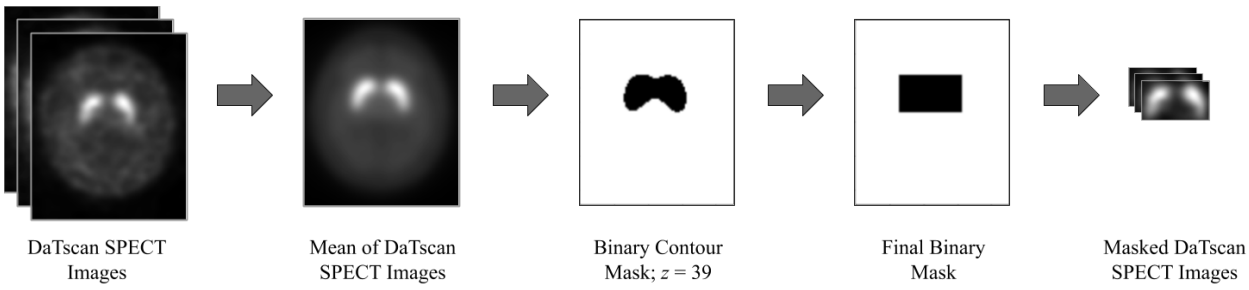

Fig. 1. Thresholding process with intensity threshold of 0.35

2.2.2 Mean Subtraction

In order to zero-center the dataset, we perform a common type of preprocessing, taking the mean of all 3D images and subtracting it from each individual image. As a result, each voxel position in the dataset is centered around zero. With mean subtraction, the model is able to optimize more efficiently without imbalanced gradients throughout the training data. We do not perform normalization by dividing each dimension by the standard deviation in the preprocessing stage, instead choosing to utilize Batch Normalization layers as part of the model framework presented in Section 2.3.4.

2.2.3 Imaging Data Augmentation

We also perform data augmentation on the DaTscan SPECT images to increase the size of the training set and, as a result, increase the generalization capabilities and prevent overfitting. This is done by reflecting each image horizontally, i.e., over the sagittal plane, effectively doubling the size of the training set (Fig. 2).

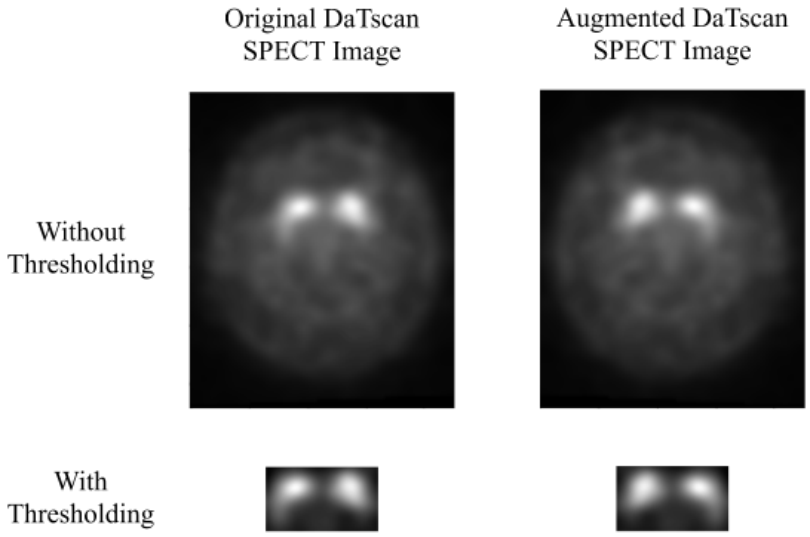

Fig. 2. Data augmentation for slice $z = 39$ of a DaTscan SPECT image

2.3 3D Convolutional Network Architecture

Convolutional Neural Networks (CNNs) are a class of Artificial Neural Networks (ANNs) primarily utilized for problems involving image classification. Rather than using each pixel or voxel value as a feature as would happen with a standard fully-connected multilayer perceptron, CNNs connect each neuron to a small section of the prior layer, drastically reducing the number of parameters and, as a result, the amount of overfitting and computational power. This process is done through the use of learnable filters and convolutions, hence the name CNN, and is elaborated on in Section 2.3.1.

A standard CNN architecture consists of convolutional layers, pooling layers, and dense layers. In general, the input images go through a series of convolutional layers and a pooling of the activations, followed by dense layers which make the classification. There are many variations of CNNs currently in practice [26, 27, 28, 29]. The learning process is done through backward propagation of errors, referred to as backpropagation, in which the gradient of the error function is calculated with respect to the network's weights. Through gradient descent, the error is minimized by updating the weights.

For the classification of Parkinson's Disease, we designed a 3D CNN framework, PD-CNN, the architecture of which is shown in Fig. 3. The DaTscan SPECT images of dimensions $(39, 23, 50, 1)$ are the inputs, as well as the added features (age,sex, and SBRs) for experiments 1B and 1C (Sec. 2.4). Layers $L_{1,4,7}$ are 3D convolutional layers (Sec. 2.3.1), each with filter size of $3 \times 3 \times 3$ and stride of 1 voxel. L_1 generates 32 filter maps, L_4 generates 64, and L_7 generates 128. Each convolutional layer is followed by Batch Normalization layers ($L_{2,5,8}$; Sec. 2.3.4) and Rectified Linear Unit (ReLU) activations (2.3.2). The activations are fed through max

pooling layers ($L_{3,6,9}$; Sec. 2.3.3), each with a pool size of $2 \times 2 \times 2$ and a stride of 2 voxels. Zero-padding was applied to the convolutional and pooling layers ($L_{1,3,4,6,7,9}$). The N features extracted from the final max pooling layer (L_9) are flattened into a $1 \times N$ vector and connected to the first dense layer (L_{10}). If called for, the added features are concatenated to the convolution-extracted features (Sec. 2.3.1) before being routed through the dense layers (Sec. 2.3.5). Dense layers $L_{10,11}$ consist of 256 neurons each, and L_{12} has 128 neurons. He weight initialization [30] is used for convolutional and dense layers ($L_{1,4,7,10,11,12}$). Each dense layer uses the ReLU activation function. The output layer uses the softmax activation function for the prediction and has as many nodes as classes, i.e., two for PD/NC classification and three for PD/NC/SWEDD classification.

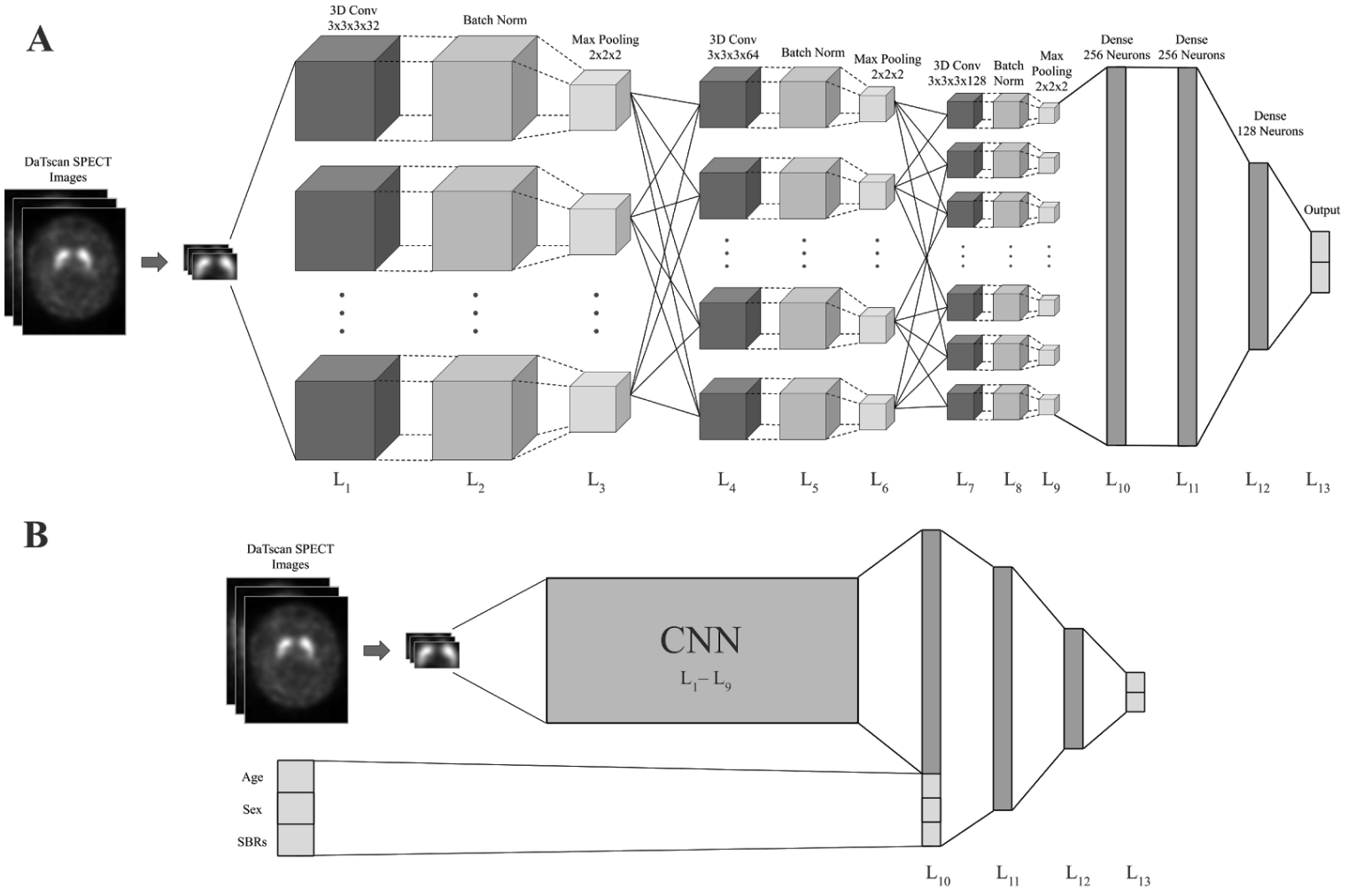

Fig. 3. Architecture of 3D Convolutional Neural Network (PD-CNN). (A) Full PD-CNN framework without added features. (B) Simplified PD-CNN framework demonstrating implementation of added features.

For the training process, a validation split of 20% is used to monitor the performance of the network. We use the Adam optimization technique [31] for gradient descent with the cross-entropy loss between the predicted diagnosis and true diagnosis of subjects. Hyperparameters such as learning rate, batch size, epochs, and the number of neurons were chosen through a grid search. The learning rate was set to 0.00001, with batch size 32 and number of epochs at 25. However, training often stopped before reaching 25 epochs due to the use of early stopping (Sec. 2.3.6).

2.3.1 3D Convolutional Layer

The convolutional layer is central to the function of CNNs, using a set of learnable filters that function similar to weights in a multilayer perceptron. These filters are relatively small compared to the image. During the forward pass, each filter shifts, or convolves, along the dimensions of the input image, computing dot products in which each element in the filter is multiplied by its respective value in the filter-sized patch of the input image and summed. These dot products are arranged spatially, producing an activation map of the image showing where the filter was detected in the input image. Through this process, convolutional layers introduce translation invariance, since the shifting of the filter across the whole input enables the identification of that feature anywhere in the image. The values within each filter are learned and updated during backpropagation with gradient descent to minimize the loss.

In a 2D convolutional layer, the activation maps produced are 2D, but are computed multiple times for the total number of filters in the layer (a hyperparameter) and stacked to create a 3D output volume. Therefore, extending to a 3D convolutional layer, a number of 3D filters convolve along the dimensions of the input volume, creating 3D activation maps which are stacked to create a 4D volume, in which the 4th dimension is the number of filters in the layer.

Convolutional layers can be applied to not only the input image, but also the outputs of other convolutional layers (following activation and pooling operations). By stacking multiple convolutional layers, deeper layers are able to detect more complex features. Initial convolutional layers can learn simple features such as lines from the input image, while subsequent layers can extract combinations of these simple features to construct more complex filters specific to the task.

Convolutional layers use four main hyperparameters: filter size, number of filters (n), stride, and zero-padding. Filter size in a 2D CNN is generally set to be a square with size of 3, 5, or 7. In this study, we will use filter dimensions of $3 \times 3 \times 3$ for each 3D convolutional layer. We use three convolutional layers, progressively increasing the number of filters for each subsequent layer, resulting in $n_1 = 32$, $n_2 = 64$, and $n_3 = 128$ filters. Stride determines how many pixels or voxels the filter shifts over for each convolution, and is usually set to 1 since any higher value reduces the amount of overlap and downsamples the input, an operation that is already performed with pooling (Sec. 2.3.3). Zero-padding adds a border of zeros around the input, enabling the voxels on the edges to be in the center of the filter during convolution and creating

an output activation map with the same dimensions as the input. We use zero-padding in each convolutional layer.

2.3.2 Activation

Activation functions are applied to every output function in ANNs. Activations enable the model to learn complex patterns in multidimensional data by introducing non-linearity. There are many activation functions, but the most popular in recent years is the Rectified Linear Unit (ReLU), especially for CNNs [32]. ReLU functions avoid the problem of saturation seen with Tanh and Sigmoid activations, which effectively kills the gradient and the learning process. Hence, ReLU has been shown to converge six times faster than other activation functions [27]. ReLU is computed by the function

$$f(x) = max(0, x) \tag{1}$$

which proves to be computationally efficient during backpropagation compared to other functions using expensive operations. In this study, we will use the ReLU activation function with every convolutional and dense layer.

2.3.3 Max Pooling

Pooling layers serve as a downsampling technique used between convolutional layers. Pooling reduces the amount of parameters which decreases the amount of computation power needed and limits overfitting to the data. In practice, max pooling layers are used, in which the max operation is applied to the pool. Traditional max pooling layers use filters/pool size of dimensions 2×2 and a stride of 2 for the downsampling operation, effectively discarding 75% of the activations. For the 3D CNN implementation, a filter size of $2 \times 2 \times 2$ will be used in each max pooling layer following the convolutional layers.

2.3.4 Batch Normalization

Batch Normalization is a technique developed to address the problem of internal covariate shift, in which training of deep learning models becomes less efficient due to the need to adapt to new input distributions as parameters are changed. Batch Normalization makes normalization a part of the model architecture, applying the operation for each training mini-batch and conforming the activations at each point in the network to a unit gaussian distribution [33]. There is some debate as to whether the effectiveness of Batch Normalization is a result of reducing internal covariate shift or if there are more significant impacts, such as the smoothening of the optimization landscape, which create more stable and traversable gradients [34]. Nevertheless, the use of mini-batch statistics for normalization has been proven to improve the training process, not only in the efficiency of convergence during training, but also through regularizing properties as a result of the influence of the whole mini-batch on each individual training

element, and is now widely used in practice. In the original paper [33], Batch Normalization was designed to be used in between the fully connected or convolutional layers and their activations, but there is some confusion as to whether better results are achieved with Batch Normalization used after the non-linearity. We achieved slightly better results with the use of Batch Normalization before the activation functions. In this study, we will use Batch Normalization following each convolutional layer.

2.3.5 Dense Layers + Added Features

Dense layers come after the series of convolutional layers, mapping the outputs of those convolutions to a final classification using one or more hidden layers. These layers, also known as fully-connected layers, consist of neurons with weights and biases as seen in a multilayer perceptron, with each neuron in a layer routed to every neuron in the next layer. Each neuron uses its inputs, weights, bias, and activation function to pass information to the next layer, where the output y is calculated with the equation

$$y = f\left(\sum_{i=0}^{n} x_i w_i + b\right) \tag{2}$$

for inputs $x_{1:n}$, weights $w_{1:n}$, bias parameter b, and activation function f. The weights and biases are then updated with gradient descent during backpropagation. At a higher level, the convolutional layers can be thought of as a feature extraction system, while the dense layers analyze those features to model the underlying relationships.

In this study, we will use three dense layers, two with 256 neurons, followed by a 128-neuron layer, each using the ReLU activation function. These layers are followed by the output layer, which has as many neurons as classes, and uses the softmax activation function. We will also implement the added features (age, sex, and SBRs) at this stage by concatenating them to the features extracted by the convolutional layers before feeding them into the dense layers.

2.3.6 Early Stopping

An effective technique to combat overfitting during training is early stopping. It works by evaluating training metrics such as the loss or accuracy of the validation set at the end of each epoch [35]. Overfitting is generally characterized by the stagnation or decline of validation metrics. Early stopping ends the training process once it notices evidence of overfitting by using two hyperparameters, patience and minimum Δ, while monitoring a validation metric. The patience parameter determines the number of epochs without improvement before training is stopped. Minimum Δ is used to set the value the validation metric must improve by over the epochs to be considered an improvement and continue training. Once training is stopped, model weights are reverted back to the epoch from when the overfitting was first detected. We will use

early stopping to monitor the validation loss during training, with a minimum Δ of 0.01 and a patience parameter of 3.

2.4 Study Design

PD-CNN was trained and evaluated using 10-fold cross validation, with a validation split of 20%. From each fold in the evaluation, metrics such as accuracy, sensitivity, and specificity are generated. For each experiment, a final confusion matrix will also be reported. PD-CNN will be tested with two experiments, each with different parts as described below.

Experiment 1: PD/NC Classification
This set of experiments will evaluate the ability of PD-CNN to differentiate PD patients from NCs using the architecture outlined in Section 2.3.

Part A. Diagnosis with only DaTscan SPECT images for each subject as input to PD-CNN.
Part B. Diagnosis with DaTscan SPECT images and the age and sex for each subject as inputs to PD-CNN.
Part C. Diagnosis with DaTscan SPECT images, age, sex, and striatal binding ratios (SBRs) for each subject as inputs to PD-CNN.

Note: The addition of Part 3 as opposed to the inclusion of all added features in one experiment is due to SBRs being a semi-quantitative measurement, and therefore subject to intra- and inter-observer variability [36]. There are also multiple methods for the calculation of SBRs [36, 37, 38, 39], some resulting in SBR values two to three times greater than others, which would affect the prediction capabilities of PD-CNN in a clinical setting. The use of a CNN for diagnosis eliminates the need for human interpretation and observer variability, which the inclusion of SBRs as added features reintroduces.

Experiment 2: SWEDD Classification
This set of experiments will evaluate the ability of PD-CNN to accurately classify SWEDD patients.

Part A. Diagnosis with SWEDD subjects as a third class using DaTscan SPECT images and the age and sex for each subject as inputs to PD-CNN. With the prediction of more than two classes, sensitivity and specificity cannot be calculated, so only accuracy and the full confusion matrix will be reported.
Part B. SWEDD subjects will be used as a test set on PD-CNN trained with PD and NC subjects, and considered clinically PD for the prediction. Only accuracy will be reported.

3 Results

Experiment 1: PD/NC Classification

Part A. This experiment demonstrates the ability of PD-CNN to differentiate PD patients from NCs using only DaTscan SPECT images as input. We achieved high performance in the diagnosis of PD with this experiment, obtaining 0.966 ± 0.016 overall accuracy, 0.973 ± 0.011 sensitivity, and 0.943 ± 0.036 specificity. The confusion matrix for this experiment is shown in Fig. 4(a).

Part B. This experiment builds upon the last one, adding the age and sex of each subject as features to PD-CNN in addition to the DaTscan SPECT images to classify PD patients and NCs. We obtained slightly improved results with the diagnosis of PD in comparison to Experiment 1A with the implementation of the added features, resulting in 0.972 ± 0.015 overall accuracy, 0.975 ± 0.008 sensitivity, and 0.965 ± 0.039 specificity. The confusion matrix for this experiment is shown in Fig. 4(b).

Part C. This experiment uses the striatal binding ratios (SBRs) from the left and right putamen and caudate for each subject as additional features to PD-CNN, as well as the DaTscan SPECT images, age, and sex as seen in the previous experiments. We still maintained high performance in this experiment, but achieved slightly worse results than seen with Experiment 1B. We obtained an overall accuracy of 0.970 ± 0.019, sensitivity of 0.975 ± 0.014, and specificity of 0.954 ± 0.042. The confusion matrix for this experiment is shown in Fig. 4(c).

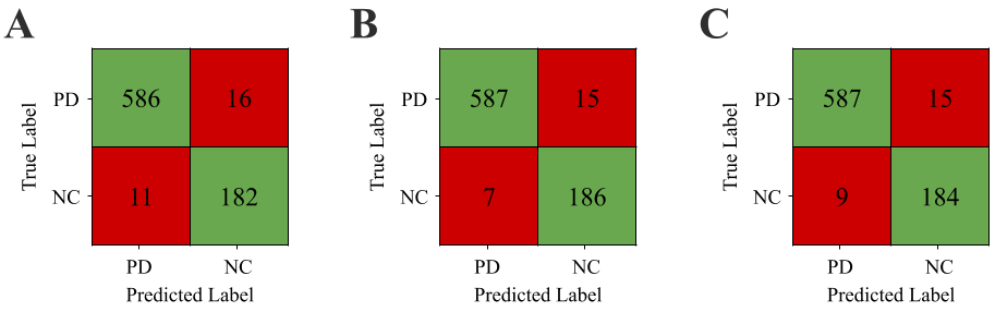

Fig. 4. Confusion matrices for Experiment 1

Experiment 2: SWEDD Classification

Part A. This experiment demonstrates the ability of PD-CNN to accurately classify SWEDD patients by implementing SWEDDs as a third class in addition to PD patients and NCs using the DaTscan SPECT images, age, and sex of each subject as inputs to PD-CNN. We achieved significantly lower performance in comparison to Experiment 1, obtaining 0.885 ± 0.031 accuracy and the resulting confusion matrix:

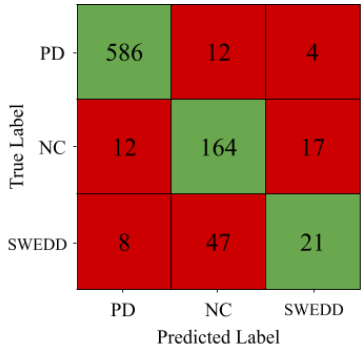

Fig. 5. Confusion matrix for Experiment 2A

As seen in the confusion matrix, only 21 out of the 76 SWEDD subjects (27.6%) in the dataset were accurately classified. Most of the SWEDD subjects (47/76, or 61.8%) were interpreted by PD-CNN as having no dopaminergic deficit and incorrectly classified as NCs. Additionally, 17 out of 193 NCs (8.8%) were misclassified as SWEDD, more than the amount misclassified as PD (12/193, or 6.2%).

Part B. This experiment uses SWEDD subjects as a test set on PD-CNN trained with PD and NC subjects. The SWEDD subjects were considered clinically PD for the prediction. We obtained 21.1% accuracy with this experiment. All SWEDD subjects show parkinsonian symptoms but do not show abnormalities in DaTscan SPECT from visual interpretation. PD-CNN classified 16 out of the 76 SWEDD subjects as PD, or showing abnormal DaT.

4 Discussion

In Experiment 1, we showed that our 3D Convolutional Neural Network, PD-CNN, was able to accurately interpret DaTscan SPECT images to diagnose PD from controls (NCs). PD-CNN showed better performance than the clinical standard of visual interpretation for the diagnosis of PD [40]. We also achieved higher accuracy and sensitivity than both Martinez-Murcia et al. [12] and Choi et al. [18] using similar methods. A comparison of these results is shown in Table 2.

	Visual Interpretation [40]	Martinez-Murcia et al. [12]	Choi et al. [18]	PD-CNN
Accuracy	93.8%	95.5%	96.0%	97.2%
Sensitivity	94.2%	96.2%	94.2%	97.5%
Specificity	91.9%	94.6%	100%	96.5%

Table 2. Accuracy of PD-CNN in comparison to other methods of diagnosis.

In Experiment 1C, we implemented Striatal Binding Ratios (SBRs) as added features to PD-CNN, but observed a slight decrease in accuracy. This may be a result of SBRs being a semi-quantitative estimate of the DaT in the striatum—a feature that is already extracted from the DaTscan SPECT images in the convolutional layers. This becomes evident when we achieve up to 94% accuracy on a simple neural network trained with only SBRs. The inclusion of SBRs in addition to the DaTscan SPECT images as input to PD-CNN does not add any important information, instead causing the model to overfit slightly and, as a result, decreasing the generalization ability of the network. In addition, one of the main benefits of PD-CNN is its ability to accurately diagnose PD by directly observing the DaTscan SPECT images, rather than with complicated feature extraction seen in other computer aided diagnosis methods. This enables PD-CNN to avoid the problem of observer variability seen with visual interpretation and feature extraction, which the inclusion of SBRs would reintroduce, due to it being a semi-quantitative measurement.

In Experiment 2, we attempted to classify SWEDD patients. Experiment 2A used SWEDD as its own class and correctly diagnosed 21/76 SWEDD patients. Additionally, 47/76 SWEDD patients were misclassified as NC and 17/193 NCs were misclassified as SWEDD, demonstrating the similarities between the two classes and the difficulty in distinguishing between them. While we did not achieve high performance (27.6%), this is still an improvement in comparison to other attempts to classify SWEDD, e.g., 6.3% accuracy reported by Martinez-Murcia et al. [12]. We obtained worse results when testing the SWEDD patients for dopaminergic deficit by training PD-CNN on PD patients and NCs in Experiment 2B. However, once again, our reported 16/76 correctly classified SWEDD patients (21.1%) is still an improvement over current attempts such as Choi et al. [18] with 6/77, or 7.8%, correct diagnoses. It should be noted that the composition of our SWEDD dataset is questionable in terms of misdiagnosed cases. Choi et al. [18] used the same SWEDD patients from the PPMI database as used in this study and found that, of 56/77 patients that had a follow-up clinical diagnosis after 2 years, only 16 were still classified as PD while the other 40 showed evidence of alternative diseases, including 23 with nonparkinsonian tremor. Out of the 16 classified as PD in the follow-up, 13 had a follow-up DaTscan SPECT. Visual interpretation of the DaTscan SPECT found 6 with evidence of dopaminergic deficit, while 7 were still classified as SWEDD. These findings may explain the low performance observed in Experiment 2. Since a majority of the SWEDD patients in this dataset were misdiagnosed originally, it would be difficult to correctly classify them due to the DaTscan SPECT of nonparkinsonian diseases being indistinguishable from NCs. Additionally, the result of Experiment 2B points to PD cases being misdiagnosed with visual interpretation, since 16 SWEDD subjects were classified by PD-CNN as PD, or showing DaT loss.

In this study, we primarily focused on differentiating between NCs and PD patients. However, in a clinical setting, DaTscan SPECT is generally used to distinguish PD from nonparkinsonian disorders with similar symptoms. Since PD-CNN achieved high accuracy in

differentiating NCs from PD, it would most likely be applicable in discriminating nonparkinsonian disorders from PD, since nonparkinsonian disorders do not show dopaminergic deficit, and therefore would present a DaTscan SPECT similar to the NCs. However, further study to evaluate this application of PD-CNN would be necessary before drawing conclusions.

5 Conclusion

In this study we designed and implemented a 3D Convolutional Neural Network, PD-CNN, for the diagnosis of Parkinson's Disease (PD) using DaTscan SPECT images. The high performance obtained by PD-CNN demonstrates the functionality of 3D Convolutional Neural Networks in avoiding observer variability while still achieving high classification accuracy. Our results show significant improvement over the current clinical standard of visual interpretation and similar implementations of CNNs for the diagnosis of PD. The decreased performance for the classification of SWEDD patients may have been impacted by the similarities between the DaTscan SPECT of SWEDD and NCs as well as misdiagnosed cases within the dataset. Future work will include the application of this methodology to differentiating between other diseases and refining the diagnosis of SWEDD patients.

Acknowledgments

PPMI – a public-private partnership (http://www.ppmi-info.org/) – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Avid Radiopharmaceuticals, Biogen Idec, Bristol-Myers Squibb, Covance, Eli Lilly & Co, F Hoffmann-La Roche, GE Healthcare, Genentech, GlaxoSmithKline, Lundbeck, Merck, MesoScale, Piramal, Pfizer, and UCB.

References

[1] P. Rizek, N. Kumar, and M. S. Jog, "An update on the diagnosis and treatment of Parkinson disease," *Canadian Medical Association Journal*, vol. 188, no. 16, pp. 1157–1165, 2016.

[2] GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 392, no. 10159, pp. 1789–1858, 2018.

[3] G. DeMaagd and A. Philip, "Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis," *Pharmacy and Therapeutics*, vol. 40, no. 8, pp. 504–510, 532, 2015.

[4] J. Darcourt, J. Booij, K. Tatsch, A. Varrone, T. V. Borght, Ö. L. Kapucu, K. Någren, F. Nobili, Z. Walker, and K. V. Laere, "EANM procedure guidelines for brain neurotransmission SPECT using 123I-labelled dopamine transporter ligands, version 2," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 2, pp. 443–450, 2009.

[5] M. C. Tondeur, A.-S. Hambye, S. Dethy, and H. R. Ham, "Interobserver reproducibility of the interpretation of I-123 FP-CIT single-photon emission computed tomography," *Nuclear Medicine Communications*, vol. 31, no. 8, pp. 717–725, 2010.

[6] N. Papathanasiou, P. Rondogianni, P. Chroni, M. Themistocleous, E. Boviatsis, X. Pedeli, D. Sakas, and I. Datseris, "Interobserver variability, and visual and quantitative parameters of 123I-FP-CIT SPECT (DaTSCAN) studies," *Annals of Nuclear Medicine*, vol. 26, no. 3, pp. 234–240, 2012.

[7] I. Mckeith, J. O'brien, Z. Walker, K. Tatsch, J. Booij, J. Darcourt, A. Padovani, R. Giubbini, U. Bonuccelli, D. Volterrani, C. Holmes, P. Kemp, N. Tabet, I. Meyer, and C. Reininger, "Sensitivity and specificity of dopamine transporter imaging with 123I-FP-CIT SPECT in dementia with Lewy bodies: a phase III, multicentre study," *The Lancet Neurology*, vol. 6, no. 4, pp. 305–313, 2007.

[8] I. Gayed, U. Joseph, M. Fanous, D. Wan, M. Schiess, W. Ondo, and K.-S. Won, "The Impact of DaTscan in the Diagnosis of Parkinson Disease," *Clinical Nuclear Medicine*, vol. 40, no. 5, pp. 390–393, 2015.

[9] K. Marek, J. Seibyl, S. Eberly, D. Oakes, I. Shoulson, A. E. Lang, C. Hyson, and D. Jennings, "Longitudinal follow-up of SWEDD subjects in the PRECEPT Study," *Neurology*, vol. 82, no. 20, pp. 1791–1797, 2014.

[10] R. Erro, S. A. Schneider, M. Stamelou, N. P. Quinn, and K. P. Bhatia, "What do patients with scans without evidence of dopaminergic deficit (SWEDD) have? New evidence and continuing controversies," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 87, no. 3, pp. 319–323, 2015.

[11] V. L. Marshall, C. B. Reininger, M. Marquardt, J. Patterson, D. M. Hadley, W. H. Oertel, H. T. Benamer, P. Kemp, D. Burn, E. Tolosa, J. Kulisevsky, L. Cunha, D. Costa, J. Booij, K. Tatsch, K. R. Chaudhuri, G. Ulm, O. Pogarell, H. Höffken, A. Gerstner, and D. G. Grosset, "Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: A 3-year European multicenter study with repeat [123I]FP-CIT SPECT," *Movement Disorders*, vol. 24, no. 4, pp. 500–508, 2009.

[12] F. J. Martinez-Murcia, A. Ortiz, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, and I. A. Illán, "A 3D Convolutional Neural Network Approach for the Diagnosis of Parkinson's Disease," *Natural and Artificial Computation for Biomedicine and Neuroscience Lecture Notes in Computer Science*, pp. 324–333, 2017.

[13] P. Schwingenschuh, D. Ruge, M. J. Edwards, C. Terranova, P. Katschnig, F. Carrillo, L. Silveira-Moriyama, S. A. Schneider, G. Kägi, F. J. Palomar, P. Talelli, J. Dickson, A. J. Lees, N. Quinn, P. Mir, J. C. Rothwell, and K. P. Bhatia, "Distinguishing SWEDDs patients with asymmetric resting tremor from Parkinson's disease: A clinical and electrophysiological study," *Movement Disorders*, vol. 25, no. 5, pp. 560–569, 2010.

[14] S. A. Schneider, M. J. Edwards, P. Mir, C. Cordivari, J. Hooker, J. Dickson, N. Quinn, and K. P. Bhatia, "Patients with adult-onset dystonic tremor resembling parkinsonian tremor have scans without evidence of dopaminergic deficit (SWEDDs)," *Movement Disorders*, vol. 22, no. 15, pp. 2210–2215, 2007.

[15] N. A. Bhalchandra, R. Prashanth, S. D. Roy, and S. Noronha, "Early detection of Parkinson's disease through shape based features from ^{123}I-Ioflupane SPECT imaging," *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015.

[16] A. Augimeri, A. Cherubini, G. L. Cascini, D. Galea, M. E. Caligiuri, G. Barbagallo, G. Arabia, and A. Quattrone, "CADA—computer-aided DaTSCAN analysis," *EJNMMI Physics*, vol. 3, no. 1, 2016.

[17] I. A. Illán, J. M. Górriz, J. Ramírez, F. Segovia, J. M. Jiménez-Hoyuela, and S. J. O. Lozano, "Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging," *Medical Physics*, vol. 39, no. 10, pp. 5971–5980, 2012.

[18] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, "Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging," *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.

[19] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," *Journal of Medical Systems*, vol. 42, no. 11, 2018.

[20] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van de Laak, B. V. Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[21] *PPMI Imaging Technical Operations Manual*, 2nd ed., The PARKINSON PROGRESSION MARKER INITIATIVE, 2010, pp. 13–15.

[22] W. Pirker, S. Asenbaum, M. Hank, S. Kandlhofer, J. Tauscher, M. Willeit, A. Neumeister,
N. Praschak-Rieder, P. Angelberger, and T. Brücke, "Imaging Serotonin and Dopamine Transporters with ^{123}I-β-CIT SPECT: Binding Kinetics and Effects of Normal Aging," *The Journal of Nuclear Medicine*, vol. 41, pp. 36–44, 2000.

[23] C. H. van Dyck, J. P. Seibyl, R. T. Malison, M. Laruelle, E. Wallace, S. S. Zoghbi, Y. Zea-Ponce, R. M. Baldwin, D. S. Charney, and P. B. Hoffer, "Age-Related Decline in Striatal Dopamine Transporter Binding with Iodine-123-β-CIT SPECT," *The Journal of Nuclear Medicine*, vol. 36, no.7, pp. 1175–1181, 1995.

[24] U. S. Food and Drug Administration Center for Drug Evaluation and Research, *DaTscan Summary Review*, Jan. 2011. [Online]. Available: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022454sOrig1s000SumR.pdf

[25] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, M. Moreno-Caballero, and M. Gómez-Río, "Parametrization of textural patterns in123I-ioflupane imaging for the automatic detection of Parkinsonism," *Medical Physics*, vol. 41, no. 1, p. 012502, 2013.

[26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.

[31] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv: 1412.6980v9 [cs.LG],* 2017.

[32] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," *arXiv: 1505.00853v2 [cs.LG]*, 2015.

[33] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *arXiv: 1502.03167 [cs.LG]*, 2015.

[34] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, “How Does Batch Normalization Help Optimization?,” *arXiv: 1805.11604v5 [stat.ML]*, 2019.

[35] R. Caruana, S. Lawrence and C. L. Giles, "Overfitting in neural nets: Backpropagation conjugate gradient and early stopping,” *Advances in Neural Information Processing Systems*, pp. 402-408, 2001.

[36] W. Koch, P. E. Radau, C. Hamann, and K. Tatsch, “Clinical Testing of an Optimized Software Solution for an Automated, Observer-Independent Evaluation of Dopamine Transporter SPECT Studies,” *Journal of Nuclear Medicine*, vol. 64, no.7, pp. 1109–1118, 2005.

[37] M. G. M. Rahman, M. M. Islam, T. Tsujikawa, Y. Kiyono, and H. Okazawa, “Count-based method for specific binding ratio calculation in [I-123]FP-CIT SPECT analysis,” *Annals of Nuclear Medicine*, vol. 33, no. 1, pp. 14–21, 2018.

[38] P. Calvini, G. Rodriguez, F. Inguglia, A. Mignone, U. P. Guerra, and F. Nobili, “The basal ganglia matching tools package for striatal uptake semi-quantification: description and validation,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 8, pp. 1240–1253, 2007.

[39] L. Tossici-Bolt, S. M. A. Hoffmann, P. M. Kemp, R. L. Mehta, and J. S. Fleming, “Quantification of [123I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 33, no. 12, pp. 1491–1499, 2006.

[40] S. J. Mo, J. Linder, L. Forsgren, and K. Riklund, “Accuracy of Visual Assessment of Dopamine Transporter Imaging in Early Parkinsonism,” *Movement Disorders Clinical Practice*, vol. 2, no. 1, pp. 17–23, 2014.

[41] Stanford University. (2020, Spring). CS231n: Convolutional Neural Networks for Visual Recognition. [Online]. Available: https://cs231n.github.io/