Expert Systems with Applications 41 (2014) 3333–3342

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging

R. Prashanth [a,*], Sumantra Dutta Roy [a], Pravat K. Mandal [b,c], Shantanu Ghosh [d]

[a] Department of Electrical Engineering, Indian Institute of Technology Delhi, India
[b] Neurospectroscopy and Neuroimaging Laboratory, National Brain Research Centre, India
[c] Department of Radiology, Johns Hopkins Medicine, MD, USA
[d] Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, MA, USA

ARTICLE INFO

Keywords:
Computer-aided early diagnosis
Parkinson's disease
Risk prediction
Pattern analysis
Support vector machine
Logistic regression

ABSTRACT

Early and accurate diagnosis of Parkinson's disease (PD) is important for early management, proper prognostication and for initiating neuroprotective therapies once they become available. Recent neuroimaging techniques such as dopaminergic imaging using single photon emission computed tomography (SPECT) with ^{123}I-Ioflupane (DaTSCAN) have shown to detect even early stages of the disease. In this paper, we use the striatal binding ratio (SBR) values that are calculated from the ^{123}I-Ioflupane SPECT scans (as obtained from the Parkinson's progression markers initiative (PPMI) database) for developing automatic classification and prediction/prognostic models for early PD. We used support vector machine (SVM) and logistic regression in the model building process. We observe that the SVM classifier with RBF kernel produced a high accuracy of more than 96% in classifying subjects into early PD and healthy normal; and the logistic model for estimating the risk of PD also produced high degree of fitting with statistical significance indicating its usefulness in PD risk estimation. Hence, we infer that such models have the potential to aid the clinicians in the PD diagnostic process.

1. Introduction

Parkinson's disease (PD) is a severe progressive neurodegenerative disorder neuropathologically characterized by the loss of dopaminergic neurons in the substantia nigra which result in substantial reduction of dopamine content in the striatum and a corresponding loss of dopamine transporters (DATs) (Booij et al., 1997). Currently, there are no definitive tests for the diagnosis of PD, and the clinical diagnosis is based on the presence of cardinal symptoms (tremor at rest, rigidity, bradykinesia, postural instability) and the response of the subject to PD medications (mainly levodopa).

The clinical diagnosis is clear-cut in the advanced stages of the disease when the symptoms are full-blown. However, in its early, when the symptoms are mild/incomplete or subtle, an accurate diagnosis becomes difficult (Booij & Knol, 2007; Cummings et al., 2011; Sixel-Doring, Liepe, Mollenhauer, Trautmann, & Trenkwalder, 2011; Tolosa, Borght, Moreno, & DaTSCAN Clinically Uncertain Parkinsonian Syndromes Study Group, 2007). For instance, the Parkinson's progression markers initiative (PPMI), which is a landmark and the first large-scale study to explore and identify PD progression biomarkers, points out that early diagnosis of *de novo* PD subjects, like those being recruited for PPMI, is difficult because characteristic signs and symptoms have not yet fully emerged and patients may present atypical signs and symptoms. (PPMI Study Protocol, http://www.ppmi-info.org/wp-content/uploads/2013/09/Attachment-4-PPMI-Protocol-AM6-V7-FINAL-Final.pdf).

Early and accurate diagnosis of PD is crucial for several reasons: early management, avoidance of unnecessary medical examinations and therapies and their associated financial costs, side-effects and safety risks (Cummings et al., 2011). Correct diagnosis is also critical for patients who are being recruited for clinical trial studies like the PPMI. Single photon emission computed tomography (SPECT) with ^{123}I-Ioflupane (DaTSCAN^{TM}, GE Healthcare; also known as [123I]FP-CIT) have shown to discriminate PD patients from healthy controls by depicting the presynaptic dopaminergic deficits in the striatum (caudate and putamen), even in the early stages of the disease and are becoming a valuable tool for the clinician in the diagnostic process (Bairactaris et al., 2009; Benamer et al., 2000; Booij et al., 1997; Seifert & Wiener, 2013; Tolosa et al., 2007; Winogrodzka et al., 2001).

Abbreviations: PD, Parkinson's disease; PPMI, Parkinson's progression markers initiative; SPECT, single photon emission computed tomography; SVM, support vector machine; SBR, striatal binding ratio.

* Corresponding author. Tel.: +91 9891279885.
E-mail address: eez108051@ee.iitd.ac.in (R. Prashanth).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.11.031

Diagnostic tools based on machine learning techniques such as support vector machine (SVM) and multivariate logistic regression (MLR) are important as they could assist the clinician in the early diagnostics, treatment planning and monitoring of disease progression (Orru, Pettersson-Yeo, Marquand, Sartori, & Mechelli, 2012). Both SVM as well as MLR are increasingly used in neuroimaging studies due to the following advantages: they allow characterization at an individual level rather than at group level, therefore yielding results with a potentially high level of clinical translation; and they are multivariate and supervised techniques which take into account characteristics of differently distributed populations encoded in a complex high-dimensional feature space, and use them to train the model and then categorize the data based on the trained model. SVM aims to find a hyperplane that classifies subjects into early PD or normal control. On the other hand, MLR determines the probability of PD for a subject based on these SBR measures and such probability estimation can be useful to classify subjects into different risk categories as studies suggest that SPECT imaging using ^{123}I-Ioflupane can also depict the progression of dopaminergic degeneration in PD (Winogrodzka et al., 2001).

Closely related works are that of Segovia et al. (2012), Illan et al. (2012), Rojas et al. (2013) and Towey, Bain, and Nijran (2011). Segovia et al. (2012) extracted out the voxels corresponding to the striatum and then performed data decomposition using partial least squares followed by classification into controls and Parkinsonian syndrome (PS) by means of an SVM classifier. Illan et al. (2012) also extracted the voxels corresponding to the striatum and then performed classification into controls and PS by using a SVM classifier with linear kernel. Rojas et al. (2013) performed selection of voxels corresponding to the striatum and then carried out three experiments. One using all features followed by classification using SVM classifier. Two using reduced number of features through independent component analysis (ICA) followed by classification using SVM. Third using reduced number of features through principal component analysis (PCA) followed by classification using SVM. They observed that the highest accuracy was obtained with the method involving PCA. Towey et al. (2011) used information from all voxels and performed feature extraction through singular value decomposition followed by classification into PS or non-PS (this study did not include controls). These studies had the following three limitations: they used huge number of features which required effective feature reduction techniques and such techniques may cause loss of information which may affect in the decision making process; their data set was of limited size; and none of the studies concentrated on prognostic or prediction models for estimating the probability of PD in subjects that might be useful in categorizing subjects into different risk categories.

The contributions of the paper are as follows:

1. We use only four features which are the striatal binding ratio (SBR) values for each of the four striatal regions (left and right caudate, left and right putamen) computed using either semi-automatic or highly accurate automated algorithms (Zubal, Wisniewski, Marek, & Seibyl, 2011), obtained from the PPMI (Marek et al., 2011) database, which is the first large-scale landmark study to identify progression biomarkers in PD, to develop automated classification and prediction/prognostic models which may aid the clinician in early diagnosis of the disease. We observe that the performance of SVM classifier using RBF kernel produced an accuracy which is higher than that obtained for the closely related works (Illan et al., 2012; Rojas et al., 2013; Segovia et al., 2012; Towey et al., 2011).
2. The sample size used for the experiments is also the highest as compared to the related works (Illan et al., 2012; Rojas et al., 2013; Segovia et al., 2012; Towey et al., 2011) and the PPMI database which we have used is a multi-centre international study involving subjects from different countries adding diversity in the database which was not there in related works, thereby making the models robust.
3. This is also first time that prediction/prognostic models based on SBR features for estimating the risk of PD is attempted using MLR and we observe that the logistic model showed high performance with statistical significance indicating its usefulness in PD risk estimation.

The rest of the paper is organized as follows. Section 2 contains the flowchart of the analysis carried out and describes about the PPMI database, study cohort, SBR values, statistical analysis of features, classification and prediction/prognostic model design. Section 3 provides the results and discussion from the experiments carried out. And finally we provide the conclusion of the work in Section 4.

2. Materials and methods

A flow chart of the analysis carried out in the paper is shown in Fig. 1. We use the SBR values from the four striatal regions (left and right caudate, and left and right putamen) as obtained from the PPMI database, for the experiments. All the four features are tested for statistical significance before performing classification and PD risk estimation.

2.1. Database and study cohort details

Data used in the preparation of this article were obtained from the Parkinson's progression markers initiative (PPMI) database (http://www.ppmi-info.org/data). For up-to-date information on the study, please visit www.ppmi-info.org. The PPMI (Marek et al., 2011) is a landmark and first large-scale, comprehensive, observational, international, multi-center study to identify PD progression biomarkers to improve understanding of the disease etiology as well as effectiveness of disease modifying therapeutic trials. For our experiments, we used the striatal binding ratio (SBR) values of the four striatal regions (left and right caudate, left and right putamen) that were computed from DaTSCAN SPECT images and were available from the PPMI database.

The database was downloaded on 18th March 2013. As per this date, the database contained SBR values from 179 normal (total of 181 observations as two of the subjects were scanned at 12 months along with the scan at screening visit) and 369 early PD (a total of 493 observations) subjects (Table 1). As PPMI is a longitudinal study, the subjects are evaluated longitudinally, i.e., evaluations occur at screening/baseline and at 3 month intervals during the first year of participation and then every 6 months thereafter.

The details of PPMI subject selection criteria, protocols in DaTSCAN imaging and the steps used by the Imaging Core (http://www.ppmi-info.org/about-ppmi/who-we-are/study-cores/) of the PPMI for calculating the SBR is provided in Appendix A. The PPMI Central SPECT Core Lab reconstructed, attenuation corrected, and analyzed the data with a standardized region of interest template

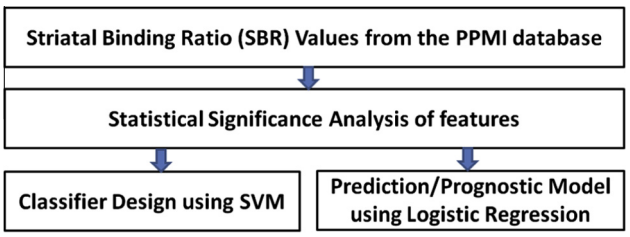

Fig. 1. Flowchart of the analysis carried out.

Table 1
SBR values (mean ± standard deviation) for right caudate, left caudate, right putamen and left putamen for normal and early PD population. In the early PD group, all subjects were in their early stage of the disease (All subjects were in Hoehn and Yahr (HY) stage 1 and 2 with mean ± standard deviation of the HY as 1.58 ± 0.49).

Case	No. of observations	Right caudate	Left caudate	Right putamen	Left putamen
Normal	181	1.93 ± 0.42	1.91 ± 0.42	1.29 ± 0.40	1.33 ± 0.40
Early PD	493	1.28 ± 0.38	1.26 ± 0.36	0.63 ± 0.24	0.63 ± 0.25

for the extraction of regional count densities in the left and right caudate and putamen. The SBR values are calculated for the right caudate, left caudate, right putamen and left putamen using occipital lobe region as the reference.

2.2. Statistical significance of SBR-based features

To assess the statistical significance of SBR-based features (predictors), univariate logistic regression analysis (Bewick, Cheek, & Ball, 2005) was used. All statistical analysis was carried out using SPSS 21 software (SPSS Inc., Chicago, IL). The threshold of significance was defined as a value of $p < 0.05$.

We also plot the histograms and notched box plots for each SBR feature to visualize its distribution for normal and early PD population (Fig. 2). The notched box plots (Fig. 2 (b,d,f,h)) show that the notches for early PD and normal control are fairly separated indicating the significance of these features. The histogram plots (Fig. 2 (a,c,e,f)) show that amount of overlap of distribution between the normal and early PD groups. The overlap is relatively higher for the caudate SBRs (both left and right) as compared to the putamenal SBRs (both left and right). The amount of overlap of the distributions of SBR measures determines the difficulty of the classification problem, i.e., classifying early PD from normal controls. Higher the overlap, the more difficult is the classification. The application of diagnostic tools comes into play here as they are capable of incorporating characteristics of these different distributed populations encoded in complex high-dimensional feature space, and using them to train the model and categorize the data.

2.3. SBR-based classification and prediction/prognostic modeling to distinguish early PD from healthy controls

Among the techniques for model building, support vector machine (SVM) and logistic regression are widely used in biomedicine (Dreiseitl & Ohno-Machado, 2002). SVMs are hard classifiers which directly target on the classification decision boundary without producing the probability estimation (Boser, Guyon, & Vapnik, 1992). On the other hand, logistic regression is a soft classifier which explicitly estimate the class-conditional probabilities and then perform classification based on estimated probabilities. The advantage with soft classifier techniques is that it produces output in the form of probabilities which allows categorization of subjects into clinically-important categories based on the predicted probability. However, the classification accuracy with logistic regression can be lower because of their linearity as compared to SVM (as we see in Sections 3.2 and 3.3) as SVMs can nonlinearly map samples into a higher dimensional space which makes it able to handle the case when the relation between the class label and features is nonlinear. We used LIBSVM (Chang & Lin, 2011) for implementing the SVM classifier and SPSS 21 software (SPSS Inc., Chicago, IL) for developing the logistic model.

2.3.1. Automatic SVM-based classification to distinguish early PD from healthy controls

Support vector machines (SVMs) are supervised techniques for classification which finds a linear hyperplane (decision boundary) with the largest margin by mapping input features to a higher dimensional space through either linear or non-linear kernel functions (Boser et al., 1992).

The dataset can be represented as $x_i \in R^d$; $i = 1, \ldots, n$, where n is the number of observations which is 674 and d is the number of features which is 4, in two classes (normal and early PD) and the class label y; $y_i \in \{\text{Early PD} = 1, \text{Normal} = 0\}$. In our case, as the data is unbalanced (i.e., different proportion of both classes: normal = 181 (26.85%) and early PD = 493 (73.15%)), we use the SVM formulation incorporating different penalty parameters for positive (early PD) and negative (normal) classes (Chang & Lin, 2011; Osuna, Freund, & Girosi, 1997) as shown below

$$\min_{w,b,\xi} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C^+\sum_{y_i=1}\xi_i + C^-\sum_{y_i=0}\xi_i \tag{1}$$

subject to

$$y_i\left(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b\right) \geq 1 - \xi_i; \quad \xi_i, > 0, \ i = 1, \ldots n \tag{2}$$

where C^+ and C^- are regularization parameters for positive and negative classes, respectively, ξ_i is the slack variable, $\phi(\boldsymbol{x}_i)$ is a function that maps $\boldsymbol{x}_i$ to a higher dimension feature space. The optimal $\boldsymbol{w}$ obtained after solving (1) is

$$\boldsymbol{w} = \sum_{i=1}^{n} y_i\alpha_i\phi(\boldsymbol{x}_i) \tag{3}$$

and the decision function is given by

$$sgn(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b) = sgn\left(\sum_{i=1}^{n} y_i\alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b\right) \tag{4}$$

where $K(\boldsymbol{x}, \boldsymbol{x}_i)$ is the kernel function defined by $K(\boldsymbol{x}, \boldsymbol{x}_i) = \phi(\boldsymbol{x})^T\phi(\boldsymbol{x}_i)$ and α_i's are the Lagrange multipliers used.

By using different kernel functions, varying degrees of nonlinearity and flexibility can be included in the model. For our experiments, we used the linear kernel defined by $K(\boldsymbol{x}, \boldsymbol{x}_i) = \boldsymbol{x}^T\boldsymbol{x}$, and the non-linear kernel of radial basis function (RBF) defined by $K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp(-\gamma||\boldsymbol{x} - \boldsymbol{x}_i||^2); \gamma > 0$ to be used with the SVM classifier. The classifier was evaluated using 10-fold cross validation.

2.3.2. Prediction/Prognostic modeling for early PD using multivariate logistic regression

In this work, multivariate binomial logistic regression technique (Bewick et al., 2005; Dreiseitl & Ohno-Machado, 2002; Kerr et al., 2010) was used for developing prediction/prognostic models for the purpose of risk prediction in PD. Binomial logistic regression models the probability of occurrence of one (early PD group, in this case) of the two classes of a dichotomous criterion.

In this technique, a linear combination of predictors (features) is used to fit a logit transformation of the probability of PD for each observation (n_i) and is given by

$$\text{logit}(\pi_i) = ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} \tag{5}$$

π_i is the likelihood of subject outcome to be PD, for each subject (n_i) which is given by $\pi_i = P(y_i = 1|\boldsymbol{x}_i, \boldsymbol{\beta}, \alpha)$; α is the intercept in the model which is a constant and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_k\}$ are the regression coefficients for the predictors and $\boldsymbol{x}_i$; $\boldsymbol{x}_i = [x_{1i}, \ldots, x_{ki}]$ is a data point which is a set of k features (can contain interaction term also). The regression

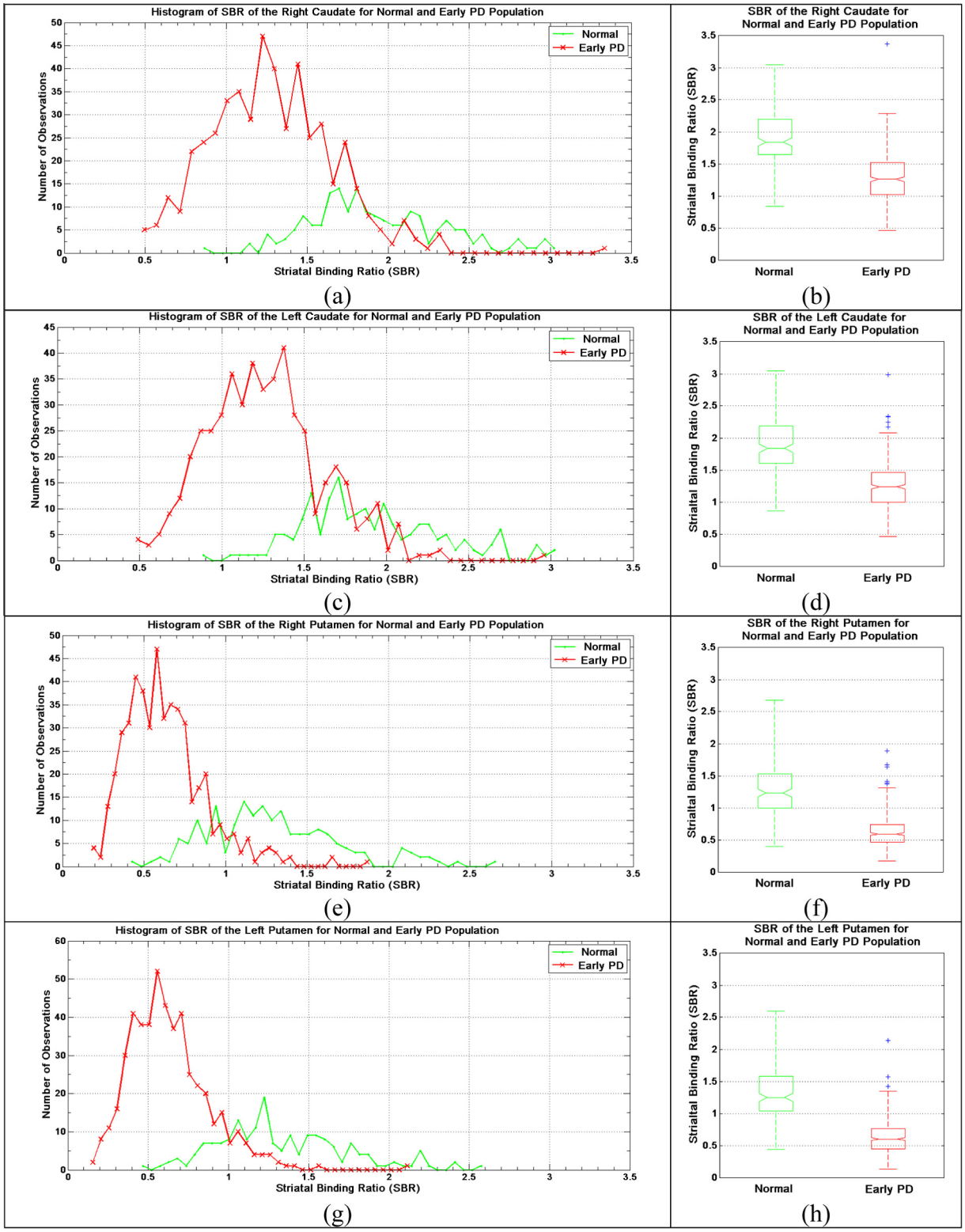

Fig. 2. Histogram plots and notched box plots of the striatal binding ratio (SBR) values for (a,b) right caudate (c,d) left caudate (e,f) right putamen (g,h) left putamen for normal and early PD population. (In each notched box plot, the central mark is the median (q_2), the edges of the box are the 25th (q_1) and 75th (q_3) percentiles, the whiskers extend to the most extreme data points that are not considered outliers, and outliers are plotted individually. The extremes of the notches or the centers of the triangular markers correspond to $q_2 \pm 1.57(q_3 - q_1)/\sqrt{n}$ where $n = 674$ is the number of observations).

coefficients are estimated using maximum likelihood estimation, and then solving the logit in order to n_i, we obtain the probability of PD for each subject which is the risk predictor as

$$\pi_i = \frac{1}{1 + \exp -(\alpha + \boldsymbol{\beta}.\mathbf{x}_i)} \tag{6}$$

Table 2
Result of statistical testing of each feature through univariate logistic regression.

Predictor	β	SE_β	Wald	*df*	*p*-value	Nagelkerke R^2
Right caudate	−4.137	.346	142.689	1	.0000	0.492
Left caudate	−4.243	.351	146.220	1	.0000	0.499
Right putamen	−6.367	.502	160.937	1	.0000	0.638
Left putamen	−7.015	.568	152.625	1	.0000	0.682

β is the value of the regression coefficient for the predictor in the univariate model, SE is its standard error, Wald is the Wald test statistic, *df* is the degree of freedom, *p*-value is the significance of the regression coefficient, Nagelkerke R^2 is the pseudo-R^2 measure that is a variation of R^2 that is used in ordinary least squares (OLS) regression. Note: the table does not show the constants obtained from univariate analyses.

Table 3
Accuracies (in %) obtained for the SVM classifier using RBF and linear kernel for each fold during 10-fold cross-validation.

Fold No.	Accuracy(%)	
	RBF	Linear
1	97.01	94.03
2	97.06	92.65
3	95.59	92.65
4	97.06	92.65
5	95.59	94.12
6	94.03	92.54
7	98.51	95.52
8	92.54	89.55
9	95.52	92.54
10	98.51	86.57
Mean Accuracy	96.14	92.28

3. Results and discussion

3.1. Statistical significance of SBR-based features

Table 2 show the results of univariate logistic regression analysis on the four SBR features (right caudate, left caudate, right putamen and left putamen). It is observed that all of the four features are statistically significant with $p < 0.05$. Table 2 also shows the Nagelkerke R^2 values (Nagelkerke, 1991), which is a variation of index R^2 that is used in ordinary least squares (OLS) regression. It is observed that higher Nagelkerke R^2 values are obtained for putamenal SBR (both right and left) as compared to the caudate SBR. This indicate that putamenal SBR had higher discriminative power than caudate SBR in distinguishing early PD from normal controls. This is also evident from the histogram plots which show lower overlaps for putamenal SBR (as shown in Fig. 2 (e,g)) than for caudate SBR (as shown in Fig. 2 (a,c)). We observe that this finding goes with the findings from following studies (Kish, Shannak, & Hornykiewicz, 1988; Piggott et al., 1999) which performed post-mortem studies to demonstrate severe reductions in the dopamine concentration in the striatum of PD patients, with greater reduction in the putamen than the caudate.

3.2. Automatic classification and prediction/prognostic modeling for early PD

We use SVM and multivariate logistic regression to develop models for distinguishing early PD from healthy normal. SVMs performed hard classification, where as MLR performed soft classification enabling us to develop prediction/prognostic model for PD risk estimation. We used SVM with RBF kernel and compared it with the SVM classifier with the linear kernel. Subsections below describes the performance measures obtained for the SVM classifiers and logistic model.

The highlights of the present study are that: We observe that the performance measure obtained by the SVM classier using RBF kernel to be the highest among all closely related works (Illan et al., 2012; Rojas et al., 2013; Segovia et al., 2012; Towey et al., 2011). We used only four features without needing any feature selection techniques. The sample size used for the experiments is also the highest as compared to the related works and the PPMI database which we have used is a multi-centre international study involving subjects from different countries adding diversity in the database which was not there in related works. This is also first time that prediction/prognostic models based on SBR features for estimating the risk of PD is attempted using MLR and we observe high performance from these models also.

3.2.1. SVM-based automatic classification to distinguish early PD from healthy controls

Table 3 shows the accuracies obtained for each fold during 10-fold cross validation of the SVM classifier using RBF and linear kernels. Table 4 shows the confusion matrix components and the performance measures obtained for the SVM classifier with the RBF and linear kernel. It is observed that the SVM classifier with RBF kernel (Accuracy = 96.14%, Sensitivity = 96.55%, and Specificity = 95.03%) performs better than the linear one (Accuracy = 92.28%, Sensitivity = 95.33%, Specificity = 83.98%). This is because RBF kernel can nonlinearly map samples into a higher dimensional space, so it is able to handle the case when the relation between the class label and features is nonlinear. Fig. 3 shows the box plots of SBRs for misclassified cases from the SVM classifier using RBF kernel which shows that the cases misdiagnosed as PD had high values of SBR (which was supposed to be low) and cases misdiagnosed as healthy normal had low values of SBR (which was supposed to be high).

3.2.2. Prediction/Prognostic modeling for early PD using multivariate logistic regression

Prediction/prognostic modeling for PD risk estimation is carried out by using multivariate logistic regression analysis. A four-predictor (using all the 4 features) logistic model was initially fitted to the data. Table 5 shows resulting model and it is observed that the two features (right caudate and left caudate) were not contributing to the model as per their *p*-value ($p > 0.05$), although they contain significant information to distinguish between early PD and normal as discussed in Section 3.1.

To tackle this issue, instead of using these two features, we included the interaction term, which is the product of the two written as (right caudate * left caudate) and then fitted the model. Table 6 shows the resulting three-predictor model and it is observed that all the three predictors (left caudate, right caudate and (right caudate * left caudate)) were significantly contributing to the model ($p < 0.05$). Table 6 also shows that among the three predictors, the SBR forthe left and right putamen had higher significance than the interaction term. The model obtained can be written as

Table 4
Confusion matrix components and performance measures for the SVM classifier with RBF kernel (SVM_{RBF}) and SVM classifier with Linear kernel (SVM_{Linear}).

Classifiers	True positive	False negative	False positive	True negative	Accuracy (%)	Sensitivity (%)	Specificity (%)
SVM_{RBF}	476	17	9	172	96.14	96.55	95.03
SVM_{Linear}	470	23	29	152	92.28	95.33	83.98

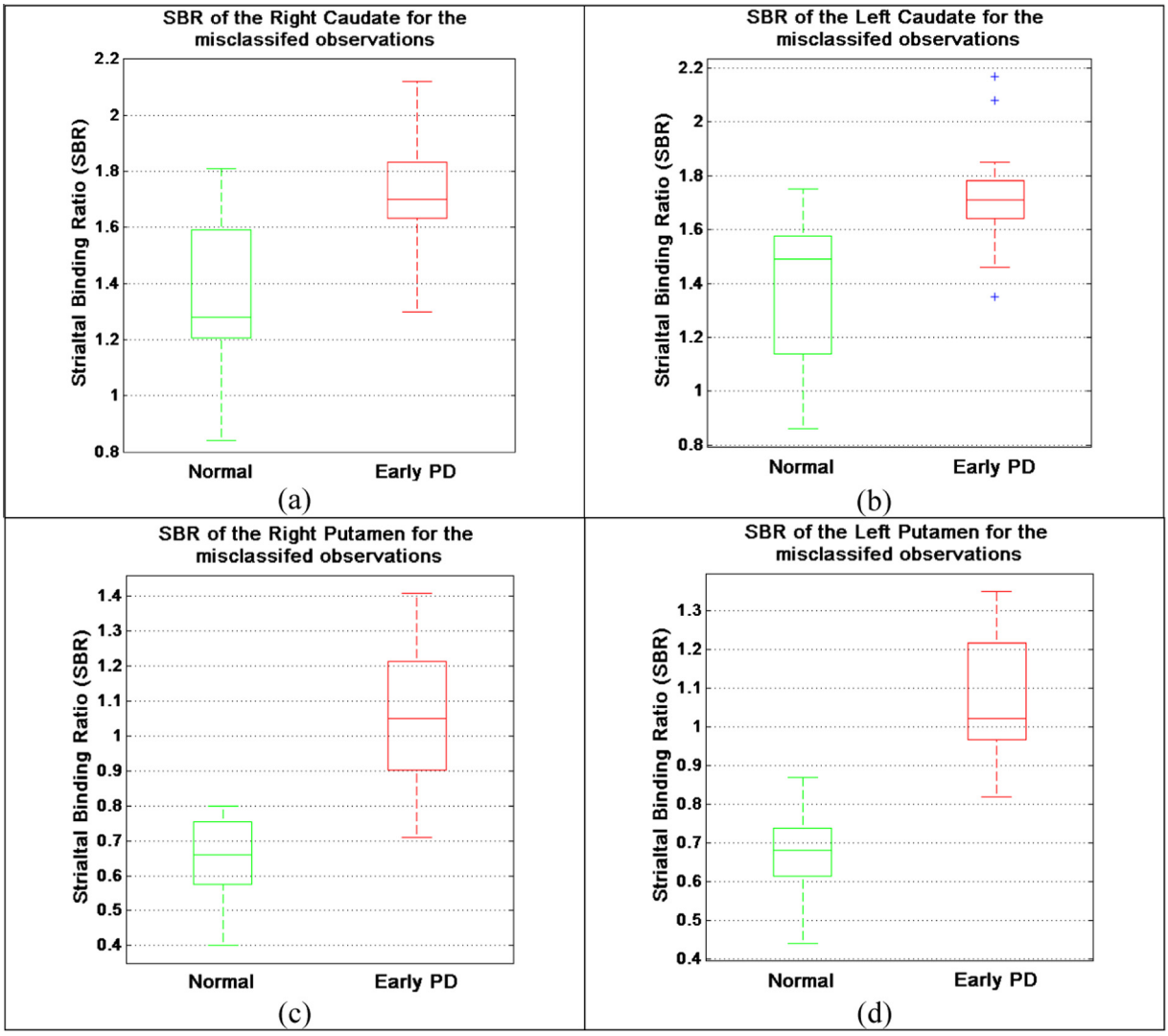

Fig. 3. Box plot of Striatal Binding Ratio (SBR) of (a) right caudate (b) left caudate (c) right putamen (d) left putamen for the misclassified samples from SVM classifier with RBF kernel.

Table 5
Four-predictor logistic model developed using predictors of left putamen, right putamen, left caudate and right caudate SBR values.

Predictors	$\beta's$	SE_β	Wald	*df*	*p*-value	exp(β) (odds ratio)	95% C.I. for exp(β)	
							Lower	Upper
Left putamen	−5.6938	0.92	38.12	1	0.0000	0.0034	0.00	0.02
Right putamen	−3.7380	0.83	20.36	1	0.0000	0.0238	0.00	0.12
Left caudate	0.5071	0.81	0.39	1	0.5337	1.6605	0.34	8.20
Right caudate	0.7960	0.82	0.94	1	0.3330	2.2167	0.44	11.11
Constant	7.4655	0.74	100.95	1	0.0000	1746.7824		

β is the value of the regression coefficient, SE_β is its standard error, Wald is the Wald test statistic, *df* is the degree of freedom, *p*-value shows the significance of each regression coefficient in the multivariate model.

$$\text{logit}(\pi_i) = 8.5132 - 5.8881 * \text{left putamen} - 3.7033 * \text{right putamen} + 0.4504 * (\text{left caudate} * \text{right caudate}) \quad (7)$$

We observe from the model that the log of the odds of the subject being early PD was negatively related to both left and right putamen SBR and positively to the interaction term (left caudate * right caudate). In other words, higher the SBR in the right and left putamen, the less likely that the subject has PD.

3.2.2.1. Evaluation of the logistic regression model. The evaluation of the logistic model is carried out as given in (Peng, Lee, & Ingersoll, 2002)

(i) Overall Model Evaluation

Table 6
Three-predictor logistic model through logistic regression using the predictors left putamen, right putamen and the interaction term which is the product of left and right caudate (written as left caudate * right caudate) SBR values.

Predictors	β's	SE_β	Wald	*df*	*p*-value	exp(β) (odds ratio)	95% C.I. for exp(β)	
							Lower	Upper
Left putamen	−5.8881	0.83	50.11	1	0.0000	0.0028	0.00	0.01
Right putamen	−3.7033	0.70	27.83	1	0.0000	0.0246	0.01	0.10
(Left caudate * right caudate)	0.4504	0.20	5.17	1	0.0230	1.5689	1.06	2.31
Constant	8.5132	0.66	167.41	1	0.0000	4980.0569		

β is the value of the regression coefficient, SE_β is its standard error, Wald is the Wald test statistic, *df* is the degree of freedom, *p*-value shows the significance of each regression coefficient in the multivariate model.

Table 7
Omnibus tests of model coefficients.

	Chi-square	*df*	*p*-value
Step	455.9850	3	.0000
Block	455.9850	3	.0000
Model	455.9850	3	.0000

Table 8
Goodness-of-fit tests.

Chi-square	*df*	*p*-value
2.9162	8	.9395

1. Hosmer and Lemeshow Test.
2. Cox & Snell $R^2 = 0.4916$.
3. Nagelkerke $R^2 = 0.7149$.

Table 7 shows the overall test results for the three-predictor logistic model. The chi-square value of 455.985 with a $p < 0.05$ indicates that this model as a whole fits significantly better than a null model (model with no predictors).

(ii) Statistical tests of individual parameters in the multivariate logistic model

The statistical significance of regression coefficients corresponding to each predictor is tested using the Wald chi-square statistic. As observed in Table 6, all the three predictors along with the constant, which is the intercept, are statistically significant ($p < 0.05$).

(iii) Goodness-of-fit statistics

The Goodness-of-fit statistics, through Hosmer and Lemeshow test (Hosmer & Lemeshow, 2000) and R^2 indices, assess the fit of a logistic model against actual outcomes. The Hosmer and Lemeshow test is a statistical test for evaluating the goodness-of-fit for logistic regression models. In Hosmer and Lemeshow test, the data is divided into approximately ten groups and then chi-square statistics is calculated using the observed and expected number of cases in each group defined by increasing order of estimated risk. Table 8 shows the result of Hosmer and Lemeshow test. It gave a small chi-square of 2.916 (with larger *p*-value closer to 1) with 8 degrees of freedom indicating that the actual outcome (early PD or normal) is not significantly different from model risk prediction and therefore, the overall model was well fit to the data. Table 8 also shows R^2 indices, defined by Cox and Snell (1989), and Nagelkerke (1991)). These indices are variations of the R^2 that is used for ordinary least squares (OLS) regression model. Cox and Snell R^2 of 0.4916 and Nagelkerke R^2 of 0.713 are high and it indicates that the model is useful in risk prediction.

Table 9
Classification table[a].

Observed	Predicted		
	Normal	Early PD	% Correct
Normal	140	41	77.35
Early PD	21	472	95.74
Overall%			90.80

Sensitivity = 95.74%, Specificity = 77.35%.
[a] cutoff value is 0.5.

(iv) Validation of the predicted probabilities

Logistic regression predicts the logit of an event outcome (early PD or normal) from a set of predictors. The predicted probabilities, which are obtained from the logit, can then be revalidated with the actual outcome to determine if high probabilities are indeed associated with higher risk of PD and low probabilities with lower risk of PD. The degree to which predicted probabilities agree with actual outcome is shown through a classification table as shown in Table 9. The overall classification accuracy was as high as 90.8% (with cutoff for classification set as 0.5) indicating that the model with 3 predictors performs well in predicting the subject outcome. It is observed that the classification accuracy is not as high as that we obtained using SVM classifier with RBF kernel. This is because logistic regression models are discriminative models for classification that produces linear decision boundaries, and are not that flexible as the non-linear SVMs.

To provide a visual demonstration of the correct and incorrect predictions, a histogram of the predicted risk probabilities is plotted (Fig. 4). It is observed that a U-shaped distribution is obtained with cases clustered at each end showing correct classification and indicating that the predictions are well-differentiated. Along with it, we also plotted these probabilities with the lowest putamenal SBR (Fig. 5). It is observed that all the observations now fit a logistic sigmoid. It is to be noted that as the logistic model is multivariate, the subject outcome (early PD or normal) is determined by all the three predictors used in the model and is not defined only by the lowest putaminal SBR. Fig. 5 is mainly for the purpose of visualizing the predicted probability and putamenal SBR was chosen instead of caudate SBR in x-axis as it was more significant in risk prediction (as per Table 6).

3.3. Future work

There are few other conditions other than PD, which show similar clinical symptoms as in PD such as Scans Without Evidence of Dopaminergic Deficit (SWEDD) and Essential Tremor (ET). Proper diagnosis of these conditions is very crucial as misdiagnosis as PD may lead to unnecessary medical examinations and therapies and associated side-effects. Other than the classification of early PD patients from healthy normal, these models may also find useful in the detection of SWEDD subjects and subjects with ET as

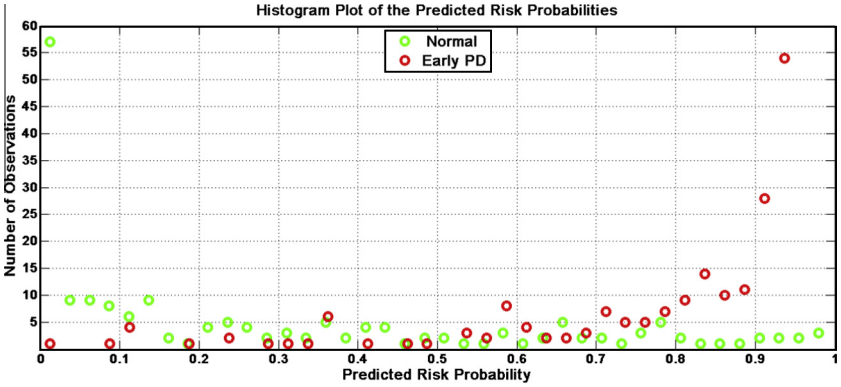

Fig. 4. Histogram of the predicted risk probabilities (plotted with 40 bins). Note: for better visualization, the maximum limit of the *y*-axis was set to 60, and the number of early PD samples with risk probability of 0.9615 (80 observations) and 0.9865 (218 observations) are not shown (0.9615 and 0.9865 are the bin centers).

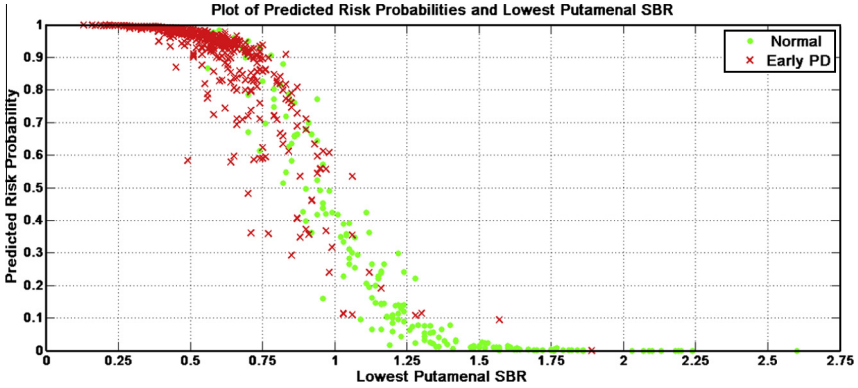

Fig. 5. Plot of the predicted risk probabilities against the lowest putamenal Striatal Binding Ratio (SBR).

these cases are very likely to have normal SPECT scans as explained in detail in the below subsections. The validity of these models in distinguishing early PD from SWEDD or ET is to be carried out as our future work.

3.3.1. Application in distinguishing SWEDD subjects from early PD

Dopaminergic studies have shown that approximately 10% of subjects thought clinically to have PD have normal dopaminergic imaging scans. These subjects are referred to as SWEDDs (Scans Without Evidence of Dopaminergic Deficit) (Schwingenschuh et al., 2010). There is substantial evidence which suggest that these subjects, most likely, do not have PD as long term follow-up of these subjects indicated poor response to PD medications (levodopa) (Marek, Jennings, & Seibyl, 2005) and lack of progression on sequential dopaminergic imaging (Schneider et al., 2007). The SBR values for SWEDD are similar to that of healthy controls and cross-sectional data across an age range show a similar age-associated reduction in SWEDD as in healthy controls (http://www.ppmi-info.org/wp-content/uploads/2012/06/Seibyl-PPMI-MDS-2012.pdf). The results of these studies indicate that an abnormal dopamine transporter imaging (like DaTSCAN imaging using SPECT), at least in cases where there is a diagnostic uncertainty, strongly supports a diagnosis of parkinsonism (Cummings et al., 2011; Schwingenschuh et al., 2010). Distinguishing SWEDD subjects from PD is important as most of the subject receive unnecessary and inappropriate treatment for many years (Schwingenschuh et al., 2010). As the proposed classifier is able to distinguish an abnormal scan (as in PD) from a normal scan with high accuracy, such system may be useful in differentiating SWEDD from PD, and act as an adjunct to other diagnostic evaluations.

3.3.2. Application in distinguishing Essential Tremor (ET) subjects from early PD

Although distinguishing PD from Essential Tremor (ET) on a clinical basis is straightforward, ET is increasingly recognized as a heterogeneous disorder that can encompass features such as rest tremor which can challenge accurate diagnosis (Cohen, Pullman, Jurewicz, Watner, & Louis, 2003; Louis, 2009). In a study with 71 patients with ET, it was observed that about 1 in 3 patients with tremor was misdiagnosed as having ET, with most frequent false diagnosis being PD and dystonia (Jain, Lo, & Louis, 2006). A series of studies, cross-sectional and longitudinal, have shown a decrease in dopamine transporter density in PD patients compared to ET and no differences between ET and healthy controls (Benamer et al., 2000; Doepp et al., 2008; Isaias et al., 2010). As the ET patients have scans in the normal range, the proposed classifier system may aid in distinguishing ET from PD and act as an aid for the clinician in the diagnostic process.

4. Conclusion

Early diagnosis of Parkinson's disease is of at most importance for early management and treatment planning as clinical symptoms in PD arise only when there is more than 60% loss of dopaminergic neurons. There are no laboratory tests for the diagnosis of PD causing a high rate of misdiagnosis especially when the disease is in the early stages and when the diagnosis is made by a non-specialist. Hence, diagnostic tools based on machine learning techniques are important as they can aid the clinician in the diagnosis process. Recent neuroimaging studies with SPECT using ^{123}I-Ioflupane (DaTSCAN) as radiotracer have shown to be a

sensitive marker even in the early stages of the disease. In this work, we developed diagnostic models using machine learning techniques for classification into early PD and healthy normal as well as prediction/prognostication of PD. We use the striatal binding ratio (SBR) values for the four striatal regions (left and right caudate, and left and right putamen) obtained from the Parkinson's Progression Marker's Initiative (PPMI) database which is a landmark and first large-scale longitudinal study of PD to explore progression biomarkers. We observe that the SVM classifier using RBF kernel produced a high accuracy in classifying early PD from normal, and the prediction/prognostic model based on multivariate logistic regression also gave high performance along with statistical significance indicating the usefulness of the model in PD risk estimation. The contributions of the present study are (1) We obtained high classification performance using only four features without needing any feature selection techniques making the system simple. (2) The sample size used for the experiments is very high and the PPMI database which we have used is a multi-centre international study involving subjects from different countries adding diversity in the database and therby making the models more robust (3) This is also first time that prediction/prognostic models based on SBR features for estimating the risk of PD is attempted using MLR and we observe high performance from these models also. Other than classifying early PD from normal, these models can also find useful in distinguishing early PD from SWEDD subjects and subjects with Essential Tremor as these subjects are likely to have normal SPECT scans. Hence, we infer from the study that these models have the potential to distinguish abnormal scans corresponding to early PD from normal scans corresponding to healthy normal or SWEDD subjects or subjects with Essential Tremor. Application of these models needs to be validated for SWEDD and ET conditions which we propose to carry out as future work.

Acknowledgement

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including [list of all of the PPMI funding partners found at www.ppmi-info.org/fundingpartners].

Appendix A

A.1. Subject selection criteria

The PPMI mainly recruits early-untreated PD subjects and age-matched healthy controls for the study. As per the PPMI subject selection criteria (Marek et al., 2011), PD subjects are recruited at disease threshold, i.e., they are required to have an asymmetric resting tremor or asymmetric bradykinesia or two of bradykinesia, resting tremor and rigidity with diagnosis within two years and to be untreated for PD (Marek et al., 2011). All subjects in the PPMI study undergo dopamine transporter (DAT) imaging and DAT deficit was required for enrollment into PD category. Healthy subjects have no significant neurologic dysfunction, no first degree family member with PD and have Montreal Cognitive Assessment > 26 (Marek et al., 2011).

A.2. DaTSCAN imaging

As per the PPMI Imaging Protocol (Schedule of Activities) (http://www.ppmi-info.org/wp-content/uploads/2010/07/Imaging-Manual.pdf), subjects with PD have DaTSCAN imaging at screening and at months 12 (Visit 4), 24 (Visit 6), and 48 (Visit 10) or at prematural withdrawal (if a scan has not been completed in the last 12 months). Healthy control subjects have DaTSCAN imaging at screening. Prior to DaTSCAN injection, subjects are pretreated with saturated iodine solution (10 drops in water) or perchlorate (1000 mg,). The target dose for subjects is 185 MBq or 5.0 mCi of DaTSCAN. The dose range for injection is 111 to 185 MBq or 3.0 to 5.0 mCi of DaTSCAN. Subjects are imaged 4 ± 0.5 h later.

Reconstructed SPECT scans from the PPMI Imaging sites are sent to the Imaging Core at Institute for Neurodegenerative Disorders (IND) at New Haven, CT for visual assessment to check for the evidence of dopamine transporter deficit. Two qualified readers assess each scan and their assessment must be in agreement. If the reader's assessments differ, the scan will be adjudicated and the agreed interpretation will be sent to the site. The imaging interpretation will serve as final criteria for enrolment into the study (http://www.ppmi-info.org/wp-content/uploads/2013/02/PPMI-Protocol-AM5-Final-27Nov2012v6-2.pdf). For each subject, the Imaging Core calculates Striatal Binding Ratios (SBR) values, which is the DaTSCAN uptake in the striata, relative to the DaTSCAN uptake in the occipital area and this ratio is the primary outcome that is used for quantitating dopamine transporters in suspected parkinsonian syndromes.

A.3. Striatal binding ratio (SBR) calculation

For calculating the striatal binding ratio (SBR) from a specific region, the Imaging Core of the PPMI carried out following steps (http://www.ppmi-info.org/wp-content/uploads/2013/06/Seibyl-PPMI-MDS-2013-Sydney__sjl.pdf):

1. Central SPECT core lab iteratively (HOSEM) reconstructs the data from raw projection data using HERMES (Hermes Medical Solutions, Skeppsbron 44, 111 30 Stockholm, Sweden) system.
2. The HOSEM reconstructed files were then transferred to the PMOD (PMOD Technologies, Zurich, Switzerland) for subsequent processing including attenuation correction based on phantoms acquired during the site visit.
3. Spatial normalization of image to standard Montreal Neurologic Institute (MNI) space is carried out next to create consistent orientation.
4. Then, the transaxial slice with the highest striatal uptake was identified and the 8 hottest striatal slices around it were averaged to generate a single slice image.
5. Apply standard region of interest template on the four striatal regions (left and right caudate, and left and right putamen), and the occipital cortex (reference region).
6. Extract count densities and calculate Striatal Binding Ratios (SBR) for each of the four striatal regions using the formula:

$$\text{SBR of a target region} = \frac{\text{Target region count density}}{\text{Reference region count density}}$$

References

Bairactaris, C., Demakopoulos, N., Tripsianis, G., Sioka, C., Farmakiotis, D., Vadikolias, K., et al. (2009). Impact of dopamine transporter single photon emission computed tomography imaging using I-123 ioflupane on diagnoses of patients with parkinsonian syndromes. *Journal of Clinical Neuroscience, 16*, 246–252.

Benamer, T. S., Patterson, J., Grosset, D. G., Booij, J., de Bruin, K., van Royen, E., et al. (2000). Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group. *Movement Disorders, 15*, 503–510.

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: logistic regression. *Critical Care, 9*, 112–118.

Booij, J., & Knol, R. J. (2007). SPECT imaging of the dopaminergic system in (premotor) Parkinson's disease. *Parkinsonism Related Disorder, 13*(Suppl. 3), S425–S428.

Booij, J., Tissingh, G., Boer, G. J., Speelman, J. D., Stoof, J. C., Janssen, A. G., et al. (1997). [123I]FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labelling in early and advanced Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry, 62*, 133–140.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In: D. Haussler (Ed.), Proceedings of the 5th Annual

Workshop on Computational Learning Theory (COLT'92) (pp. 144–152). Pittsburgh, PA: ACM Press.
Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 1–27.
Cohen, O., Pullman, S., Jurewicz, E., Watner, D., & Louis, E. D. (2003). Rest tremor in patients with essential tremor: prevalence, clinical correlates, and electrophysiologic characteristics. *Archives of Neurology, 60*, 405–410.
Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London: Chapman and Hall.
Cummings, J. L., Henchcliffe, C., Schaier, S., Simuni, T., Waxman, A., & Kemp, P. (2011). The role of dopaminergic imaging in patients with symptoms of dopaminergic system neurodegeneration. *Brain, 134*, 3146–3166.
Doepp, F., Plotkin, M., Siegel, L., Kivi, A., Gruber, D., Lobsien, E., et al. (2008). Brain parenchyma sonography and 123I-FP-CIT SPECT in Parkinson's disease and essential tremor. *Movement Disorders, 23*, 405–410.
Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics, 35*, 352–359.
Hosmer, D. W. J., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
Illan, I. A., Gorrz, J. M., Ramirez, J., Segovia, F., Jimenez-Hoyuela, J. M., & Ortega Lozano, S. J. (2012). Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging. *Medical Physics, 39*, 5971–5980.
Isaias, I. U., Marotta, G., Hirano, S., Canesi, M., Benti, R., Righini, A., et al. (2010). Imaging essential tremor. *Movement Disorders, 25*, 679–686.
Jain, S., Lo, S. E., & Louis, E. D. (2006). Common misdiagnosis of a common neurological disorder: how are we misdiagnosing essential tremor? *Archives of Neurology, 63*, 1100–1104.
Kerr, G. K., Worringham, C. J., Cole, M. H., Lacherez, P. F., Wood, J. M., & Silburn, P. A. (2010). Predictors of future falls in Parkinson disease. *Neurology, 75*, 116–124.
Kish, S. J., Shannak, K., & Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. *New England Journal of Medicine, 318*, 876–880.
Louis, E. D. (2009). Essential tremors: a family of neurodegenerative disorders? *Archives of Neurology, 66*, 1202–1208.
Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., et al. (2011). The parkinson progression marker initiative (PPMI). *Progress of Neurobiology, 95*, 629–635.
Marek, K., Jennings, D., & Seibyl, J. (2005). Long-term follow-up of patients with scans without evidence of dopaminergic deficit (SWEDD) in the ELLDOPA study. *Neurology, 64*(Suppl. 1), A274.
Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691–692.
Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews, 36*, 1140–1152.
Osuna, E., Freund, R., & Girosi, F. (1997). Support vector machines: Training and applications. In: AI Memo 1602: Massachusetts Institute of Technology.
Peng, C.-Y. J., Lee, K. D., & Ingersoll, G. M. (2002). An Introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*, 3–14.
Piggott, M. A., Marshall, E. F., Thomas, N., Lloyd, S., Court, J. A., Jaros, E., et al. (1999). Striatal dopaminergic markers in dementia with Lewy bodies, Alzheimer's and Parkinson's diseases: rostrocaudal distribution. *Brain, 122*(Pt 8), 1449–1468.
Rojas, A., Górriz, J. M., Ramírez, J., Illán, I. A., Martínez-Murcia, F. J., Ortiz, A., et al. (2013). Application of empirical mode decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson disease. *Expert Systems with Applications, 40*, 2756–2766.
Schneider, S. A., Edwards, M. J., Mir, P., Cordivari, C., Hooker, J., Dickson, J., et al. (2007). Patients with adult-onset dystonic tremor resembling parkinsonian tremor have scans without evidence of dopaminergic deficit (SWEDDs). *Movement Disorders, 22*, 2210–2215.
Schwingenschuh, P., Ruge, D., Edwards, M. J., Terranova, C., Katschnig, P., Carrillo, F., et al. (2010). Adult onset asymmetric upper limb tremor misdiagnosed as Parkinson's disease: a clinical and electrophysiological study. *Movement Disorders, 25*, 560–569.
Segovia, F., Gorriz, J. M., Ramirez, J., Alvarez, I., Jimenez-Hoyuela, J. M., & Ortega, S. J. (2012). Improved parkinsonism diagnosis using a partial least squares based approach. *Medical Physics, 39*, 4395–4403.
Seifert, K. D., & Wiener, J. I. (2013). The impact of DaTscan on the diagnosis and management of movement disorders: a retrospective study. *American Journal of Neurodegenerative Disease, 2*, 29–34.
Sixel-Doring, F., Liepe, K., Mollenhauer, B., Trautmann, E., & Trenkwalder, C. (2011). The role of 123I-FP-CIT-SPECT in the differential diagnosis of Parkinson and tremor syndromes: a critical assessment of 125 cases. *Journal of Neurology, 258*, 2147–2154.
Tolosa, E., Borght, T. V., Moreno, E., & DaTSCAN Clinically Uncertain Parkinsonian Syndromes Study Group. (2007). Accuracy of DaTSCAN (123I-Ioflupane) SPECT in diagnosis of patients with clinically uncertain parkinsonism: 2-year follow-up of an open-label study. *Movement Disorders, 22*, 2346–2351.
Towey, D. J., Bain, P. G., & Nijran, K. S. (2011). Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images. *Nuclear Medicine Communications, 32*, 699–707.
Winogrodzka, A., Bergmans, P., Booij, J., van Royen, E. A., Janssen, A. G., & Wolters, E. C. (2001). [123I]FP-CIT SPECT is a useful method to monitor the rate of dopaminergic degeneration in early-stage Parkinson's disease. *Journal of Neural Transmission, 108*, 1011–1019.
Zubal, G., Wisniewski, G., Marek, K., & Seibyl, J. (2011). Automated program for analyzing striatal uptake of DaTSCAN SPECT images in humans suspected of Parkinson's disease. *Journal of Nuclear Medicine, 52*, 2098 [Meeting Abstracts].