NeuroImage 152 (2017) 299–311

Contents lists available at ScienceDirect

NeuroImage

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

Voxel-based logistic analysis of PPMI control and Parkinson's disease DaTscans

Hemant D. Tagare[a,b,*], Christine DeLorenzo[c], Sudhakar Chelikani[a], Lawrence Saperstein[a], Robert K. Fulbright[a]

[a] Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA
[b] Department of Biomedical Engineering, Yale University, New Haven, CT, USA
[c] Department of Psychiatry, Stony Brook University, NY, USA

ARTICLE INFO

Keywords:
Parkinson's disease
DaTscan
PPMI
Logistic Lasso
Logistic Principal Components

ABSTRACT

A comprehensive analysis of the Parkinson's Progression Markers Initiative (PPMI) Dopamine Transporter Single Photon Emission Computed Tomography (DaTscan) images is carried out using a voxel-based logistic lasso model. The model reveals that sub-regional voxels in the caudate, the putamen, as well as in the globus pallidus are informative for classifying images into control and PD classes. Further, a new technique called logistic component analysis is developed. This technique reveals that intra-population differences in dopamine transporter concentration and imperfect normalization are significant factors influencing logistic analysis. The interactions with handedness, sex, and age are also evaluated.

Introduction

Parkinson's disease, PPMI, and machine learning

Dopamine transporter imaging by [123I]-FP-CIT SPECT (also known as DaTscan) is used to diagnose Parkinson's disease (PD) and to distinguish it from other movement disorders, such as essential tremor (Benamer et al., 2000). In the clinic, most DaTscans are usually interpreted visually by experts, but automated quantitative analysis is likely to improve the interpretation. The European Association of Nuclear Medicine Neuroimaging Committee recommends quantitative analysis in addition to visual analysis (Darcourt et al., 2010). Because PD primarily affects dopaminergic neurons, most previous quantitative analysis of DaTscans focused on the striatum. We too focus on the striatum, but also include the globus pallidus and the thalamus in the analysis. The motivation for including these extra-striatal structures is discussed below in detail.

The Parkinson's Progression Markers Initiative (PPMI) is a study that offers an unprecedented number of DaTscans for analysis (www.ppmi.org). As of April 2016, over 600 subjects (control+PD) have been scanned, and after reconstruction and registration to a common space, their images are available for download and analysis. This large amount of data opens the door to using machine learning techniques to classify control and PD images.

Much of the previous work on machine learning/automated analysis of DaTscans is region-based, e.g. Prashanth et al. (2014); Zubal et al. (2007). The regional striatal binding ratios are calculated for the right and left putamen and the right and left caudate, and these four numbers are used in all subsequent analysis. Region-based analysis is usually justified on the grounds that there is dopamine loss in the putamen relative to the caudate in PD. However, from a statistical point of view, there are limitations to region-based analysis. First, it is unclear whether a single number calculated from a region is statistically optimal for analysis or for classification. Certain voxels within the region may have higher dopamine transporter loss than others and hence may be more informative. Second, it is not clear why only the putamen and the caudate should enter into quantification. Extra-striatal regions can contain significant amount of dopamine, and using these regions can improve the statistical reliability of the method. This could happen, for example, if the extra-striatal regions are pooled with the caudate to provide a reference to compare the putamen with. Pooling data improves statistical reliability. Dopaminergic neurons are known to densely occur in the primate thalamus (Sánchez-González et al., 2005), and the thalamus was added to our analysis for that reason. Extra-striatal structures are also known to be involved in PD. The globus pallidus is known to be involved in PD subtypes (Rajput et al., 2008), and was included in our analysis for that reason. Including all voxels from these regions in the data and letting an algorithm decide which voxels are most informative is likely to be statistically more meaningful than assuming *a priori* which voxels are

* Corresponding author at: Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA.
E-mail address: hemant.tagare@yale.edu (H.D. Tagare).

http://dx.doi.org/10.1016/j.neuroimage.2017.02.067
Received 20 October 2016; Accepted 23 February 2017
Available online 27 February 2017

important. Such an algorithm is *voxel-based* rather than region based.

We use the logistic lasso (Tibshirani, 1996) as the machine learning method for classification. The logistic lasso is voxel-based. It works by using a linear combination of a sparse set of voxels to calculate the probabilities of belonging to the control and PD classes. Voxels in the sparse set are chosen purely based on training data. Below, we use the informal term "informative voxels" to denote those voxels that are statistically useful for classification. Informative voxels are likely to be a subset of all voxels that are analyzed, and possibly also a subset of all voxels affected by PD.

Understanding the heterogeneity of the PPMI data set is also important because heterogeneity in the DaT signal can be confounding; this can happen, for example, in a clinical trial where response to dopaminergic therapies is measured. We need a greater understanding of how the image features that distinguish controls from PDs vary within each population. To achieve this, we introduce the concept of logistic principal components (LPC). LPCs are particularly illuminating for the PPMI data, as we show in the Results section. We also investigate the interaction of the discriminatory image feature with handedness, sex, and age and establish the significance of the interaction with p-values.

To our knowledge, such a comprehensive analysis of PPMI DaTscans at the voxel level has not yet been carried out.

Previous work

Imaging holds considerable promise in evaluating pre-motor PD, assessing disease progression, and in differential diagnosis. Excellent reviews are available in Booij and Knol (2007); Tatsch and Poepperl (2013). While our paper is focused on analyzing DaTscan images, techniques for analysis of other SPECT methods have also been developed; one example is the IBZM tool (Buchert et al., 2006).

Machine learning/automated classification of DaTscan images has been applied to non-PPMI data (Illan et al., 2012; Koch et al., 2005; Morton et al., 2005; Segovia et al., 2012; Tossici-Bolt et al., 2006; Towey et al., 2011) as well as to PPMI data (Kuo et al., 2013, 2014; Oliviera and Castelo-Branco, 2015; Prashanth et al., 2014; Zubal et al., 2007). Pioneering studies of automated classification of PPMI images were carried by Zubal, Kuo and co-workers (Kuo et al., 2013, 2014; Zubal et al., 2007) starting in 2007. In their technique, each three-dimensional DaTscan is projected onto a two-dimensional plane by summing voxels along the vertical dimension. A rudimentary striatal "atlas" containing the caudate, the putamen, and the occipital lobe is placed and adjusted on the two-dimensional image. The mean striatal binding ratios in the left and right caudate nuclei and putameni are calculated. In Zubal et al. (2007), these ratios are compared with corresponding ratios calculated from manual tracings, validating the automated placement of the atlas. In Kuo et al. (2013), the smaller of the left and right striatal binding ratios are used in an ROC analysis for classification. In Kuo et al. (2014), the difference between left and right striatal binding ratios is used as a laterality measure and compared with clinical symptoms and visual reads. Similar atlas or template-based approaches using non-PPMI data are Koch et al. (2005); Morton et al. (2005); Tossici-Bolt et al. (2006).

Regional-level support-vector and logistic analysis of the PPMI images was carried out by Prashanth et al. (2014) using the mean striatal binding ratios in the left and right caudate and putamen. Interestingly, the authors find that an interaction term (a product of the binding ratios of the two caudate nuclei) is necessary for accurate logistic classification.

One exception to the region-based analysis, is the voxel-based analysis of PPMI images carried out by Oliviera and Castelo-Branco (2015) using a support-vector machine. Our approach is similar in spirit to this approach, but differs from it in several important aspects: First, support-vector machines provide a binary output while the logistic model provides a probability of classification, which is more nuanced. Second, the support-vector machine used in Oliviera and Castelo-Branco (2015) cannot identify informative voxels, while the logistic model can. The authors of Oliviera and Castelo-Branco (2015) use a post processing step using a voxel-wise z-score to identify voxels in a PD image that differ from corresponding voxels in the control images. This identifies only the most strongly differing voxels; it does not does not identify all voxels that contribute to the classification. In contrast, the logistic lasso model explicitly identifies informative voxels, and only uses the identified voxels for classification. Third, an extension of the logistic formulation that we propose provides a mechanism (LPCs) to understand the source of variation in the data as it pertains to classification. No such formulation is available for support-vector machines. And finally, the logistic model provides a simple mechanism to understand interactions with age, gender, etc. These analyses give significant additional insight into the data.

Machine learning has also been applied successfully to classify controls from PD subjects using non-DaTscan information. An excellent survey of machine learning approaches for PD using voice recordings, MR images, gait patterns etc. can be found in Bind et al. (2015).

Materials and methods

PPMI images

As of April 2016, DaTscans from 658 subjects (210 controls + 448 PD) were available from PPMI and were downloaded for this study. The PD subjects had multiple longitudinal scans, and only the first of these longitudinal scans was used. Controls do not have longitudinal scans; they are scanned only once.

The imaging protocol for the PPMI scans is documented in http://www.ppmi-info.org/wp-content/uploads/2013/02/PPMI-Protocol-AM5-Final-27Nov2012v6-2.pdf. All scans are co-registered and re-sampled in a common image space. A sketch of this procedure is available in Wisniewski et al. (2013) (however, the registration algorithm is not described in detail). The DICOM headers for the images reveal that the images are of size 109×91×91 voxels, with a voxel size of 2 mm×2 mm in the xy plane, a z-slice thickness of 2 mm., and a z-slice spacing of 2 mm. All PPMI images are in MNI space. The age, sex, and handedness of all subjects are available and were also downloaded.

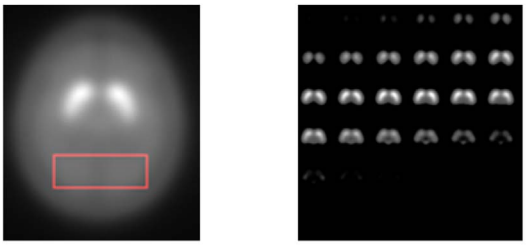
Table 1 summarizes the demographics of the control and PD populations. The proportions of males and females in controls (M=64%, F =36%) are very similar to the proportions in PDs (M=65%, F=35%). A χ^2-test of consistency in a 2×2 table shows no significant difference in the proportions at a p-level of 0.05. Similarly, the age ranges and medians are well matched. The handedness for controls (RH=82%, LH=12%, Ambi=6%) and PDs (RH=86%, LH=11%, Ambi=3%) are also well matched. A χ^2-test of consistency in a 2×3 table shows no significant difference in proportions at a p-level of 0.05.

Preprocessing

All downloaded PPMI images were preprocessed. The first preprocessing step retained those image slices that bracket the striatum

Table 1
Demographics of the Control and PD populations.

	Sex (No. of subjects)	Age (Yrs.)	Handedness (No. of subjects)
Control	M: 137, F: 73	Range: 31–84 Median: 62	RH: 172, LH: 26, Ambi: 12
PD	M: 289, F: 159	Range: 34–85 Median: 63	RH: 307, LH: 39, Ambi: 12



(a) ROI for normalization (b) Masked mean control image

Fig. 1. Preprocessing. (a) One axial slice showing the boundaries of the ROI used for normalization. The ROI extends as a cube throughout the volume. Every voxel in the image is divided by the mean in the ROI. (b) Axial slices through the masked mean control image. The mask loosely envelopes the striatum, the globus palllidus, and the thalamus.

(slices 25–55), the globus pallidum and the thalamus. Next, we calculated the striatal binding ratios (SBR) for all voxels in the retained slices. The SBR in any voxel is defined (Innis et al., 2007) as the ratio of bound tracer in that voxel to non-displaceable tracer in tissue. SBR is calculated by choosing a reference volume in the tissue:

$$\text{SBR in any voxel} = \frac{\text{count in the voxel}}{\text{mean count in the reference volume}} - 1. \tag{1}$$

We drop the −1 term in the above equation because adding or subtracting the same constant from all voxels has no effect on the logistic model or any other calculations reported below. We take the reference volume to be a cube containing the occipital region and excluding the striatum (Fig. 1a). This reference region is similar to the reference regions used in Kuo et al. (2013, 2014); Zubal et al. (2007). Finally, a loose mask is created to contain the striatum, the globus pallidum and the thalamus. All voxels outside the mask are set to zero (Fig. 1b). The mask is created by averaging all control SBR images, thresholding the average image at 0.3 times the maximum voxel value in the average image, and then convolving the thresholded image with a 5×5×5 Gaussian kernel.

Notation and terminology

We fix some notation and terminology before proceeding to describe the main analysis. The image of the ith subject is x_i, i=1, …,N (N=658). All images have the same size; the number of voxels in any image is V (V=60791). If u is a voxel in an image, then $x_i(u)$ is the value of the image x_i in voxel u.

By scanning the voxels in a raster fashion, the image x_i can also be thought of as a V×1 vector. Then, $x_i(u)$ is the uth component of this vector. Conversely, any V×1 vector a can be filled raster fashion into V voxels and displayed as a 3d image. The 3d image has value $a(u)$ in the uth voxel. This can be used to visualize the vector, as we do below.

Each subject belongs to one of two classes – control or PD. The class for the ith subject is y_i, with y_i=0 for control, and y_i=1 for PD.

The training set is a set of M image-class pairs $\{x_i, y_i\}$, $i = 1, \ldots, M$. If a subset of the data is used for training, then $M < N$. The number of controls and PDs in the training set may not be equal, and we let M_0 denote the number of controls and M_1 denote the number of PDs in the training set.

The logistic lasso

The logistic model (McCullagh and Nelder, 1989) gives the conditional probability of y_i given x_i as

$$p(y_i \mid x_i) = \begin{cases} \frac{1}{1+\exp(a^T x_i + b)} & \text{if } y_i = 0 \\ \frac{\exp(a^T x_i + b)}{1+\exp(a^T x_i + b)} & \text{if } y_i = 1, \end{cases} \tag{2}$$

where a is vector of coefficients of size V×1, and b is a scalar. Together, a and b are the parameters of the model. Learning the logistic model is equivalent to estimating a and b from the training set.

Training

Given the M training examples $\{x_i, y_i\}$, $i = 1, \ldots, M$ containing M_0 samples from class 0 and M_1 samples from class 1, the parameters a and b can be estimated by maximizing a penalized, weighted, log-likelihood function

$$\hat{a}, \hat{b} = \underset{a,b}{\operatorname{argmax}} \sum_{i=1}^{M} \alpha_i \log p(y_i \mid x_i) - \lambda \parallel a \parallel_1, \tag{3}$$

where, the weight $\alpha_i = 1/M_0$ if y_i=0, else $\alpha_i = 1/M_1$. In Eq. (3), $\parallel a \parallel_1$ is the L_1 norm of a, $\lambda > 0$ is a scalar, and the $\lambda \parallel a \parallel_1$ term promotes sparsity of a. That is, loosely speaking, the $\lambda \parallel a \parallel_1$ term biases the answer $\hat{a}$ to have many zero components. The components of $\hat{a}$ that are not zero correspond to voxels that are informative. The scalar λ is a free parameter, and is determined by cross-validation. With the addition of the L_1 penalty, the logistic model is referred to as the *logistic lasso* (Tibshirani, 1996).

The objective function to be maximized on the right hand side of Eq. (3) is strictly concave and hence has a single maxima. But the objective function is not differentiable because of the L_1 norm. Nevertheless, the alternating directions method of multipliers (ADMM) can be used to maximize the objective function. Details of ADMM and its use in logistic lasso models can be found in Boyd et al. (2010). We use ADMM to obtain the estimates $\hat{a}$, $\hat{b}$.

Once the estimates $\hat{a}$, $\hat{b}$ are available, they can be substituted for a and b in Eq. (2) to calculate the probabilities for any new image x_i.

Cross validation

We use 10-fold cross validation to choose a value of the parameter λ from a set of possible values. For each value of λ in this set, the following procedure is repeated 10 times: A random 10% of the data is held out and the logistic lasso model of equation (3) is used to estimate $\hat{a}$ and $\hat{b}$ from the remaining 90%. The estimated $\hat{a}$, $\hat{b}$ are used to calculate the log-likelihood of the hold-out data.

The 10 repeats give a mean and a standard deviation of the hold-out log-likelihood for each λ. The maximum value of the mean hold-out log-likelihood is noted, and the values of λ whose mean hold-out log-likelihood are within $\pm$ one standard deviation of the maximum are taken as feasible values of λ. Any value of λ in this range can be used to obtain the final estimate of $\hat{a}$, $\hat{b}$.

Further analysis of the model

The class probabilities of image x_i are determined by the scalar term $a^T x_i + b$. In other words, $a^T x_i + b$ is the scalar "discriminatory feature" of the image which distinguishes controls from PD. Visualizing and analyzing this feature for the PPMI dataset gives additional insight into the control and PD classes. We analyze $a^T x_i + b$ in the following way:

1. The probability that x_i belongs to the PD class increases monotonically with $a^T x_i + b$. This implies that a can be used to visualize the location of dopamine transporter loss. To see how, let $\Omega_+ = \{u \mid a(u) > 0\}$ be the set of voxels where $a(u)$ is positive, and let $\Omega_- = \{u \mid a(u) < 0\}$ be the set of voxels where the $a(u)$ is negative. Then x_i has a higher probability of belonging to the PD class if the weighted average value (weighted by absolute value of a) of x_i in Ω_- is less than the weighted average value of x_i in Ω_+. That is, greater loss of dopamine transporters in Ω_- relative to Ω_+ increases the probability of PD. Visualizing Ω_- in 3d identifies anatomical regions of dopamine transporter loss, and visualizing Ω_+ in 3d identifies the reference

1. Create the $V \times 1$ vector 1_V, according to $1_V(u) = \text{sign}(a(u))$.
2. Convert every image x_i to $\bar{x}_i$ by $\bar{x}_i(u) = \mid a(u) \mid x_i(u)$ for all pixels u in the image (in Matlab notation: $\bar{x}_i = \text{abs}(a). * x_i$).
3. Find ordinary principal components e_k and eigenvalues λ_k of all $\bar{x}_i$ in the given class.
4. Sort the principal components found in step 2 in decreasing order of $\{1_V^T e_k\}^2 \lambda_k$. Let $[k]$, $k = 1, \cdots$ represent this order. Then, $e_{[k]}$ are the logistic principal components (LPCs).
5. Visualize $e_{[k]}$ as 3d images to understand which voxels contribute to the LPCs.

Fig. 2. Calculating logistic principal components. The algorithm for calculating and visualizing the logistic principal components.

region to which Ω_- is compared.

2. The variance of $a^T x_i + b$, calculated for all x_i in the PD class, and separately for all x_i in the control class, indicates how "heterogenous" the discriminatory feature is in each class. A small variance for both classes would suggest that the feature is present (or absent) to the same degree in each class. A larger variance implies a more graded change of the feature within each class. Even more insight can be obtained if the variance of the feature is explained in terms of variances of the voxels in the control and PD images. We do this by developing the notion of *logistic principal components* below.

3. Finally, correlating $a^T x_i + b$ in each class with age, sex, and handedness sheds light on how these variates influence the discriminating feature.

Since b is a constant, we can drop it from $a^T x_i + b$ to carry out all of the above analysis.

Logistic principal components

Ordinary principal components model the covariance of a random variable efficiently, but ordinary principal components may not be efficient at explaining the variance of $a^T x$. To see why, suppose we take images x_i from one class (control or PD) to be samples of a random variable x, and further suppose that e_k and λ_k are the principal components and eigenvalues of the covariance of x. The principal component e_k is a $V \times 1$ vector, and can be thought of as a 3d volume image. Now, since $\text{var}(a^T x) = \sum_k (a^T e_k)^2 \lambda_k$, the amount of variance explained by the principal component e_k depends on $a^T e_k$. We expect a to be sparse, and if e_k happens to take large values in those voxels where a is zero (so that e_k takes small values in those voxels where a is not zero), then terms in the above sum are likely to be small. That is, ordinary principal components may not be efficient at explaining $\text{var}(a^T x)$.

What we want are "principal components" that are more efficient than ordinary principal components at explaining the variance of $a^T x$. To do this, we simply rewrite the inner product as

$$a^T x = \sum_u a(u)x(u) = \sum_u \text{sign}(a(u)) \times \Big|a(u)\Big| x(u) = 1_V^T \bar{x}, \tag{4}$$

where, by slight abuse of notation, we take 1_V to be a $V \times 1$ vector $1_V(u) = \text{sign}(a(u))$, and we take $\bar{x}$ to be a $V \times 1$ vector whose uth component is $\bar{x}(u) = |a(u)|x(u)$ (i.e., in Matlab notation $\bar{x} = \text{abs}(a). * x$). Thus, $\bar{x}$ is a voxel-wise, non-negatively weighted version of x, and $\bar{x}$ is zero for all voxels where a is zero.

Let e_k, λ_k, $k = 1, \ldots, P$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ be the eigenvectors and eigenvalues of the covariance of $\bar{x}$. Then, a simple generative model for $\bar{x}$ is

$$\bar{x} = E[\bar{x}] + \sum_{k=1}^{P} e_k \sqrt{\lambda_k} z_k, \tag{5}$$

where z_k are univariate, zero mean, uncorrelated random variables with unit variance. Moreover,

$$a^T x = 1_V^T \bar{x} = 1_V^T E[\bar{x}] + \sum_{k=1}^{P} 1_V^T e_k \sqrt{\lambda_k} z_k$$

where the random variables $1_V^T e_k \sqrt{\lambda_k} z_k$ are also uncorrelated. Therefore,

$$\text{var}(a^T x) = \sum_{k=1}^{P} \{1_V^T e_k\}^2 \lambda_k, \tag{6}$$

where the right hand side is a sum of variances of uncorrelated random variables. Sorting the terms on the right hand side in decreasing order explains the variance of $a^T x$ in terms of contributions from uncorrelated "sources" $1_V^T e_k \sqrt{\lambda_k} z_k$. We call the sorted e_k, *logistic principal components* (LPC). Each LPC is a $V \times 1$ vector and can be visualized as an image to understand which voxels contribute to the component.

LPCs have two relevant properties: First, because $\bar{x}$ is identically zero for those voxels where a is zero, every LPC is also zero for the same voxels. This avoids the problem with ordinary principal components mentioned at the start of this section. Second, $\bar{x}$ is simply the random variable x scaled with a non-negative scaling at every voxel. Thus any LPC represents variation in the underlying data, albeit scaled at every voxel. These two properties, and the decomposition in Eq. (6) suggest that LPCs can efficiently explain $\text{var}(a^T x)$ in terms of the underlying data.

The logistic component analysis algorithm is displayed in Fig. 2.

Results

Applying the logistic lasso

The logistic lasso model was applied to the preprocessed PPMI DaTscan images in four different ways: First, all of the images were used as training images, with 10-fold cross validation to determine the λ parameter. Then the ADMM algorithm was used to fit the logistic lasso using the cross validated λ to all images. We refer to this as the *all-data* case. Next, the images were divided into three equal sized groups, each group containing the same fraction of control and PD images as the original set. Then, holding back one group at a time as a test set, the remaining two groups were merged to form a training set. This gave three training + test sets. As above, for each training set, 10-fold cross validation was used to determine the λ parameter, and ADMM used to fit the logistic lasso to the training set. The classification accuracy of the fitted model was then evaluated over the test set. We refer to this as the *split-data* case. Finally, all images were summed along the z-axis to create 2d images, and in analogy with the above, *2d*

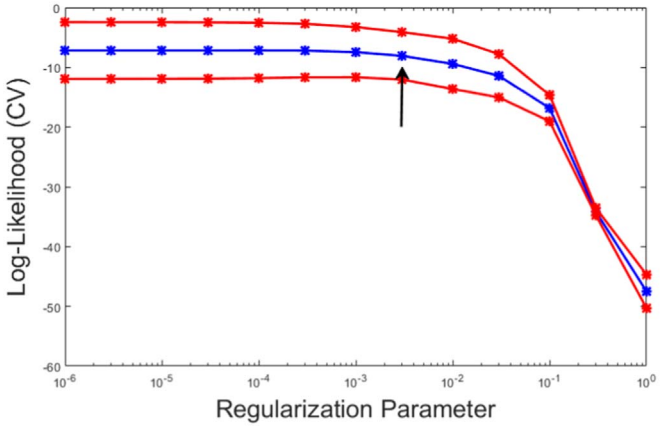


Fig. 3. Cross Validation of λ. The mean log-likelihood (blue curve) ± one standard deviation (red curves) for 10-fold cross validation for fitting the logistic lasso to all PPMI DAT images. The arrow shows the value of λ used in subsequent analysis.

all-data and *2d split-data* cases were created. The 2d cases are similar to Kuo et al. (2013, 2014), Zubal et al. (2007) and were created in order to evaluate the 3d vs. 2d classification performance.

Fig. 3 shows cross validation results for the 3d all-data case. The figure plots the mean (blue curve) ± one standard deviation (red curves) of the hold-out log-likelihood for λ in the range $[10^{-6} - 10^{0}]$. As the figure shows, the mean of the log-likelihood is quite flat till about $\lambda = 3 \times 10^{-3}$, after which it drops off. The initial flat part of the curve suggests that the voxel-wise data in the PPMI DAT images is highly correlated, so that increasing λ (which causes a to be more sparse) does not affect the log-likelihood till $\lambda = 3 \times 10^{-3}$. Since we are interested in finding the minimal set of voxels which are statistically informative, we use $\lambda = 3 \times 10^{-3}$ for all subsequent analysis with the 3d all-data case. Almost identically shaped cross validation curves were obtained for three pairs of training+testing sets in the 3d split-data case as well as the 2*d* all-data and split-data cases. These curves are not shown to conserve space. For all of these cases, the value of λ was chosen as the value at which the initial flat curve begins to drop.

The logistic-lasso model was fit to all cases using the cross validated λs. Fig. 4 shows the logistic lasso PD class probability for all images in the 3d all-data case. For easy comprehension, the subjects are arranged so that the first 210 are controls and the remaining are PD. The 0.5 probability is indicated by a dashed orange line. All images with probabilities below this line are classified as controls, and all images with probabilities above this line are classified as PDs by the model. To conserve space, the figures for the 3d split-data data cases are not shown. The split-data figures are very similar to Fig. 4.

Table 2
Training Errors. FP=False Positive, FN=False Negative, ERR=Net Error.

	3d			2d		
Data Type	FP	FN	ERR	FP	FN	ERR
All Data	1.43%	2.90%	2.43%	4.76%	6.25%	5.78%
Split Data 1	2.16%	3.36%	2.97%	5.76%	6.71%	6.41%
Split Data 2	0.72%	3.02%	2.29%	5.04%	6.38%	5.95%
Split Data 3	2.14%	3.02%	2.74%	5.00%	6.38%	5.95%
Mean ± Std.	1.67% ± 0.83%	3.13% ± 0.20%	2.67% ± 0.35%	5.27% ± 0.43%	6.49% ± 0.19%	6.10% ± 0.27%

Table 3
Test Errors. FP=False Positive, FN=False Negative, ERR=Net Error.

	3d			2d		
Data Type	FP	FN	ERR	FP.	FN	ERR
Split Data 1	2.82%	2.67%	2.71%	7.04%	4.00%	4.98%
Split Data 2	2.82%	6.00%	4.98%	4.23%	10.00%	8.14%
Split Data 3	2.86%	2.67%	2.73%	5.71%	6.67%	6.36%
Mean. ± Std.	2.83% ± 0.023%	3.78% ± 1.92%	3.47% ± 1.31	5.66% ± 1.41	6.89% ± 3.01	6.49% ± 1.58

Table 2 shows the training errors for 3d and 2d all-data and the three split-data cases. The last row of the table shows the mean errors ± standard deviation of the three split-data cases. Table 3 shows the test errors for 3d and 2d split-data cases. The last row of Table 3 shows the mean errors ± standard deviation of the three split-data cases.

The results in Tables 2 and 3 can be summarized thus:

1. The mean training errors for the 3d split-data cases are approximately 0.24% higher than the all-data case. This is close to, or within, one standard deviation of the split-data cases. This holds for the 2d cases as well: the split-data cases have higher mean errors, but the differences between the all-data and split-data cases are close to, or within, one standard deviation of the split-data cases.
2. The mean 3d split-data test errors exceed the training errors by 1.16% or less. Except for false positives, the differences between training and test errors are within one standard deviation of the test errors. Also, the mean 3d split-data test errors exceed the all-data training errors by 1.4% or less.

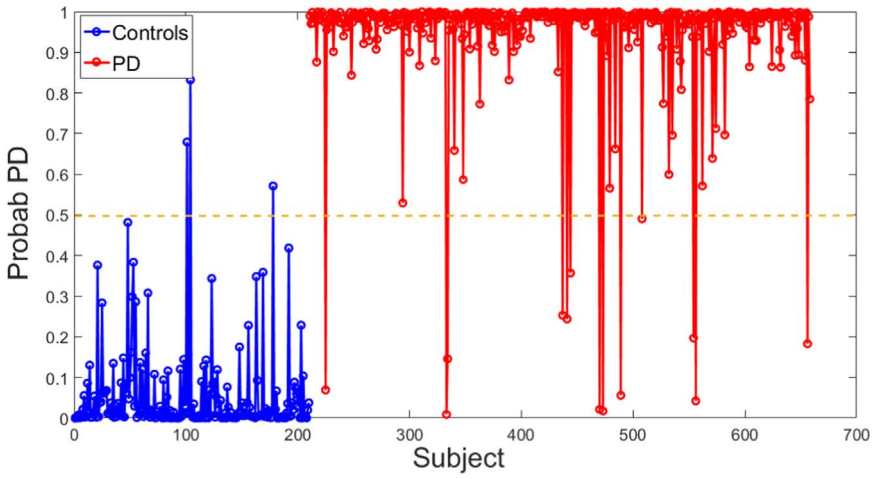


Fig. 4. PD probability. The probability of belonging to the PD class according to the logistic lasso model for the all-data case. The first 210 subjects are controls, the rest are PD. The dashed, horizontal line is probability=0.5. For this case, the false positive rate was 1.43%, the false negative rate was 2.90%, and the overall net error was 2.43% (see Table 2).

For the 2d cases, the mean split-data test errors exceed the training errors by 0.4% or less. All of the training-test differences are within one standard deviation of the test errors. The mean 2d split-data test errors exceed the all-data training errors by 0.9% or less. These differences too are within one standard deviation of the test errors.

3. Finally, the 2d all-data errors and mean of the 2d split-data errors are significantly greater than the corresponding 3d errors. In all cases, the 2d errors are greater by a factor of 1.82 or more. In most cases, the 2d errors are greater by a factor of 2 or more.

The 3d results are significantly better than the 2d results, consequently we only focus on 3d for the rest of the paper.

Similarity of all-data and split-data results

As reported above, the all-data errors and the split-data errors are similar to each other. There are other similarites as well: The same images tend to be misclassified in both cases. Further, visualizing $\hat{a}$ for both cases as a volume shows that $\hat{a}$ is very similar for the 3d all-data and the three split-data cases.

Given this similarity between the all-data and split-data cases, below we present additional analysis of the 3d all-data case only. This has the advantage that only one set of results need be presented rather than three, and all images contribute to the conclusions drawn from the analysis. From now on, we simply refer to this case as the all-data case.

Visualization of Informative voxels

To visualize informative voxels found by the logistic lasso, a high resolution MRI T1 structural image template (Holmes et al., 1998) in MNI space was segmented using FreeSurfer 5.3 http://surfer.nmr.mgh.harvard.edu. Since PPMI images are already in MNI space, the T1 image and its segmentation served as an atlas for displaying the logistic lasso informative voxels. The results are displayed in Fig. 5a–b. Fig. 5a shows axial slices which are displayed in raster fashion. They are also referred to by numbering them in raster fashion. Thus, the topmost slice is the left-most image on the top row and referred to as slice 1. Subsequent slices go from left to right, and from the top row to the bottom row. The last slice, slice 25, is the slice at the bottom-right. As an aid to remembering the numbering convention, note that there 5 slices in each row. Thus, the top row has slices 1–5, the second row has slices 6–10, etc.

The colored voxels in Fig. 5a correspond to the FreeSurfer segmentation and are taken as atlas regions: the pale purple region is the caudate, the pale green region is the putamen and the pale yellow is the globus pallidus.

Fig. 5b shows Ω_+ and Ω_- overlayed on the atlas. Recall that Ω_+ and Ω_- are voxels where $\hat{a}$ takes positive and negative values respectively. Ω_+ voxels are rendered in red and Ω_- voxels are rendered in blue. The colormap strips shown in the right of the figure indicate how the value of $\hat{a}$ is converted to color.

Note that Ω_+ extends from the second slice through till the 21st slice. Slices 2–15 show clearly that Ω_+ occupies only a part of the caudate. Further, slices 14–21 clearly show that Ω_+ contains voxels in the globus pallidus. (Note that boundaries of the some of the atlas regions are occluded by Ω_+ and Ω_- in Fig. 5b. Fig. 5a is provided to be a handy reference for these occluded boundaries.).

Slices 10–16 show that Ω_- overlaps with the putamen, but Ω_- does not contain all of the putamen. Most of Ω_- occupies the posterior voxels of the putamen.

Fig. 5b clearly shows that only a subset of voxels in the caudate and putamen are informative. Furthermore, some voxels in the globus pallidus are also informative. The algorithm suggests that loss of dopamine transporter in the identified voxels of the putamen compared to the identified voxels in the caudate+globus pallidus is indicative of PD.

Visual inspection of misclassified images

Next, we turned to visually examining images that were misclassified by the algorithm. There were 3 misclassified control images (the false positives) and 13 misclassified PD images (the false negatives). One of the authors (LS) is a radiologist with expertise in reading DaTscans, and the misclassified images were visual inspected by him and manually classified. LS was blinded to the true class of the images, and also to the fact that these images were misclassified. All 3 misclassified control images were visually identified as normals, but of the 13 misclassified PD images, only 2 were identified as PD. Visually, the other 11 images strongly resembled control images. This is not surprising. A small percent of patients meeting the clinical criteria for PD do not show dopaminergic deficit on SPECT scans (Varrone et al., 2013). Thus, at least some of the misclassifications are not because of algorithm limitations, but because some images are unusual.

Comparison with ROI analysis

How much is the classification improved by using voxel-based methods over ROI-methods? Recall that all images are already registered to each other in the PPMI dataset, and that the Freesurfer atlas was registered by us to the mean control image. The predefined regions in the Freesurfer atlas were used as ROIs; specifically, the left- and right-caudate and the left- and right-putamen were used as defined in the Freesurfer atlas. Then, the symmetric and asymmetric differences between the average SBR in the caudate and putamen were calculated as follows:

First, the union of the left- and right-caudate ROIs was taken as a single caudate ROI. Similarly, the union of the left- and right-putamen ROIs was taken as a single putamen ROI. Then the difference: mean SBR in the caudate ROI − mean SBR in the putamen ROI was calculated. We call this the symmetric difference between the caudate and putamen.

Second, the difference between the mean SBR of left-caudate and left-putamen, and the difference between the mean SBR of right-caudate and right-putamen were calculated. Of these two differences, the one with the largest magnitude was taken as the asymmetric difference.

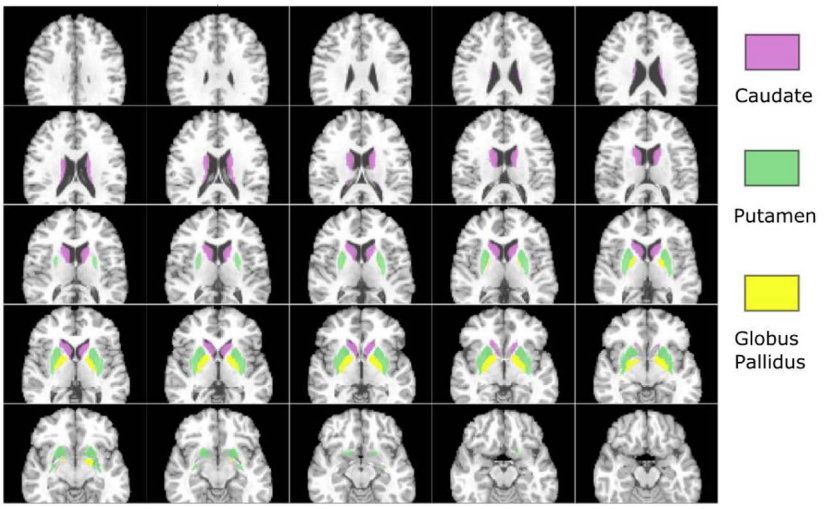
The symmetric difference calculates the net difference between the mean SBRs in the caudate and putamen ROIs, and is blind to hemispheric asymmetry in disease. The asymmetric difference uses the maximum of the difference between caudate and putamen in the two hemispheres and is sensitive to hemispheric difference.

Fig. 6 shows ROCs calculated by using our logistic lasso 3d all-data model and using the ROI based symmetric and asymmetric differences. The area under the curve for all three ROCs (AUROC) are also shown in the figure. The ROCs clearly show the advantage of using voxel-based processing over ROI-based processing.

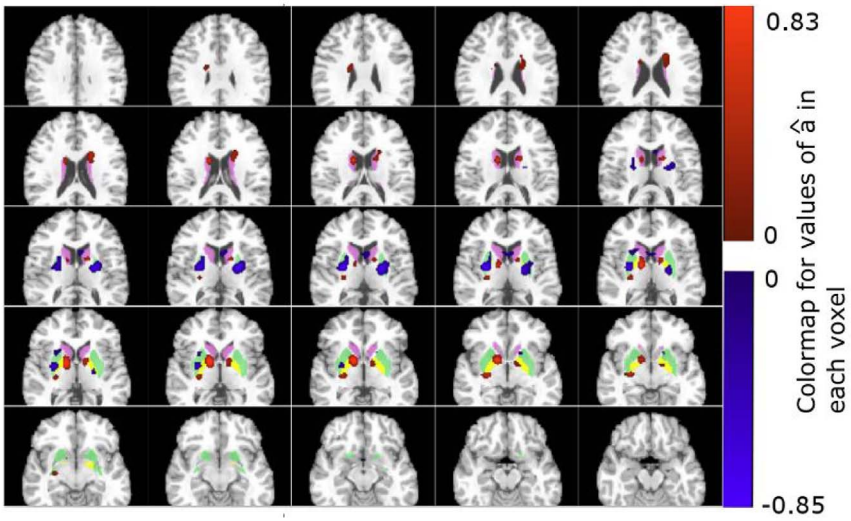
Logistic principal components

Fig. 7 shows histograms of $\hat{a}^T x_i + b$ for controls as well as PD subjects. Both histograms have significant variance, suggesting that the feature $\hat{a}^T x_i + b$ varies within both groups. It is useful to understand the source of this variance via logistic principal component (LPC) analysis. Table 4 shows the fraction of variance of $\hat{a}^T x_i + b$ explained cumulatively by the first n LPCs as a function of n. The table shows that the first two LPCs explain almost 95% of the variance for controls as well as PDs.

Recall that each LPC is a $V\times 1$ and can be rendered as a volume image. Fig. 8 shows control LPCs rendered as volumes. Fig. 9 shows PD LPCs. Both figures mimic Fig. 5: the LPCs are rendered on the Freesurfer atlas and the T1 image. The atlas regions are rendered pale purple, pale green, and pale yellow as before. The LPC voxels are



(a) Freesurfer atlas regions superimposed on a T1-structural image. Pale Purple: caudate, Pale green: putamen, Pale yellow: globus pallidus.



(b) Ω_+ and Ω_- overlayed on the above atlas. Ω_+ voxels are rendered in red. Ω_- voxels are rendered in red. The color bar at the right indicates the values of $\hat{a}$ in Ω_+ and Ω_-.

Fig. 5. Informative voxels found by the algorithm overlayed on the Freesurfer atlas. (a) Freesurfer atlas regions superimposed on a T1-structural image. Pale Purple: caudate, Pale green: putamen, Pale yellow: globus pallidus. (b) Ω_+ and Ω_- overlayed on the above atlas. Ω_+ voxels are rendered in red. Ω_- voxels are rendered in red. The color bar at the right indicates the values of $\hat{a}$ in Ω_+ and Ω_-.

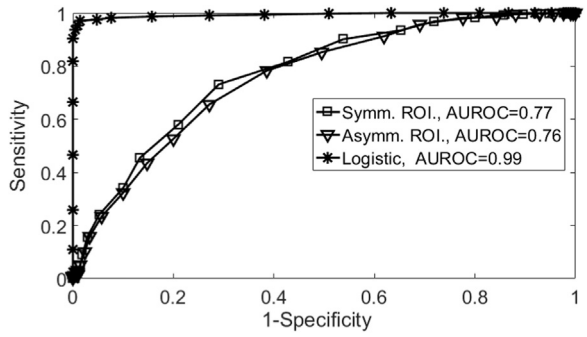


Fig. 6. ROCs. The ROCs for the 3d all-data logistic model, and the symmetric and asymmetric difference based on ROIs.

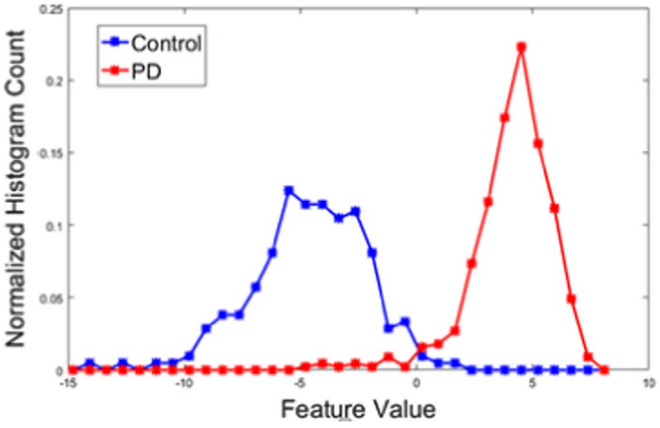


Fig. 7. Histograms of $\hat{a}^T x_i + b$ for control and PD.

Table 4
Fractional variance of $a^T x + b$ explained cumulatively by the first n logistic principal components.

	Fractional variance explained by 1 to n components				
	n=1	n=2	n=3	n=4	n=5
Control	0.82	0.95	0.98	0.99	1.00
PD	0.54	0.93	0.99	1.00	1.00

rendered red when the LPC voxel has a positive sign, and blue when it has a negative sign. The color maps in the right part of the figures show how voxel values relate to color.

Control LPCs

Almost all voxels of the first control LPC (Fig. 8a) are positive. This LPC actually has three, isolated, very mildly negative voxels. They do not appear to be significant and are not rendered in the figure.

A slice by slice comparison shows that the first control LPC is remarkably similar to the lasso coefficients of Fig. 5b if we ignore the sign of the lasso coefficients. This suggests that the dominant component in the control variance is due to simultaneous (i.e. "in phase") brightening/dimming of all informative voxels.

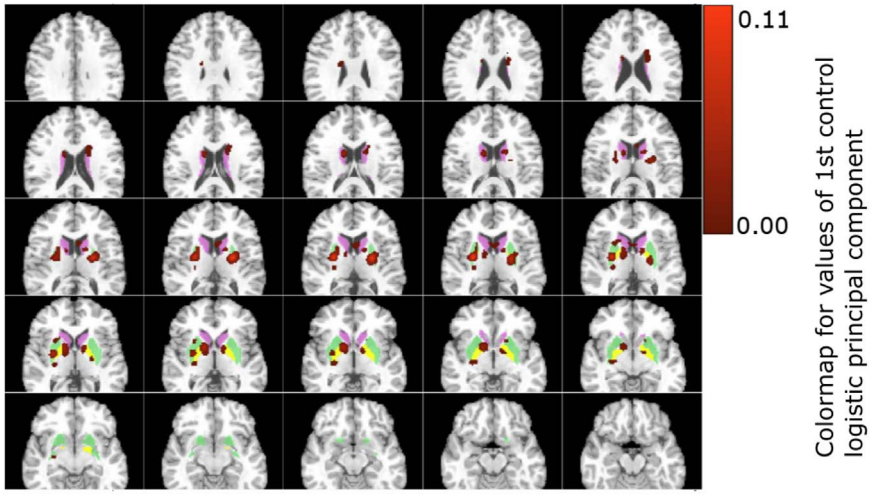
The second control LPC has positive voxels in the caudate and globus pallidus, and negative voxels in the putamen. Thus this LPC suggests that the second largest source of variance in controls is variable "contrast" between the caudate+globus pallidus and the putamen.

PD LPCs

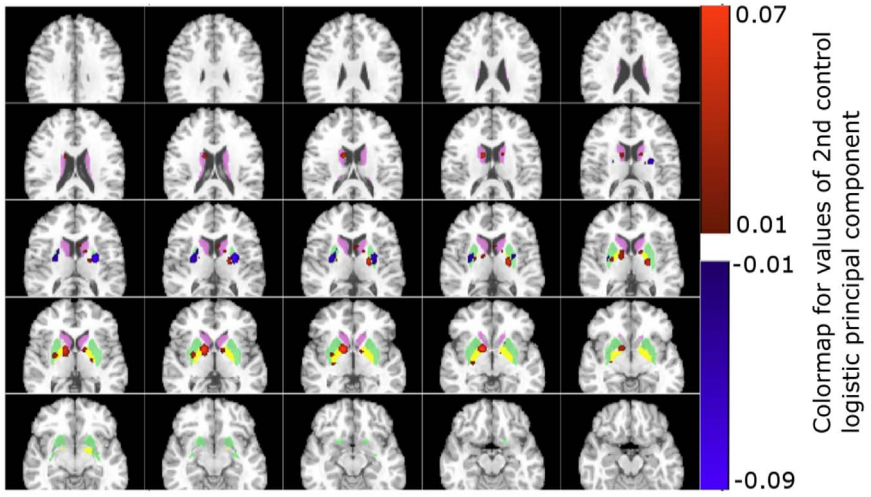
A slice by slice comparison of the PD LPCs and the control LPCs reveals that the first PD LPC is similar to the second control LPC but with a larger range of negative values. The second PD LPC is similar to the first PD. This suggests that the sources of variance in the PD category are similar to the sources of variance in the control category. But, in PDs, the variable "contrast" between the caudate+globus pallidus and the putamen is a more significant source of variance than simultaneous brightening/dimming of voxels.

Interactions

Fig. 10a-c show the interactions of $\hat{a}^T x_i + b$ with handedness, sex, and age. The values of $\hat{a}^T x_i + b$ are plotted in blue for controls and in red for PD. Tables 5a–c summarize the interactions.

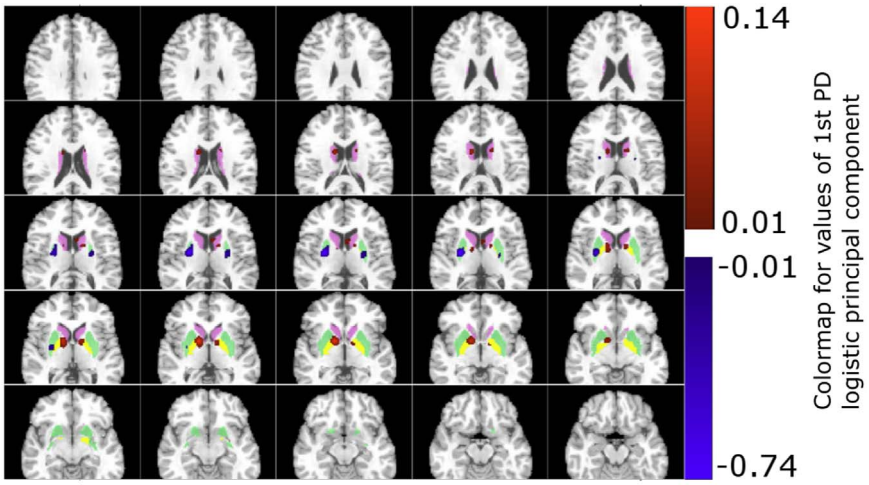


(a) First logistic principal component for controls.

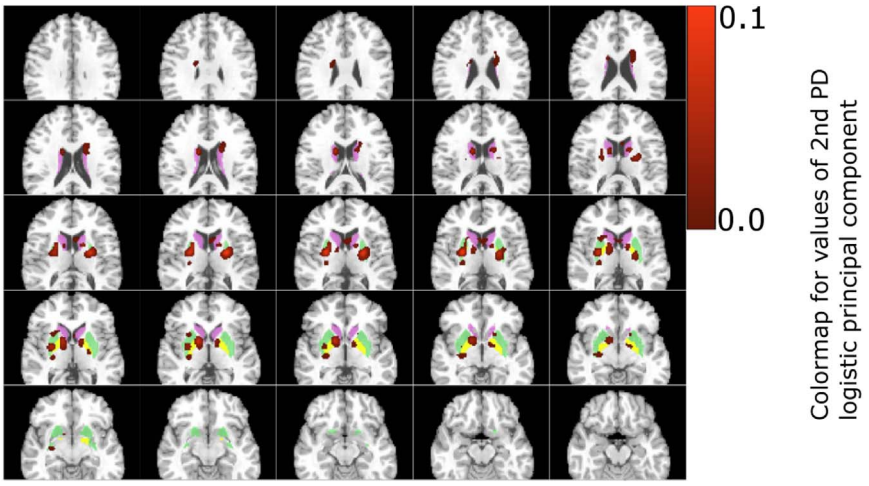


(b) Second logistic principal component for controls.

Fig. 8. Logistic principal components for controls. The components are displayed using the colormaps on the right and the T1-image and atlas of Fig. 5a. (a) The first component. The first component is mostly positive (there are three, isolated, mildly negative voxels which are not rendered). (b) The second component has positive and negative components. The positive components are localized in or near the caudate and the globus pallidus. The negative components are in or near the putamen.



(a) First logistic principal component for PDs.



(b) Second logistic principal component for PDs.

Fig. 9. Logistic principal components for PDs. The components are displayed using the colormaps on the right and the T1-image and atlas of Fig. 5a. (a) The first component. (b) The second component. The first component is similar to the second control LPC of Fig. 8b, and the second component is similar to the first control LPC of Fig. 8a.

Fig. 10a shows the scatter plot of $\hat{a}^T x_i + b$ vs. handedness of the subject for controls and PD. Table 5a shows the means and standard deviations of $\hat{a}^T x_i + b$ for controls and PDs. T-tests, carried out assuming equal as well as unequal variances, show that the differences between the means of RH, LH, and Ambi are not significant at the 0.05 level for controls as well as PDs. F-tests for equality of variance show that the difference between the variances for controls are not significant at the 0.05 level. The differences in the variances for PD are significant at the 0.05 level between RH-LH and LH-Ambi.

Fig. 10b shows the scatter plot of $\hat{a}^T x_i + b$ vs. sex for control and PD subjects. Table 5b shows the means and standard deviations of $\hat{a}^T x_i + b$ for the same. T-tests, assuming equal as well as unequal variances, and F-tests show that the means as well as the variances of control females are significantly different from control males at the 0.05 level. On the other hand, the means and variances of PD females are not significantly different from PD males at the 0.05 level.

Fig. 10c shows the interaction of $\hat{a}^T x_i + b$ with age for controls and PD subjects. The figure shows straight lines that are linear regressions of $\hat{a}^T x_i + b$ with age for controls and PDs. Table 5c shows that $\hat{a}^T x_i + b$ has a positive correlation coefficient with age for controls, and a negative correlation coefficient with age for PDs. Both correlation coefficients are significantly different from 0 at the 0.05 level. As Fig. 10c shows, the separation between the two classes decreases with increasing age.

Relation to MDS-UPDRS Part III scores

How does the logistic feature relate to clinical features of PD? To address this, we turned to the MDS-UPDRS Part III scores of controls and PDs. These scores are available from PPMI.

The Part III scores contains 36 scores/ratings. Two of these scores were set aside (“Constancy of rest – Did these movements interfere with ratings?”, and “Hoehn and Yahr Stage”), and the remaining 34 ratings were summed to get a score that we refer to as the *total movement score* (TMS) for each individual. Higher TMS indicates greater movement disorder. The “Constancy of rest – Did these movements interfere with ratings?” score was not used because it was unscored for all individuals. Presumably, there was no interference with ratings. The Hoehn and Yahr Stage was set aside for a separate analysis, reported below.

Fig. 11a shows a scatter of the logistic feature $\hat{a}^T x_i + b$ vs. TMS for controls and PDs. Scatter data for controls is in blue, while scatter data for PDs is in red. The scatter plot has two large clusters – one to the left, containing controls with TMSs close to 0, and another cluster to the right, containing PDs with TMSs that are significantly higher. There is also a scattering of data in between the two clusters.

To analyze the relation between the logistic feature and MS in more detail, we grouped controls and PDs into four classes:

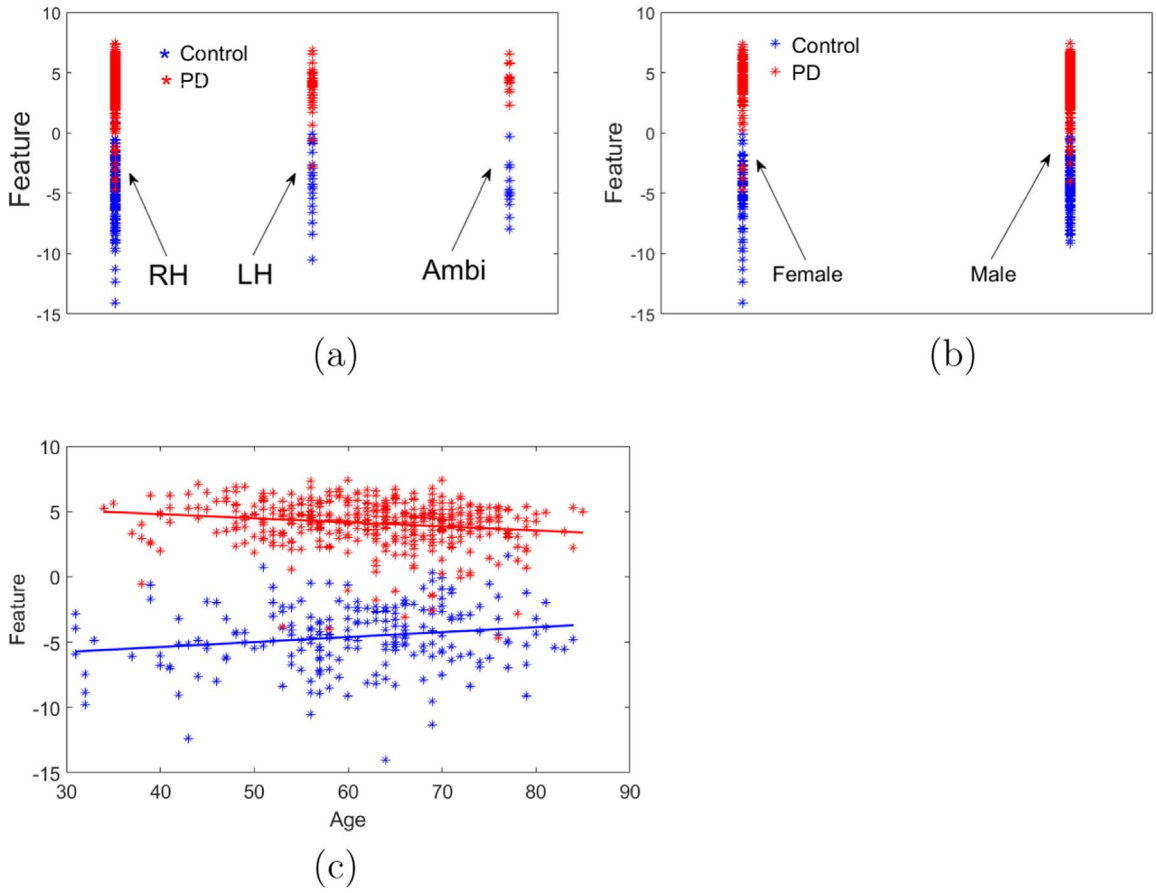


Fig. 10. Interaction of the logistic feature $\hat{a}^T x_i + b$ with handedness, sex, and age. Controls plotted in blue, PD in red. (a) Scatter plots of the feature values for controls and PD for right-handed (RH) subjects, left-handed subjects (LH) and ambidextrous (Ambi.) subjects. (b) Interaction with sex. There is significant difference between the control males and females, but not between PD males and females. (c) Interaction with age. The control and PD populations approach each other with increasing age.

1. Class 1: All controls whose probability of belonging to the PD class, as estimated by the logistic model, is less than or equal to 0.2. These are typical control images.
2. Class 2: All controls whose probability of belonging to the PD class, as estimated by the logistic model, is greater than 0.2. These are atypical control images.
3. Class 3: All PDs whose probability of belonging to the PD class, as estimated by the logistic model, is less than or equal to 0.8. These are atypical PD images.
4. Class 4: All PDs whose probability of belonging to the PD class, as estimated by the logistic model, is greater than 0.8. These are typical PD images.

Table 5
Interaction with Handedness, Sex, and Age.

	Cntrl mean ± std. dev.	PD mean ± std. dev.
(a) Interaction with Handedness		
Right Handed	−4.63 ± 2.46	4.13 ± 1.76
Left Handed	−4.23 ± 2.70	3.62 ± 1.78
Ambidextrous	−4.69 ± 2.05	4.49 ± 1.14
	Cntrl mean ± std. dev.	PD mean ± std. dev.
(b) Interaction with Sex		
Male	−4.27 ± 2.21	4.00 ± 1.71
Female	−5.15 ± 2.80	4.25 ± 1.82
	Cntrl	PD
(c) Interaction with Age		
Corr. Coeff.	0.17	−0.18

The logistic probabilities of 0.2 and 0.8 correspond to logistic feature values of −1.39 and 1.39. These boundaries are shown in Fig. 11a as vertical, dotted, orange and green lines respectively. Thus, Class 1 is all control images to the left of the vertical line at −1.39, Class 2 is all control images to the right of the vertical line at at −1.39. Class 3 is all PD images to the left of the vertical line at 1.39, and Class 4 data is all PD images to the right of the vertical line at 1.39.

Fig. 11b shows the cumulative probability of TMS for each class. That is, for any value, say t, on the x-axis, the y-axis of Fig. 11b shows the fraction of data in each class with TMS less than or equal to t. The cumulative probabilities for Classes 1 and 2 are very similar. On the other hand, the cumulative probability curve for Class 4 is shifted to the right of the curve for Class 3 almost everywhere. For any score t on the x-axis, Class 3 has a greater fraction of population with scores less than or equal to t. In other words, subjects with atypical PD images, having classification probabilities less than or equal to 0.8, show lower TMS than subjects with typical PD images having classification probabilities greater than 0.8.

Fig. 12 shows the scatter plot of the logistic feature vs. the Hoehn and Yahr (HY) Stage for PDs. It appears that there is a greater incidence of HY Stage 2 with higher logistic feature value. This is quantified in Table 6 which shows the fraction of data points with HY Stages 1,2 and 3 for Class 3 and 4 as defined above. Class 4 has a greater fraction of HY Stage 2 data than Class 3.

Discussion

We now turn to discussing implications of the above results, starting with the logistic lasso results of Tables 2 and 3 and the visualization of logistic lasso coefficients in Fig. 5. But, a word of

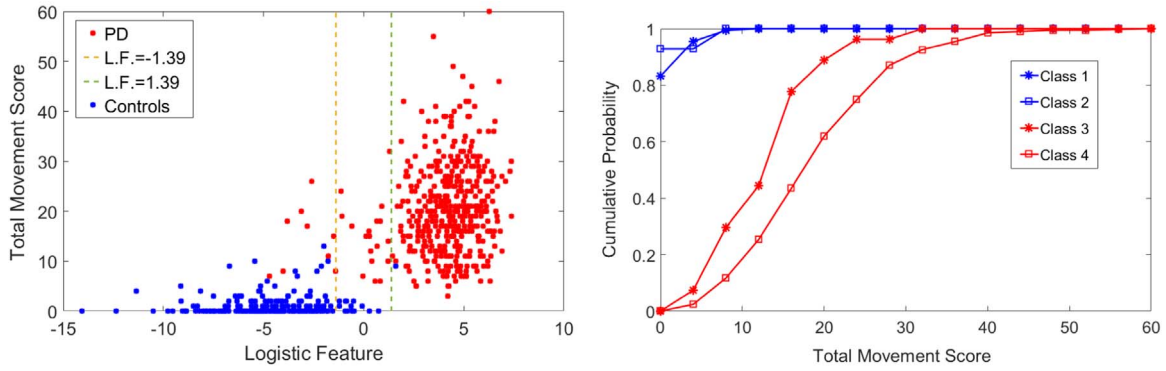


Fig. 11. The relation of logistic features to the total movement score. Total movement score (TMS) is a sum of all MDS-UPDRS Part III scores for an individual (excluding two scores, as discussed in the text). (a) Scatter plot of logistic features vs. TMS for controls (blue) and PDs (red). Vertical, dotted, orange and green lines indicate logistic feature boundaries at −1.39 and +1.39 respectively. These boundaries are used to define four classes, as discussed in the text. (b) Cumulative probabilities of TMS for the four classes. The cumulative probability curve for class 4 is shifted to the right of the curve for class 3, showing that subjects in class 4 have a higher TMS than subjects in class 3.

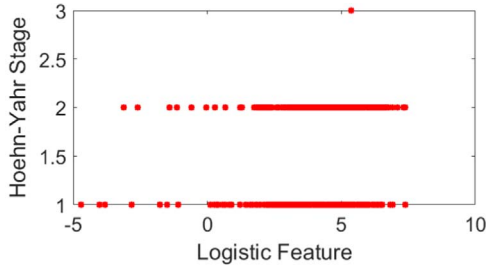


Fig. 12. Scatter plot of the logistic feature vs. Hoehn and Yahr Stage for PDs.

Table 6
Hoehn and Yahr Stage population fractions for PD Class 3 and Class 4.

	Frac.(Stage=1)	Frac.(Stage=2)	Frac.(Stage=3)
Class 3	0.63	0.37	0.0
Class 4	0.49	0.51	0.0 (0.0024)

caution before we proceed. The logistic lasso model is focused on calculating classification probabilities. All of the above results, as well as the discussion below, should be interpreted only in the context of classification. The results and the discussion do not bear a more general interpretation.

The ROCs of Fig. 6 clearly show that the voxel-based analysis is more optimal than ROI based-analysis. However, it may well be possible to improve the ROI results by using more sophisticated techniques.

As mentioned in the literature review, support-vector analysis of the PPMI data set using ROIs has been carried out in Prashanth et al. (2014). And their results show that high discrimination is not obtained by ROI data unless an interaction term is included. No such term is needed for the voxel-based analysis. Support for voxel-level analysis also comes from the visualization in Fig. 5 which shows that good classification can be obtained from sub-regional data in the right and left caudate, pallidum, and putamen. Our analysis suggests that pooling the pallidum with the caudate nuclei (which together define the "red" voxels in Fig. 5b) is statistically more robust than just using voxels from the caudate. Voxels in the pallidum have often been ignored in the past.

Because PPMI images are available registered to a common space, the effect of potential misregistration on the logistic lasso is a concern. The effect of misregistration is likely to be stronger in those voxels that are close to the caudate, putamen, and pallidum boundaries. These voxels are easily identified in Fig. 5a. Unfortunately, details of the registration algorithm and estimates of registration error are not available for the PPMI DaTscan images (Wisniewski et al., 2013), and thus its effect on the logistic lasso is difficult to quantify.

It is also useful to note that two characteristics of the DaTscan data, and a property of the lasso algorithm, tend to ameliorate the effect of misregistration: First, the striatal and extra-striatal structures we consider are close to the center of the image, and this reduces the effect of rotation and scaling errors in registration. Second, the signal in neighboring voxels in a DaTscan image is highly correlated. This also reduces the effect of small misregistration. Finally, the sparsity prior acts to eliminate poorly registered voxels, because poor registration increases variance (variance calculated at a voxel across different subjects). If the variance is large (compared to the mean), then the sparsity prior is likely to eliminate the voxel.

Turning to LPC analysis, note that it provides additional insights. The similarity of the first two LPCs for controls and PDs is striking, and suggests similar underlying causes for the variance of $\hat{a}^T x_i + b$ in both groups. As mentioned before, the first LPC for controls (and the 2nd LPC for PDs) suggests a simultaneous brightening/dimming of voxels. Two potential mechanisms explain this. The first explanation is that it is likely that the striatal binding ratio in the informative voxels has an intrinsic variation within the control population, and furthermore, this variation is not completely wiped out by PD. The second explanation is that the occipital region normalization used in the striatal binding ratio is not perfect. This is not to suggest that the occipital normalization is useless; it does provide a useful overall normalization. Normalization in nuclear imaging is a surprisingly complex topic. In analyzing FDG-PET images, for example, results are sensitive to the normalization strategy (e.g. global normalization vs. cerebellar normalization) (Guenther et al., 2011; Martino et al., 2013). Similar observations about normalization have also been made for SPECT imaging (Soonawala et al., 2002). Whether occipital lobe normalization is optimal for DaTscans, or a better normalization exists appears to be an unexplored area.

The second control LPC (and the first PD LPC) suggests a varying "contrast" between the dopamine transporter levels in the caduate +pallidum and the putamen. For the control group, there are two possible explanations: First, the contrast variation may simply be due to intrinsic differences between individual subjects. Second, age may play a role. It is likely that as the control population ages, the differential contrast in dopamine levels changes naturally. As recorded in Table 1, controls have a wide age range, 31 − 84 years.

The above two explanations are valid for the PD group as well, but an additional factor may also be responsible: it is possible that at the time of the scan, the PD population was in different stages of the disease causing a additional variable degree of dopaminergic neuronal loss. This may account for the larger negative values in the first PD LPC, as compared to the second control LPC.

The lack of interaction with handedness is interesting, but should be interpreted cautiously. Our results do not say that there is no

interaction of the disease with handedness. The results only say that, as far as classification into groups is concerned, a set of voxels can be chosen that have no interaction with handedness. Whether handedness plays a role in PD is a more complex matter (Scherfler et al., 2012).

The interaction with sex, especially the difference between the female and male controls is striking. This difference has been noted in the literature before (Haaxma et al., 2007; Varrone et al., 2013). Two possible explanations for the difference, offered in Varrone et al. (2013), are that on an average females have a slightly smaller volume for the striatum, so that if the net amount of dopamine transporter is more or less equal for females and males, then control females will exhibit a higher dopamine transporter concentration than control men. Alternately, it may be that women express dopamine transporters at a higher level.

The interaction with age is as expected. There is loss of dopaminergic neurons in the normal aging population (van Dyke et al., 2002), and the aging controls and PDs can be expected to approach each other. The fact that the correlation coefficients for controls and PD in Table 5c are negatives of each other is a result of using the logistic model. The classification boundary in the logistic model is $\hat{a}^T x + b = 0$, and for high separability the two populations are mapped approximately symmetrically around this line.

Can the classification rate be improved by taking interactions into account? We added interaction terms to the logistic analysis, but this did not improve the classification performance. Along similar lines, one can ask whether accounting for laterality improves classification performance? Parkinson's disease is known to present and progress asymmetrically in the two hemispheres. This laterality is manifest in DaTscans and can be quantified (Kuo et al., 2014). The logistic model can be easily modified to account for laterality, but we found that the modified model does not improve classification performance. Others too have observed this; the authors of Oliviera and Castelo-Branco (2015) comment that the support-vector machine classification performs slightly poorly when laterality is taken into account. This too should be interpreted in the narrowest sense: It does not imply that laterality is not important for understanding the disease, only that the logistic lasso and support-vector machines do not explicitly need to model laterality to perform well.

The relation between the logistic feature and the Total Movement Scores suggests that the logistic feature does correlate with clinical observations. Especially interesting are the observations – in Fig. 11b and Table 6 - that PD Class 3 has lower total motion scores and lower percentage of Hoehn and Yahr Stage 2 subjects than Class 4. These observations demonstrate the advantage of using a logistic model, which gives a continuous estimate of probability, over a support-vector machine model which gives only a binary classification.

The relation between the logistic feature and the total movement scores is quite intriguing, and it suggests that a voxel-wise analysis that is focused on finding additional such relations (rather than focused on classification) is likely to be fruitful. We hope to address this in the future.

This paper is meant to be methodological, but in the supplementary information we apply the logistic lasso to another classification problem: that of classifying subjects with scans without evidence of dopaminergic deficit (SWEDDs) from controls and PDs. As expected, SWEDDs are difficult to distinguish from controls (based only on DaTscan images), but SWEDDs can be distinguished from PDs.

Conclusion

In conclusion, 3d voxel-wise logistic analysis of the PPMI control and PD population provides accurate classification. The analysis shows that sub-regional voxels in the caudate, the globus pallidus, and the putamen are informative for classification.

Logistic principal component analysis reveals two uncorrelated sources which explain most of the variance of the logistic feature. Finally, there are significant interactions of the logistic feature with sex (for controls) and with age, but not with handedness.

Acknowledgements

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.

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and multiple funding partners. The full list of PPMI funding partners can be found at ppmi-info.org/fundingpartners.

We would like to add that the research presented in this paper was not support by any grant from PPMI, the Michael J. Fox foundation, their funding partners, or any other agency.

We are grateful to the staff of MNI Imaging for help with downloading PPMI data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2017.02.067.

References

Benamer, T.S.1., 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. Mov. Disord. 15 (May (3)), 503–510.

Bind, S., Tiwari, A.K., Sahani, A.K., Koulibaly, P.M., Nobili, F., Pagani, M., Sabri, O., Borght, T.V., Laere, K.V., Tatsch, K., 2015. A survey of machine learning based approaches for Parkinson disease prediction. Int. J Comp. Sci. Inform. Technol. 6 (2), 1648–1655.

Booij, J., Knol, R.J.J., 2007. SPECT imaging of the dopaminergic system in (premotor) Parkinson's disease. Park. Relat. Disord. 13, S425–S428.

Buchert, R., Berding, G., et al., 2006. IBZM Tool: a fully automated expert system for evaluation of IBZM SPECT studies. Eur. J. Nucl. Med. Mol. Imaging 33 (September (9)), 1073–1083.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2010. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3 (1), 1–122.

Darcourt, J., Booij, J., Tatsch, K., Varrone, A., Vander Borght, T., Kapucu, O.L., Någren, K., Nobili, F., Walker, Z., Van Laere, K., 2010. EANM procedure guidelines for brain neurotransmission SPECT using (123)I-labelled dopamine transporter ligands, version 2. Eur. J. Nucl. Med. Mol. Imaging 37 (February (2)), 443–450.

Guenther, T., Küntzelmann, A., Haberkorn, U., Essig, M., Schröder, J., Schönknecht, P., 2011. Comparison between global and cerebellar intensity normalization of FDG PET images in patients with mild cognitive impairment or Alzheimer's disease. Pharmacopsychiatry, 21–A45.

Haaxma, C.A., Bloem, B.R., Borm, G.F., Oyen, W.J.G., Leenders, K.L., Eshuis, S., Booij, J., Dluzen, D.E., MWIM, Horstink., 2007. Gender differences in Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 78, 819–824.

Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C., 1998. Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 22, 324–333.

⟨http://surfer.nmr.mgh.harvard.edu/⟩.

⟨http://www.ppmi-info.org/wp-content/uploads/2013/02/PPMI-Protocol-AM5-Final-27Nov2012v6-2.pdf⟩.

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. Med. Phys. 39 (October (10)), 5971–5980.

Innis, R.B., Cunningham, V.J., Delforge, J., Fujita, M., Gjedde, A., et al., 2007. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. J. Cereb. Blood Flow Metab. 27 (September (9)), 1533–1539.

Koch, W., Radau, P.E., Hamann, C., Tatsch, K., 2005. Clinical testing of an optimized software solution for an automated, observer-independent evaluation of dopamine transporter SPECT studies. J. Nucl. Med. 46 (July (7)), 1109–1118.

Kuo, P.H., Avery, R., Krupinski, E., Hong, L., Bauer, A., Sherman, S., McMillan, N., Seibyl, J., Zubal, G., 2013. Receiver-operating-characteristic analysis of an automated program for analyzing striatal uptake of ^{123}I-Ioflupane SPECT images: calibration using visual reads. J. Nucl. Med. Technol. 41, 26–31.

Kuo, P.H., Hong, L.H., Avery, R., Krupinski, E., Bauer, A., Sherman, S., McMillan, N., Seibyl, J., Zubal, G.I., 2014. Evaluation of an objective striatal analysis program for determining laterality in uptake of ^{123}I-Ioflupane SPECT: comparison to clinical symptoms and to visual reads. J. Nucl. Med. Technol. 42, 1–4.

Martino, M.E., de Villoria, J.G., Lacalle-Aurioles, M., Olazarán, J., Cruz, I., Navarro, E., García-Vázquez, V., Carreras, J.L., Desco, M., 2013. Comparison of different

methods of spatial normalization of FDG-PET brain images in the voxel-wise analysis of MCI patients and controls. Ann. Nucl. Med. 27 (August (7)), 600–609.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models 2nd Ed.. Chapman and Hall.
Morton, R.J., Guy, M.J., Clauss, R., Hinton, P.J., Marshall, C.A., Clarke, E.A., 2005. Comparison of different methods of DatSCAN quantification. Nucl. Med. Commun. 26 (December (12)), 1139–1146.
Oliviera, F.P.M., Castelo-Branco, M., 2015. Computer-aided diagnosis of Parkinson's disease based on [^{123}I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines. J. Neural Eng. 12, 026008.
Prashanth, R., Roy, S.D., Mandal, P.K., Ghosh, S., 2014. Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. Exp. Syst. Appl. 41, 3333–3342.
Rajput, A.H., Sitte, H.H., Rajput, A., Fenton, M.E., Pifl, C., HornyKiewicz, O., 2008. Globus pallidus dopamine and Parkinson motor subtypes. Neurology 70, 1403–1410.
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. Med. Phys. 39 (July (7)), 4395–4403.
Sánchez-González, M.A., García-Cabezas, M.A., Rico, B., Cavada, C., 2005. The primate thalamus is a key target for brain dopamine. J. Neurosci. 25 (26), 6076–6083, (June 29).
Scherfler, C., Seppi, K., Mair, K.J., Donnemiller, E., Virgolini, I., Wenning, G.K., Poewe, W., 2012. Left hemispheric predominance of nigrostriatal dysfunction in Parkinson's disease. Brain 135 (November (Pt 11)), 3348–3354.
Soonawala, D., Amin, T., Ebmeier, K.P., Steele, J.D., Dougall, N.J., Best, J., Migneco, O., Nobili, F., Scheidhauer, K., 2002. Statistical parametric mapping of (99 m)Tc-HMPAO-SPECT images for the diagnosis of Alzheimer's disease: normalizing to cerebellar tracer uptake. Neuroimage 17 (November (3)), 1193–1202.
Tatsch, K., Poepperl, G., 2013. Nigrostriatal dopamine terminal imaging with dopamine transporter SPECT: an update. J. Nucl. Med. 54 (August (8)), 1331–1338.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58 (1), 267–288.
Tossici-Bolt, L., Hoffmann, S.M., Kemp, P.M., Mehta, R.L., Fleming, J.S., 2006. Quantification of [123I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio. Eur. J. Nucl. Med. Mol. Imaging 33 (December (12)), 1491–1499, (Epub 2006 July 21).
Toweya, D.J., Bain, P.G., Nijran, K.S., 2011. Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images. Nucl. Med. Commun. 32 (August (8)), 699–707.
van Dyke, C.H., Seibyl, J.P., Malison, R.T., Laruelle, M., Zogbi, S.S., Baldwin, R.M., Innis, R.B., 2002. Age related decline in dopamine transporters. Am. J. Geriatr. Psychiatry 10, 36–43.
Varrone, A., Dickson, J.C., Tossici-Bolt, L., Sera, T., Asenbaum, S., Booij, J., Kapucu, O.L., Kluge, A., Knudsen, G.M., Koulibaly, P.M., Nobili, F., Pagani, M., Sabri, O., Borght, T.V., Laere, K.V., Tatsch, K., 2013. European Multicenter Database of Healthy Controls for [^{123}I] FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. Eur. J. Nucl. Med. Mol. Imaging 40, 213–227.
Wisniewski, G., Seibyl, J., Marek, K., August 2013. DatScan SPECT Image Processing Methods for Calculation of Striatal Binding Ratio (SBR). Parkinson's Progression Markers Initiative.
⟨www.ppmi.org⟩.
Zubal, G.I., Early, M., Yuan, O., Jennings, D., Marek, K., Seibyl, J.P., 2007. Optimized automated striatal uptake analysis applied to SPECT brain scans of Parkinson's disease patients. J. Nucl. Med. 48, 857–864.