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

Hemant D. Tagare⁎, Christine DeLorenzo, Sudhakar Chelikani, Lawrence Saperstein, Robert K. Fulbright

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
1053-8119/ © 2017 Elsevier Inc. All rights reserved.
important. Such an algorithm is \textit{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 (LPCs). 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.

\textit{Previous work}

Imaging holds considerable promise in evaluating pre-motor PD, assessing disease progression, and in differential diagnosis. Excellent reviews are available in Booj 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 applied; 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).

\textit{Materials and methods}

\textit{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-AMS-Final-27Nov2012v6-2.pdf. All scans are co-registered and resampled 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 \( \chi^2 \)-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 \( \chi^2 \)-test of consistency in a 2×3 table shows no significant difference in proportions at a p-level of 0.05.

\textit{Preprocessing}

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

<p>| Table 1: Demographics of the Control and PD populations. |
|-------------------------|-----------------|----------------------|</p>
<table>
<thead>
<tr>
<th>Sex (No. of subjects)</th>
<th>Age (Yrs.)</th>
<th>Handedness (No. of subjects)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median: 62</td>
<td></td>
</tr>
<tr>
<td>M: 289, F: 159</td>
<td>Median: 63</td>
<td></td>
</tr>
</tbody>
</table>
(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:

\[
\text{SBR in any voxel} = \frac{\text{count in the voxel}}{\text{mean count in the reference volume}} - 1. \quad (1)
\]

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 pallidus, 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 \(i\)th subject is \(x_i\), \(i=1, \ldots, N\) (\(N=658\)). 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 \(V\times1\) vector. Then, \(x_i(u)\) is the \(u\)th component of this vector. Conversely, any \(V\times1\) 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 \(u\)th 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 \(i\)th 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

\[
p(y_i | x_i) = \begin{cases} 
\frac{1}{1 + \exp(-a^T x_i + b)} & \text{if } y_i = 0 \\
\frac{\exp(a^T x_i + b)}{1 + \exp(a^T x_i + b)} & \text{if } y_i = 1, 
\end{cases} \quad (2)
\]

where \(a \in \mathbb{R}^V\) is a vector of coefficients of size \(V\times1\), 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

\[
\hat{a}, \hat{b} = \arg \max_{a, b} \sum_{i=1}^{M} a_i \log(\gamma_i) - \lambda \|a\|_1, \quad (3)
\]

where, the weight \(a_i = 1/M_0\) if \(y_i=0\), else \(a_i = 1/M_1\). In Eq. (3), \(|a|_1\) is the \(L_1\) norm of \(a\), \(\lambda > 0\) is a scalar, and the \(\lambda \|a\|_1\) term promotes sparsity of \(a\). That is, loosely speaking, the \(\lambda \|a\|_1\) 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 \(\lambda\) 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 \(\lambda\) from a set of possible values. For each value of \(\lambda\) 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 \(\lambda\). The maximum value of the mean hold-out log-likelihood is noted, and the values of \(\lambda\) whose mean hold-out log-likelihood are within one standard deviation of the maximum are taken as feasible values of \(\lambda\). Any value of \(\lambda\) 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 \(\Omega_a = \{a(u) > 0\}\) be the set of voxels where \(a(u)\) is positive, and let \(\Omega_c = \{a(u) < 0\}\) be the set of voxels where \(a(u)\) 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 \(\Omega_a\) is less than the weighted average value of \(x_i\) in \(\Omega_c\). That is, greater loss of dopamine transporters in \(\Omega_a\) relative to \(\Omega_c\) increases the probability of PD. Visualizing \(\Omega_a\) in 3d identifies anatomical regions of dopamine transporter loss, and visualizing \(\Omega_c\) in 3d identifies the reference region associated with the normal transporter value.

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

3. The probability that \(x_i\) belongs to the PD class increases with \(b\). This implies that \(b\) can be used to visualize the location of dopamine transporter loss. To see how, let \(\Omega_b = \{b > 0\}\) be the set of voxels where \(b\) is positive, and let \(\Omega_l = \{b < 0\}\) be the set of voxels where \(b\) 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 \(b\)) of \(x_i\) in \(\Omega_b\) is greater than the weighted average value of \(x_i\) in \(\Omega_l\). That is, greater gain of dopamine transporters in \(\Omega_b\) relative to \(\Omega_l\) increases the probability of PD. Visualizing \(\Omega_b\) in 3d identifies anatomical regions of dopamine transporter gain, and visualizing \(\Omega_l\) in 3d identifies the reference region associated with the normal transporter value.
1. Create the $V \times 1$ vector $l_V$, according to $l_V(u) = \text{sign}(a(u))$.
2. Convert every image $x_i$ to $\tilde{x}_i$ by $\tilde{x}_i(u) = a(u) x_i(u)$ for all pixels $u$ in the image (in Matlab notation: $\tilde{x}_i = \text{abs}(a) \ast x_i$).
3. Find ordinary principal components $e_k$ and eigenvalues $\lambda_k$ of all $\tilde{x}_i$ in the given class.
4. Sort the principal components found in step 2 in decreasing order of $(e_k \lambda_k)^2$. 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.

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 $\lambda_k$ 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 $\text{var}(a^T x) = \sum(a e_k)^2 \lambda_k$, 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

$$a^T x = \sum_{a} a(u) x(u) = \sum_{a} \text{sign}(a(u)) \times a(u) x(u) = \tilde{l}_V^T \tilde{x},$$

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

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

$$x = E^T \tilde{x} + \sum_{k=1}^{P} e_k \sqrt{\lambda_k} z_k,$$

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

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

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

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

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” $\tilde{l}_V^T e_k \sqrt{\lambda_k} z_k$. 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 $\tilde{x}$ 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, $\tilde{x}$ 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 $\lambda$ parameter. Then the ADMM algorithm was used to fit the logistic lasso using the cross validated $\lambda$ 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 $\lambda$ 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

302
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 $\lambda$ in the range $[10^{-6} \text{ to } 10^0]$. 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 $\lambda$ (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 2d all-data and split-data cases. These curves are not shown to conserve space. For all of these cases, the value of $\lambda$ 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 $\lambda$s. 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 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.

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.
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 similarities as well: The same images tend to be misclassified in both cases. Further, visualizing \( a \) for both cases as a volume shows that \( a \) 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 PMPI 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 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 \( \Omega_l \) and \( \Omega_r \) overlayed on the atlas. Recall that \( \Omega_l \) and \( \Omega_r \) are voxels where \( a \) takes positive and negative values respectively. \( \Omega_l \) voxels are rendered in red and \( \Omega_r \) voxels are rendered in blue. The colormap strips shown in the right of the figure indicate how the value of \( a \) is converted to color.

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

Slices 10–16 show that \( \Omega_r \) overlaps with the putamen, but \( \Omega_l \) does not contain all of the putamen. Most of \( \Omega_l \) 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 \( a^T x_i + b \) for controls as well as PD subjects. Both histograms have significant variance, suggesting that the feature \( a^T x_i + b \) 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 \( a^T x_i + b \) 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 significantly different between the two hemispheres and is sensitive to hemispheric difference.
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) $\Omega_+$ and $\Omega_-$ overlayed on the above atlas. $\Omega_+$ voxels are rendered in red. $\Omega_-$ voxels are rendered in red. The color bar at the right indicates the values of $\hat{a}$ in $\Omega_+$ and $\Omega_-$.  

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 $\hat{a}^2 z_i + b$ for control and PD.
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 $a^T x + b$ with handedness, sex, and age. The values of $a^T x + b$ are plotted in blue for controls and in red for PD. Tables 5a–c summarize the interactions.

<table>
<thead>
<tr>
<th>Fractional variance explained by 1 to n components</th>
<th>n=1</th>
<th>n=2</th>
<th>n=3</th>
<th>n=4</th>
<th>n=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.82</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>PD</td>
<td>0.54</td>
<td>0.93</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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

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.
Fig. 10a shows the scatter plot of $l^x + T_i$ vs. handedness of the subject for controls and PDs. Table 5a shows the means and standard deviations of $l^x + T_i$ 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 $l^x + T_i$ vs. sex for control and PD subjects. Table 5b shows the means and standard deviations of $l^x + T_i$ 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 $l^x + b$ with age for controls and PD subjects. The figure shows straight lines that are linear regressions of $l^x + b$ with age for controls and PDs. 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.

**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 $l^x + b$ 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. 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.
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.

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. 11 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 $-1.39$. 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.

Table 5: Interaction with Handedness, Sex, and Age.

<table>
<thead>
<tr>
<th>Interaction with Handedness</th>
<th>Ctrl mean ± std. dev.</th>
<th>PD mean ± std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Right Handed</td>
<td>-4.63 ± 2.46</td>
<td>4.13 ± 1.76</td>
</tr>
<tr>
<td>Left Handed</td>
<td>-4.23 ± 2.70</td>
<td>3.62 ± 1.78</td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>-4.69 ± 2.05</td>
<td>4.49 ± 1.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Interaction with Sex</th>
<th>Ctrl mean ± std. dev.</th>
<th>PD mean ± std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-4.27 ± 2.21</td>
<td>4.00 ± 1.71</td>
</tr>
<tr>
<td>Female</td>
<td>-5.15 ± 2.80</td>
<td>4.25 ± 1.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Interaction with Age</th>
<th>Corr. Coeff.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctrl PD</td>
<td>0.17</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

Table 5 shows the fraction of data points with HY Stage 2 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
The similarity of the first two LPCs for controls and PDs is striking, and the first PD LPC suggests a varying 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 without interaction with handedness. Whether handedness plays a role in PD is a more complex matter (Scherrer 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 \( a^T x + b = 0 \), 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 Oliveira 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 supported by any grant from PPMI, the Michael J. Fox foundation, their funding partners, or any other agency.

We are grateful to the staff of MRI 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


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...