Journal of Neuroscience Methods 222 (2014) 230–237

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Clinical Neuroscience

# Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy

C. Salvatore[a], A. Cerasa[b], I. Castiglioni[c,*], F. Gallivanone[c], A. Augimeri[b], M. Lopez[d], G. Arabia[e], M. Morelli[e], M.C. Gilardi[c], A. Quattrone[b,e]

[a] Department of Physics, University of Milan – Bicocca, Piazza della Scienza 3, 20126 Milan, Italy
[b] Neuroimaging Research Unit, Institute of Neurological Sciences, National Research Council, Germaneto, CZ, Italy
[c] Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), via F.lli Cervi 93, 20090 Segrate, MI, Italy
[d] DITEN, University of Genoa, Via Opera Pia 11A, 16145 Genoa, Italy
[e] Institute of Neurology, University "Magna Graecia", Germaneto, CZ, Italy

## HIGHLIGHTS

- The algorithm allows individual differential diagnosis of PD and PSP by means of MR images.
- The algorithm does not require a priori hypotheses of where useful information may be coded in the images.
- Classification accuracy was significantly higher compared to other published methods.
- The algorithm was able to obtain voxel-based morphological biomarkers of PD and PSP.

## ARTICLE INFO

*Article history:*
Received 1 July 2013
Received in revised form
14 November 2013
Accepted 17 November 2013

*Keywords:*
Support Vector Machine (SVM)
Parkinson's disease (PD)
Progressive Supranuclear Palsy (PSP)
Magnetic resonance imaging (MRI)
Machine learning

## ABSTRACT

*Background:* Supervised machine learning has been proposed as a revolutionary approach for identifying sensitive medical image biomarkers (or combination of them) allowing for automatic diagnosis of individual subjects. The aim of this work was to assess the feasibility of a supervised machine learning algorithm for the assisted diagnosis of patients with clinically diagnosed Parkinson's disease (PD) and Progressive Supranuclear Palsy (PSP).
*Method:* Morphological T1-weighted Magnetic Resonance Images (MRIs) of PD patients (28), PSP patients (28) and healthy control subjects (28) were used by a supervised machine learning algorithm based on the combination of Principal Components Analysis as feature extraction technique and on Support Vector Machines as classification algorithm. The algorithm was able to obtain voxel-based morphological biomarkers of PD and PSP.
*Results:* The algorithm allowed individual diagnosis of PD versus controls, PSP versus controls and PSP versus PD with an Accuracy, Specificity and Sensitivity > 90%. Voxels influencing classification between PD and PSP patients involved midbrain, pons, corpus callosum and thalamus, four critical regions known to be strongly involved in the pathophysiological mechanisms of PSP.
*Comparison with existing methods:* Classification accuracy of individual PSP patients was consistent with previous manual morphological metrics and with other supervised machine learning application to MRI data, whereas accuracy in the detection of individual PD patients was significantly higher with our classification method.
*Conclusions:* The algorithm provides excellent discrimination of PD patients from PSP patients at an individual level, thus encouraging the application of computer-based diagnosis in clinical practice.


## 1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting millions of people worldwide. The primary objective in the clinical practice of PD is to achieve an individual differential diagnosis, in order to tailor the best individual treatment. PD clinical diagnosis is particularly prone to errors (Tolosa et al., 2006), as an array of motor symptoms can also

* Corresponding author at: IBFM-CNR, via F.lli Cervi 93, 20090 Segrate, MI, Italy. Tel.: +39 02 21717511; fax: +39 02 21717558.
*E-mail addresses:* christian.salvatore@unimib.it (C. Salvatore), a.cerasa@unicz.it (A. Cerasa), isabella.castiglioni@ibfm.cnr.it (I. Castiglioni), miriamlp@ugr.es (M. Lopez).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.11.016

be present in other parkinsonian conditions such as Progressive Supranuclear Palsy (PSP). For instance, PSP patients are clinically similar to PD patients, however, are less responsive to treatment and they have a more rapid disease progression. Among the various forms of parkinsonisms, PSP is one of the most difficult to clinically disentangle from idiopathic PD, particularly in early disease stages, when the typical clinical signs are not clearly evident (Gelb et al., 1999; Litvan et al., 1996).

To date, individual diagnosis of PSP is predominantly based on patients' clinical history where standard brain magnetic resonance imaging (MRI) protocols are routinely employed only to exclude concomitant diseases, resulting in poor diagnostic accuracy, sensitivity and specificity, given that images are only visually inspected (Tolosa et al., 2006; Jankovic, 2008). In the past 20 years, a considerable effort has been put into the development of advanced neuroimaging processing techniques in order to identify neuroimaging biomarkers which could be then used for enhancing the diagnostic confidence of clinical diagnosis. Although significant results have been obtained (Shi et al., 2013), most studies have reported differences between patients and controls only at a group level, thus with a very limited translation to an individual diagnosis in more clinical settings. For this reason, attention has recently been directed toward alternative approaches to the analyses of neuroimaging data.

In the last few years, there has been a growing interest within the neuroimaging community in classification methods, including machine-learning algorithms. These techniques are based on algorithms able to automatically extract multiple information from image sets without requiring a priori hypotheses of where this information may be coded in the images. The main aim of these methods is to classify individual structural or functional brain images by maximizing the distance between groups of images. Several studies have assessed the diagnostic value of these techniques, e.g. for the diagnosis of the Alzheimer's disease (Kloppel et al., 2008; Magnin et al., 2009) and Mild Cognitive Impairment (Teipel et al., 2007), and have showed promising findings.

The aim of this study is to implement a supervised machine-learning method able to perform individual differential diagnosis of PD and PSP by means of structural T1-weighted MRIs. This method was based on Principal Component Analysis (PCA) underlying the feature extraction technique (Habeck et al., 2008; Salas-Gonzalez et al., 2010) and on Support Vector Machines (SVMs) for classification purposes (Scholkopf and Smola, 2001; López et al., 2011). In order to identify potential MRI-related biomarkers useful for the diagnosis of PD and PSP, we also generated image maps of pattern distribution of brain structural differences, which reflect the importance of each image voxel for the SVM classification.

## 2. Materials and methods

### 2.1. Clinical and MRI studies

In this retrospective study we enrolled 56 patients and 28 healthy control subjects. All study procedures and ethical aspects were approved by the institutional review board. Written informed consent was obtained from all subjects.

All patients and the healthy control subjects were examined by neurologists with more than ten years of experience in movement disorders. Age at onset, disease duration and severity of symptoms, as assessed by the Unified Parkinson's Disease Rating Scale (UPDRS), and the Hoehn–Yahr (H&Y) stage, were recorded. The Mini Mental State Examination (MMSE) was used to assess general cognitive status. The group of 56 patients consisted of 28 patients with clinically diagnosed PD (Gelb et al., 1999) and 28 patients with clinical diagnosis of probable or possible PSP (Litvan et al., 1996). The healthy control subjects had no history of neurologic or psychiatric diseases, with normal neurological examinations. The 28 healthy control subjects were of similar age as both patient groups.

One brain structural MRI study was performed for each subject by a 1.5-T unit (Signa NV/I; GE Medical Systems, USA). MRI data were acquired using a 3D T1-weighted spoiled gradient echo sequence with the following parameters: TR = 15.2 ms; TE = 6.7 ms; flip angle = 15°; FOV = 24 cm. Slice thickness was of 1.2 mm and each slice had a resolution of 256 × 256 pixels. A T1-weighted 3D dataset was obtained for each subject. Motion artifacts were negligible for all scans by visual inspection.

### 2.2. The machine-learning method

A machine-learning method able to perform individual differential diagnosis of PD and PSP by means of structural T1-weighted Magnetic Resonance Images (MRI) was implemented. This method was based on Principal Component Analysis (PCA) as feature extraction technique and on Support Vector Machines (SVMs) as classification algorithm (Scholkopf and Smola, 2001; López et al., 2011). Both these phases were implemented using the Matlab platform (Matlab version R2011b, The MathWorks, Natick, MA).

#### 2.2.1. Image pre-processing

Original datasets were cropped, re-oriented and converted from DICOM format to 3D NIfTI format using the dcm2nii tool included in the MRICron software (http://www.mccauslandcenter.sc.edu/mricro/mricron/). After that, the pre-processing procedure mainly consisted of 2 steps: (1) skull stripping, which was achieved using the BET tool of the FSL 4.1 software (Smith et al., 2004; Jenkinson et al., 2012), and (2) normalization to MNI space, which was performed by co-registration to the MNI template (MNI152_T1_1mm_brain) (Grabner et al., 2006) included in the FSL 4.1 software.

Images were then imported into the Matlab platform using the 'Tools For NIfTI And ANALYZE Image' toolbox (http://www.mathworks.com/matlabcentral/fileexchange/8797). No smoothing or segmentation were applied. Resulting images were limited within a bounding box. Final whole-brain volumes consisted of 145 × 178 × 133 voxels.

#### 2.2.2. Feature extraction

Feature extraction was implemented by applying spatial transformations on the images in order to reduce data dimensions without losing relevant information.

We used PCA to extract the most significant features from our MRI datasets (Habeck et al., 2008; Alvarez et al., 2009; Salas-Gonzalez et al., 2010). PCA is a standard technique which consists in applying an orthogonal transformation to a dataset of (possibly) correlated variables to obtain a set of values of linearly uncorrelated (orthogonal) variables. These values are called 'principal components' of the original dataset.

The application of PCA involves a second step: the projection itself, which reduces the original number of features to a much lower number of so-called PCA coefficients. These coefficients are the ones used for classification.

Mathematically, let us define $X$ to be a dataset of 3D brain images $X_i$, where $i$ varies from 1 to $N$, $N$ being the number of images in the dataset. Let's suppose that each image $X_i$ is given in the form of a vector of dimension $V$ (in our study $V$ is the total number of voxels of each image, $V = 145 \times 178 \times 133$), so that the dimension of $X$ is $N \times V$, and that the dataset $X$ has zero mean (in case dataset $X$ had non-zero mean, then the average $X_M$ would be subtracted from each

image $X_i$). Now, PCA-space is defined as the space which is spanned by the eigenvectors of the covariance matrix $C$ of the dataset $X$:

$$C = X \cdot X^{T}$$

Finally, PCA coefficients can be extracted by projecting each image onto the PCA-space (Alvarez et al., 2009).

Application of PCA to a given dataset results in a number of principal components which is at most equal to the number of the lower dimension of the data matrix −1. If $N$ is the number of subjects in the dataset, there will only be $N-1$ eigenvectors (principal components) with non-zero eigenvalues. The other eigenvectors have a zero eigenvalue associated, so it does not make sense to consider them.

Once the original data are projected onto PCA coefficients or scores, these coefficients (low dimension representation of the samples) and associated labels can be considered to understand which principal components are more discriminative. For this purpose principal components were ordered in a decreasing order, according to their Fisher Discriminant Ratio (FDR):

$$\mathrm{FDR} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

where $\mu_i$ and $\sigma_i^2$ denote the mean and the variance of the $i$-th class, respectively.

#### 2.2.3. Classification algorithm

We used a classification algorithm based on Support Vector Machines (SVMs) (Scholkopf et al., 2002; Vapnik, 1995, 1998, 1999). Let's suppose that we have a set of training data each one consisting of a pair: an input vector $x_n \in \mathrm{R}^N$, $n = 1, \ldots N$ and the corresponding label or target value $t_n \in \{+/-1\}$. The aim of a SVM is to estimate a decision function which will correctly classify unseen examples $(x, t)$. To do this, SVMs design the optimal separating hyper-plane in terms of distance between classes, that is, a hyper-plane which is maximally distant from the two classes to be separated (this is why it is also called maximal margin hyper-plane). The decision function for an unseen example $x$ is then given by

$$y(x) = \sum_{n=1}^{N} a_n \cdot t_n \cdot k(x, x_n) + b$$

where $N$ is the number of samples in the training set, $a_n$ is a weight constant, $k(x, x_n)$ is a kernel function and $b$ is a threshold parameter. This decision function returns the predicted class $y(x)$ of the unseen example $x$. In this way, this hyper-plane is able to discriminate binary labeled training datasets and to subsequently classify unseen data as belonging to one or the other of the two training classes.

In our work, the SVM classifier was implemented using algorithms of the biolearning toolbox of Matlab. Datasets were divided into three binary labeled groups: PD versus Controls, PSP versus Controls and PSP versus PD. The training step was performed using the principal components and the associated labels of each dataset as features for the SVM (López et al., 2011). A linear kernel was chosen as the more general form and for the purpose of improving computational efficiency. The training of the classifier was carried out for a number $i$ of principal components ranging from 1 to PC, where PC is the whole number of extracted principal components.

### 2.3. Validation of the classifier

Validation of the classification algorithm was performed by a cross-validation strategy. In cross-validation, the data is partitioned into complementary subsets: the training set and the testing set. The training of the classifier is performed on the training set; the validation of the classifier is performed on the testing set. Multiple rounds of cross-validation can then be performed using different partitions. In this study we used Leave-One-Out (LOO) validation, that is a particular case of cross-validation in which the testing set is made up of only one sample of the dataset and the training set is made up of the remaining samples of the dataset $(N-1)$. In this way, all samples in the dataset can be tested in turn if the total number of rounds equals the number of samples in the dataset. LOO is a widely used validation approach because it returns an almost unbiased estimate of the probability of test error of a classification algorithm (e.g. Vapnik, 1998; Chapelle et al., 1999).

Moreover, the use of the same number of subjects for training and classification is somewhat controversial. Given the fair amount of subjects available in our study, we also adopted another validation method using $N/2$ randomly chosen subjects for training the classifier and the remaining N/2 subjects for testing.

Validation was carried out by the two methods for each of the three binary labeled groups and for a number $i$ of principal components ranging from 1 to PC, where PC is the whole number of extracted principal components ($N-1$ or $N/2$). Accuracy, Specificity and Sensitivity rates were computed for each binary labeled group over the first $i$ principal components as follows:

$$\mathrm{Accuracy}_i = \frac{T_{\mathrm{RC}}}{T}$$

$$\mathrm{Specificity}_i = \frac{X_{\mathrm{RC}}}{X_{\mathrm{RC}} + Y_{\mathrm{WC}}}$$

$$\mathrm{Sensitivity}_i = \frac{Y_{\mathrm{RC}}}{Y_{\mathrm{RC}} + X_{\mathrm{WC}}}$$

$T$ being the total number of classified images; $T_{\mathrm{RC}}$ being the total number of classified images that underwent Right Classification (RC); $X_{\mathrm{RC}}$ being the number of images belonging to the first binary labeled group that underwent Right Classification (RC); $X_{\mathrm{WC}}$ being the number of images belonging to the first binary labeled group that underwent Wrong Classification (WC); $Y_{\mathrm{RC}}$ being the number of classified images belonging to the second binary labeled group that underwent RC; $Y_{\mathrm{WC}}$ being the number of images belonging to the second binary labeled group that underwent WC.

Overall Mean Accuracy, Specificity and Sensitivity rates were obtained as mean values calculated over a number of principal components ranging from 1 to PC, where PC is the whole number of extracted principal components.

The dependence of Accuracy, Specificity and Sensitivity on the number of principal components was studied.

$\mathrm{Accuracy}_{>80}$, $\mathrm{Specificity}_{>80}$ and $\mathrm{Sensitivity}_{>80}$ rates (mean, minimum and maximum values) were calculated over a range of principal components for which Accuracy, Specificity and Sensitivity fell above 80% for each of the three binary labeled groups.

### 2.4. Voxel-based pattern distribution

Using the whole number of subjects in the dataset (28 PD, 28 PSP and 28 Controls) and for each binary labeled group, we generated image maps of voxel-based pattern distribution of brain structural differences, which reflect the importance of each image voxel for the SVM classification. This allowed the identification of MRI-related biomarkers useful for the diagnosis of PD and PSP.

During the training step, the SVM classifier calculates a specific weight for each of the images, reflecting the importance of that image for binary separation. This weight is non-zero only for support vectors, while its sign is positive for support vectors belonging to the first binary labeled group and negative for support vectors belonging to the second binary labeled group.

**Table 1**
Demographic and clinical data of enrolled subjects.

| Variables | HC | PD | PSP | *p* values |
|---|---|---|---|---|
| N° | 28 | 28 | 28 | |
| Gender (% males) | 54% | 54% | 64% | |
| Age (years) | 67.5 ± 7.1 | 68.2 ± 5 | 69.4 ± 5.7 | n.s. |
| Disease duration (years) | – | 8.0 ± 4.8 | 3.0 ± 1.6 | <0.001[a] |
| Age at onset (years) | – | 60.8 ± 5.6 | 67.2 ± 3.0 | <0.001[a] |
| UPDRS-ME | – | 24 (10–45) | 34 (24–47) | <0.001[b] |
| H&Y | – | 3 (1–4) | 4 (3–5) | <0.001[b] |
| MMSE | 26.5 ± 2.1 | 25.7 ± 1.9 | 24 ± 4.8 | <0.001[c] |

*Note*: Data are given as mean values (SD) or median values (range) when appropriate.
UPDRS-ME, Unified Parkinson Disease Rating Scale-Motor Examination in "off" phase (off medications overnight); H&Y, Hoehn-Yahr; MMSE, Mini Mental State Examination.
[a] Unpaired *t* test.
[b] Mann–Whitney test.
[c] One-way ANOVA.

Each image of the training set was multiplied with the corresponding weight and summed on a voxel basis, resulting in a map of values reflecting the importance of each voxel for SVM binary group discrimination (Kloppel et al., 2008; Focke et al., 2011). The resulting map was superimposed onto a standard stereotactic brain for visualization and localization purposes.

## 3. Results

### 3.1. Clinical and MRI studies

Data on demographic and clinical features of all patients and control subjects included in the study are listed in Table 1. Although no significant differences were detected among groups in demographical data, as expected, PSP patients were characterized by a more rapid disease progression (with a fatal prognosis after few years) and by a more critical clinical status with respect to PD patients in terms of motor disability, as assessed by higher UPDRS and H&Y scores.

Considering MRI studies, all subjects had no evidence of vascular lesions as evaluated in Fluid Attenuated Inversion Recovery (FLAIR) and by T2-weighted MRI. Healthy control subjects had normal MRI scanning. Both PD and PSP patients showed no evident structural abnormalities.

### 3.2. The machine-learning method

#### 3.2.1. Features extraction

Fig. 1 shows the 1-st, 2-nd and 3-rd PCA coefficients extracted, as representative examples, for the PSP (28) versus PD (28) binary labeled group (for this comparison, the total number of extracted PCA coefficients was equal to 55).

#### 3.2.2. Classification algorithm

Fig. 2 shows the optimal separating hyper-plane resulting from the SVM classification algorithm trained for the PSP (28) versus PD (28) binary labeled group.

### 3.3. Validation of the classifier

Table 2 shows Accuracy, Specificity and Sensitivity of the SVM classifier for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups obtained using LOO validation. Overall Mean Accuracy, Specificity and Sensitivity rates were calculated over a number of principal components ranging from 1 to 55. Overall Mean Accuracy (Specificity/Sensitivity) were 85.8 (86.0/86.0), 89.1 (89.1/89.5) and 88.9 (88.5/89.5)% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively.

Fig. 3 shows how the metrics depend on the number of principal components in LOO validation, as a representative example. In this case, results are shown for a number of principal components ranging from 1 to 55 and for the PSP versus PD binary labeled group. As expected, Accuracy, Specificity and Sensitivity rates increase with the number of considered principal components. This occurs because, in our method, the dimension of the feature space is enhanced with a covariance eigenvalue criterion coupled with an FDR criterion, allowing noisy information to be contained only in a small fraction of the eigenvectors (the last eigenvectors). For instance, without an FDR process, the relevant information would be contained in the few first eigenvectors and the rest of the eigenvectors would only contribute with noisy information (Alvarez et al., 2009).

The range of principal components for which Accuracy, Specificity and Sensitivity fell above 80% ($Accuracy_{>80}$, $Specificity_{>80}$ and $Sensitivity_{>80}$) was found from 30 to 52 components in LOO validation. In this range, for each of the binary labeled groups, $Accuracy_{>80}$ was found > 83.9% (mean $Accuracy_{>80}$: 92.7, 97.0 and 98.2% for PD versus Controls, PSP versus Controls and PSP

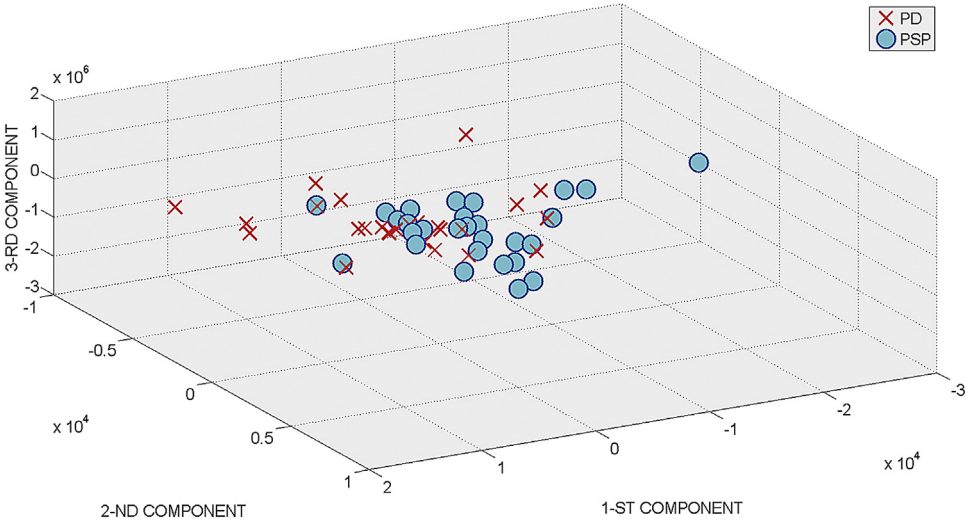


**Fig. 1.** PCA coefficients for the PSP versus PD binary labeled group (1st, 2nd and 3rd components).

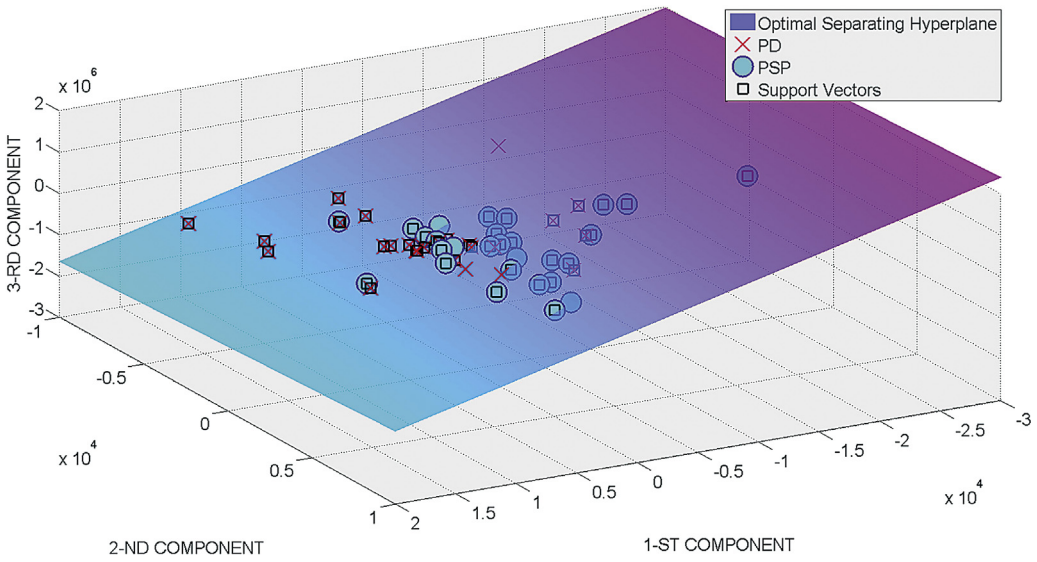


**Fig. 2.** Optimal separating hyper-plane for the PSP versus PD binary labeled group (1st, 2nd and 3rd components).

**Table 2**
Accuracy, Specificity and Sensitivity rates of SVM using LOO validation.

| | Overall Mean Accuracy (%)<br>$Accuracy_{>80}$ (%) Mean (Min/Max) | Overall Mean Specificity (%)<br>$Specificity_{>80}$ (%) Mean (Min/Max) | Overall Mean Sensitivity (%)<br>$Sensitivity_{>80}$ (%) Mean (Min/Max) |
|---|---|---|---|
| PD vs. Controls | 85.8<br>92.7 (83.9/100.0) | 86.0<br>92.3 (81.3/100.0) | 86.0<br>93.4 (80.6/100.0) |
| PSP vs. Controls | 89.1<br>97.0 (92.9/100.0) | 89.1<br>98.2 (92.9/100.0) | 89.5<br>95.9 (90.0/100.0) |
| PSP vs. PD | 88.9<br>98.2 (94.6/100.0) | 88.5<br>98.8 (96.3/100.0) | 89.5<br>97.8 (93.1/100.0) |

versus PD binary labeled groups, respectively). $Specificity_{>80}$ was found > 81.3% (mean $Specificity_{>80}$: 92.3, 98.2 and 98.8% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively). $Sensitivity_{>80}$ was found > 80.6% (mean $Sensitivity_{>80}$: 93.4, 95.9 and 97.8% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively). Furthermore, there is always at least one number of employed principal components for which $Accuracy_{>80}$, $Specificity_{>80}$ and $Sensitivity_{>80}$ values reach 100%.

Table 3 shows Accuracy, Specificity and Sensitivity of the SVM classifier for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups obtained using N/2 randomly chosen subjects for training the classifier and the remaining N/2 subjects for testing. Overall Mean Accuracy, Specificity and Sensitivity

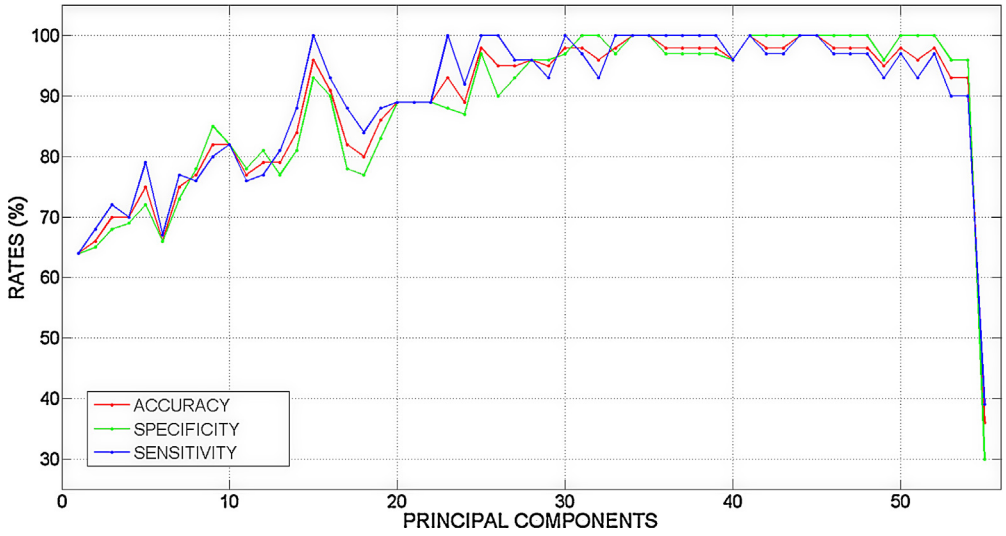


**Fig. 3.** Accuracy, Specificity and Sensitivity rates (%) of SVM versus Number of PCA components in LOO validation.

**Table 3**
Accuracy, Specificity and Sensitivity rates of SVM using N/2 subjects for training and N/2 subjects for testing.

| | Overall Mean Accuracy (%)<br>$Accuracy_{>80}$ (%) Mean (Min/Max) | Overall Mean Specificity (%)<br>$Specificity_{>80}$ (%) Mean (Min/Max) | Overall Mean Sensitivity (%)<br>$Sensitivity_{>80}$ (%) Mean (Min/Max) |
|---|---|---|---|
| PD vs. Controls | 83.2<br>93.5 (89.3/100) | 81.9<br>90.6 (82.4/100) | 85.4<br>97.4 (92.3/100) |
| PSP vs. Controls | 86.2<br>92.2 (85.7/96.4) | 92.1<br>92.5 (85.7/100) | 82.9<br>92.4 (85.7/100) |
| PSP vs. PD | 84.7<br>92.2 (89.3/96.4) | 87.5<br>91.3 (82.4/100) | 83.8<br>94.4 (86.7/100) |

rates were calculated over a number of principal components ranging from 1 to 28. Overall Mean Accuracy (Specificity/Sensitivity) were 83.2 (81.9/85.4), 86.2 (92.1/82.9) and 84.7 (87.5/83.8)% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively.

The range of principal components for which Accuracy, Specificity and Sensitivity fell above 80% ($Accuracy_{>80}$, $Specificity_{>80}$ and $Sensitivity_{>80}$) was found from 16 to 26 components when using N/2 randomly chosen subjects for training the classifier and the remaining N/2 subjects for testing. In this range, for each of the binary labeled groups, $Accuracy_{>80}$ was found > 85.7% (mean $Accuracy_{>80}$: 93.5, 92.2 and 92.2% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively). $Specificity_{>80}$ was found > 82.4% (mean $Specificity_{>80}$: 90.6, 92.5 and 91.3% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively). $Sensitivity_{>80}$ was found > 85.7% (mean $Sensitivity_{>80}$: 97.4, 92.4 and 94.4% for PD versus Controls, PSP versus Controls and PSP versus PD binary labeled groups, respectively). All the considered parameters proved the good performance of the classification algorithm with respect to all three binary labeled groups.

### 3.4. Voxel-based pattern distribution

Fig. 4 shows, maps of voxel-based pattern distribution displaying the influence of each voxel for the classification in the PD (28) versus Controls (28), in the PSP (28) versus Controls (28) and in the PSP (28) versus PD (28) comparisons. The pattern of differences is expressed according to the color scales. The most relevant finding concerned the separation of patients with PD with respect to patients with PSP. As showed in the bottom part of the Fig. 4, voxels influencing the classification between PD and PSP patients are localized in the midbrain, pons, corpus callosum and thalamus. When considering instead the direct comparisons with healthy controls, the voxel-based pattern distribution was similar for both PD and PSP patients, although it is important to highlight that SVM classification revealed significant voxels within the medial part of the midbrain (encompassing the substantia nigra) and the caudal part of the pons only for PD as compared to controls (upper part of the Fig. 4).

## 4. Discussion

Effective and accurate diagnosis of PD or PSP, a critical clinical variant of PD, by means of MRI biomarkers has recently attracted strong attention. So far, several biomarkers have been shown to be sensitive to the diagnosis of PD as compared to PSP. For instance, morphological abnormalities in the brainstem as well as in the cerebellar peduncles, have been demonstrated to be useful markers in a clinical context (Massey et al., 2013; Quattrone et al., 2008; Schulz et al., 1999). However, the only validated MRI-based measurement employed in clinical practice of PD derives from conventional MRI using manual morphometric quantification. In the above - mentioned studies it has been shown by using different approaches,

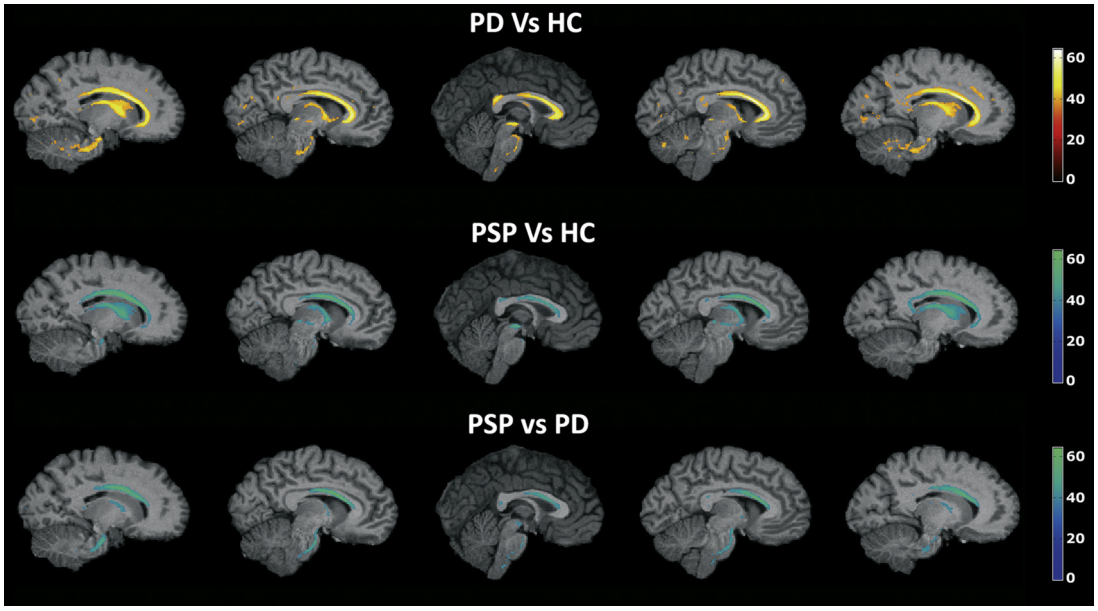


**Fig. 4.** Maps of voxel-based pattern distribution of brain structural differences (sagittal view, threshold = 60%). The importance of each voxel in the SVM classification is expressed according to the color scale. A, PD versus Controls; B, PSP versus controls; C, PSP versus PD; PD, Parkinson's disease; PSP, Progressive Supranuclear Palsy.

that the diameter or area of the brainstem subregions represent the simplest morphological feature on the basis of which a differential diagnosis of PD with respect to PSP can be performed. Although manual segmentation is currently considered as the gold standard approach to determine the morphology of brain regions, this technique is time-consuming, dependent on the raters' experience and limited to few specific regions-of-interest. For this reason, the implementation of supervised whole-brain automatic classification methods (Haller et al., 2011) is an essential step for improving clinical management of neurological patients, as well as, in longitudinal and prospective studies.

SVM has been proposed as a revolutionary approach for identifying sensitive biomarkers (or a combination between them) that allow for automatic discrimination of individual subjects. In this study, we considered a SVM algorithm using structural neuroimaging data at a whole brain level, reaching an excellent individual classification, both in the comparisons between PSP patients with PD patients (mean accuracy > 92.5) and, in the comparison between PD and healthy controls (mean accuracy > 92.7%). To the best of our knowledge, the classification accuracy in the discrimination of individual PSP patients was consistent with previous manual morphological metrics (Massey et al., 2013; Quattrone et al., 2008; Oba et al., 2005; Schulz et al., 1999) and with other SVM application to MRI data (Focke et al., 2011; Haller et al., 2012), whereas the accuracy in the detection of individual PD patients was significantly higher. For instance, in one first study (Focke et al., 2011), the best classification accuracy achieved ~ 97% (considering only white matter tissue) and was obtained for the comparison between PSP vs. PD using T1-weighted MRI data, while separation from the PD and the control cohort was only marginally better than chance. Successively, Haller and colleagues (2012, 2013) applying SVM to other MRI parameters (DTI and susceptibility-weighted images) reported classification accuracies in the individual diagnosis of PD comparable to our findings. However, it is important to bear in mind that in these later works, the clinical classification was made without considering healthy controls and including a heterogeneous cohort of parkinsonisms in the same group of patients, i.e. joining PSP cases with patients affected by multiple system atrophy (MSA) without considering that these disorders are characterized by distinct patterns of brain atrophy (Messina et al., 2011).

Our study has several strengths that may have improved our discrimination analysis. First, the larger sample of PSP patients investigated (n° 28) with respect to other studies (n° 10 in Focke et al., 2011; n° 1 in Haller et al., 2012 and n° 1 in Haller et al., 2013). Second, the clinical consistency of our group selection where only PSP patients were enrolled without conflating our analysis with the inclusion of other parkinsonian variants, such as MSA, dementia with Lewy Body, vascular Parkinsonism and atypical tremor (Focke et al., 2011; Haller et al., 2012, 2013). Third, the balance between classes (i.e., we choose the same number of samples in each class), allowing the system to learn without biases due to unequal samples.

Another point of relevance of our work is that we also studied the relevance of each brain voxel with respect to the classification analysis, thus allowing to identify regions critically involved in the pathophysiological mechanisms of PD and PSP. Indeed, brain regions that allowed to perform the best discrimination between PD and PSP were: midbrain, pons, corpus callosum and thalamus. These features are highly consistent with typical neuropathological (Steele et al., 1964) and imaging findings described in patients with PSP (Shi et al., 2013; Messina et al., 2011), where a key role is played by the volumetric atrophy of the brainstem, which represents a hallmark of PSP (Oba et al., 2005). More recently, studies using diffusion tensor imaging (DTI) for quantifying white matter pathology in PSP, highlighted the involvement of the corpus callosum as well. This finding is of particular relevance since corpus callosum is the largest white matter tract in the brain, enabling interhemispheric communication, particularly with respect to motor coordination, and is one of the tract that is known to be damaged in PSP (Knake et al., 2010; Canu et al., 2011).

As far as the spatial pattern of the SVM classification between PD and controls is concerned, more widespread patterns involving several cerebral regions were found in our analysis, thus confirming that PD is a more heterogeneous clinical phenotype that might be characterized by several and topographically separated neural pathologies. Of note, SVM revealed significant voxels within the medial part of the midbrain (encompassing the substantia nigra) and the caudal part of the pons. This finding is consistent with the Braak's neuroanatomical model of the PD (Braak et al., 2003). Post-mortem studies by Braak and colleagues (Del Tredici et al., 2002), based on the analysis of Lewy neuritis and Lewy bodies accumulation, a proteic hallmark of PD, have shown that various cerebral structures are damaged before substantia nigra in a consistent and repeated pattern. In a six-stage model (Braak et al., 2004), PD would initially begin in the medulla oblongata (stage 1) and in the olfactory bulb, and progresses in a caudo-rostral pattern (stage 2), affecting substantia nigra in stage 3 only, corresponding to the onset of the motor symptoms, often revealed by the first visit of the patient to a neurologist. However, previous structural neuroimaging studies investigating the neural basis of PD did not describe the occurrence of anatomical changes in this region (a part from a single study, Jubault et al., 2009). It must be acknowledged, that most of the studies that have investigated structural abnormalities in PD, as those previously reported, are based on standard mass-univariate analytical methods, applied in structural neuroimaging (i.e. voxel based morphometry). The SVM technique has one main advantage over these techniques: it takes inter-regional correlations into account and therefore is sensitive to differences that are subtle and spatially distributed; as such, it provides an ideal framework for investigating neurological disorders that affect a distributed network of regions.

In conclusion, our findings, together with those provided by other colleagues (Focke et al., 2011; Haller et al., 2012, 2013) offer new avenues for encouraging the application of computer-based diagnosis in clinical practice of PD.

## References


Alvarez I, Gorriz JM, Ramirez J, Salas-Gonzalez D, Lopez M, Puntonet CG, et al. Alzheimer's diagnosis using eigenbrains and support vector machines. Electron Lett 2009;45:342–3.

Braak H, Del Tredici K, Rub U, De Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003;24:197–211.

Braak H, Ghebremedhin E, Rub U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res 2004;318:121–34.

Canu E, Agosta F, Baglio F, Galantucci S, Nemni R, Filippi M. Diffusion tensor magnetic resonance imaging tractography in progressive supranuclear palsy. Mov Disord 2011;26:1752–5.

Chapelle O, Vapnik V. Model selection for support vector machines. Adv Neural Inform Process Syst 1999;12:230–6.

Del Tredici K, Rub U, De Vos RA, Bohl JR, Braak H. Where does parkinson disease pathology begin in the brain? J Neuropathol Exp Neurol 2002;61:413–26.

Focke NK, Helms G, Scheewe S, Pantel PM, Bachmann CG, Dechent P, et al. Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic parkinson syndrome and healthy controls. Hum Brain Mapp 2011;32:1905–15.

Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. Arch Neurol 1999;56:33–9.

Grabner G, Janke AL, Budge MM, Smith D, Pruessner J, Collins DL. Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. Med Image Comput Comput Assist Interv 2006;9(Pt 2):58–66.

Habeck C, Foster NL, Perneczky R, Kurz A, Alexopoulos P, Koeppe RA, et al. Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. Neuroimage 2008;40:1503–15.

Haller S, Badoud S, Nguyen D, Barnaure I, Montandon ML, Lovblad KO, et al. Differentiation between Parkinson disease and other forms of Parkinsonism using support vector machine analysis of susceptibility-weighted imaging (SWI): initial results. Eur Radiol 2013;23(1):12–9.

Haller S, Badoud S, Nguyen D, Garibotto V, Lovblad KO, Burkhard PR. Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results. AJNR Am J Neuroradiol 2012;33(11):2123–8.
Haller S, Lovblad KO, Giannakopoulos P. Principles of classification analyses in mild cognitive impairment (MCI) and Alzheimer disease. J Alzheimers Dis 2011;26(Suppl 3):389–94.
Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008;79:368–76.
Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62:782–90.
Jubault T, Brambati SM, Degroot C, Kullmann B, Strafella AP, Lafontaine AL, et al. Regional brain stem atrophy in idiopathic Parkinson's disease detected by anatomical MRI. PLoS One 2009;4:e8247.
Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131:681–9.
Knake S, Belke M, Menzler K, et al. In vivo demonstration of microstructural brain pathology in progressive supranuclear palsy: a DTI study using TBSS. Mov Disord 2010;25:1232–8.
Litvan I, Agid Y, Calne D, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology 1996;47:1–9.
López M, Ramírez J, Górriz JM, Álvarez I, Salas-Gonzalez D, Segovia F, et al. Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. Neurocomputing 2011;74:1260–71.
Magnin B, Mesrob L, Kinkingnéhun S, Pélégrini-Issac M, Colliot O, Sarazin M, et al. Support vector machine-based classification of Alzheimers disease from whole-brain anatomical MRI. Neuroradiology 2009;51:73–83.
Massey LA, Jäger HR, Paviour DC, O'Sullivan SS, Ling H, Williams DR, et al. The midbrain to pons ratio: a simple and specific MRI sign of progressive supranuclear palsy. Neurology 2013;80:1856–61.
Messina D, Cerasa A, Condino F, Arabia G, Novellino F, Nicoletti G, et al. Patterns of brain atrophy in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Parkinsonism Relat Disord 2011;17(3):172–6.
Oba H, Yagishita A, Terada H, Barkovich AJ, Kutomi K, Yamauchi T, et al. New and reliable MRI diagnosis for progressive supranuclear palsy. Neurology 2005;64:2050–5.
Quattrone A, Nicoletti G, Messina D, Fera F, Condino F, Pugliese P, et al. MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy1. Radiology 2008;246:214–21.
Salas-Gonzalez D, Górriz JM, Ramírez J, Illán IA, López M, Segovia F, et al. Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images. Med Phys 2010;37:6084–95.
Scholkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. 1st ed. Cambridge: MIT Press; 2001.
Schulz JB, Skalej M, Wedekind D, Luft AR, Abele M, Voigt K, et al. Magnetic resonance imaging based volumetry differentiates idiopathic Parkinson's syndrome from multiple system atrophy and progressive supranuclear palsy. Ann Neurol 1999;45:65–74.
Shi HC, Zhong JG, Pan PL, Xiao PR, Shen Y, Wu LJ, et al. Gray matter atrophy in progressive supranuclear palsy: meta-analysis of voxel-based morphometry studies. Neurol Sci 2013;34(7):1049–55, http://dx.doi.org/10.1007/s10072-013-1406-9.
Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004;23(S1):208–19.
Steele JC, Richardson JC, Olszewski J. Progressive supranuclear palsy. A heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Arch Neurol 1964;10:333–59.
Teipel SJ, Born C, Ewers M, Bokde AL, Reiser MF, Moller HJ, et al. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 2007;38:13–24.
Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. Lancet Neurol 2006;5:75–86.
Vapnik V. An overview of statistical learning theory. IEEE Trans Neural Netw 1999;10:988–1000.
Vapnik V. Statistical learning theory. New York: John Wiley and Sons, Inc; 1998.
Vapnik V. The nature of statistical learning theory. 2nd ed. New York: Springer-Verlag; 1995.