

Assisted Diagnosis of Parkinsonism Based on the Striatal Morphology

Fermín Segovia, Juan M. Górriz*, Javier Ramírez,
Francisco J. Martínez-Murcia and Diego Castillo-Barnes

*Department of Signal Theory
Networking and Communications
DASCI Institute, University of Granada
Granada 18071, SPAIN
gorriz@ugr.es

Accepted 1 March 2019

Published Online 14 May 2019

Parkinsonism is a clinical syndrome characterized by the progressive loss of striatal dopamine. Its diagnosis is usually corroborated by neuroimaging data such as DaTSCAN neuroimages that allow visualizing the possible dopamine deficiency. During the last decade, a number of computer systems have been proposed to automatically analyze DaTSCAN neuroimages, eliminating the subjectivity inherent to the visual examination of the data. In this work, we propose a computer system based on machine learning to separate Parkinsonian patients and control subjects using the size and shape of the striatal region, modeled from DaTSCAN data. First, an algorithm based on adaptative thresholding is used to parcel the striatum. This region is then divided into two according to the brain hemisphere division and characterized with 152 measures, extracted from the volume and its three possible 2-dimensional projections. Afterwards, the Bhattacharyya distance is used to discard the least discriminative measures and, finally, the neuroimage category is estimated by means of a Support Vector Machine classifier. This method was evaluated using a dataset with 189 DaTSCAN neuroimages, obtaining an accuracy rate over 94%. This rate outperforms those obtained by previous approaches that use the intensity of each striatal voxel as a feature.

Keywords: Parkinsonism; DaTSCAN; striatum; machine learning; striatal morphology.

1. Introduction

Parkinsonism refers to a family of disorders causing similar clinical syndromes: tremor, hypokinesia, rigidity and postural instability.¹ Most of the patients with Parkinsonism suffer from Parkinson's disease (PD), a neurodegenerative disorder that affects about 1–2% of people over 65 years and whose prevalence is increasing in developed nations due to the growth of the older population.

The diagnosis of Parkinsonism is usually corroborated by means of neuroimaging data of several

modalities. Because PD and related disorders cause a progressive loss of striatal dopamine, a big branch of neuroimaging modalities used for PD diagnosis focuses on visualizing that potential dopamine deficiency. An example is the ¹²³I-ioflupane or ¹²³I-FP-CIT (also known by the trademark name DaTSCAN), a radiopharmaceutical drug that binds the dopamine and serotonin transporters *in vitro* with high affinity. Thus, along with a Single Photon Emission Computer Tomography (SPECT) scanner, this drug allows us to visualize the loss of striatal

*Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

dopamine.^{2,3} Nevertheless, many researches still consider the ^{18}F -DOPA as the standard radiotracer for PD diagnosis. Since the decade of 1980, this drug has been successfully used to evaluate patients affected by PD.⁴ ^{123}I -IBZM SPECT is another neuroimaging modality widely used to characterize the dopamine $D_{2/3}$ receptors. A recent study confirmed that this radioligand is highly sensitive and accurately detects mild to subtotal striatal lesions by measuring loss of $D_{2/3}$ receptor in rats.⁵ In addition, non-dopamine focused radiopharmaceuticals as ^{18}F -FDG have been successfully used to assist the diagnosis of PD.⁶

Traditionally, neuroimaging data were visually analyzed by experienced clinicians in order to detect the specific patterns that characterize the disease. Nowadays, the visual examination is no longer acceptable since it is subjective and often misses important information. The striatal binding ratio, computed from the estimation of the radiopharmaceutical uptake using specific computer software is probably the “gold standard” to diagnose Parkinsonism using neuroimaging data. During the last decade, several computer-aided diagnosis (CAD) systems^{7,8} have been developed to automatically separate parkinsonian patients and healthy subjects, assisting the diagnosis that way.^{9,10} These systems are based on modern statistical classifiers^{11–15} and are able to take advantage of the huge amount of information contained in the neuroimages.^{16,17}

One of the major problems of statistical classification is the overfitting problem in high-dimensional settings with a small sample size, where the designed classifiers are inevitably over-adjusted to the training set, providing an almost vanishing training error, but with a poor generalization ability. Unfortunately, in neuroimaging this situation is the rule rather than the exception, since the dimensionality of each neuroimage (millions of voxels) in relation to the size of available samples (hundreds of scans) implies a high risk of overfitting. This issue is commonly referred as the *curse of dimensionality* or *small sample size problem* and can be also explained in terms of the high probability of the training set to be separable by a given surface in high dimensional spaces.⁴⁷ Several approaches have been proposed to address this problem and most of them consist on reducing the dimensionality of the data.^{18–20} They could be divided into two big branches: (i) Those that perform a voxels selection and are commonly based on previous

knowledge and (ii) The ones that summarize the information contained in the data into a reduced set of features and often do not require previous knowledge (they are based on ANOVA).^{9,21} In addition, to preserve complex models from overfitting, some solutions can be adopted that are well established on cross-validation (CV) methods. In this sense, the most common method to assess the validity of classification models in real-word datasets with a high number of attributes is 10-fold stratified CV.^{22,23}

In this work, we propose a novel approach to analyze DaTSCAN neuroimages using statistical classification in order to assist the diagnosis of Parkinsonism. It performs a dimensionality reduction by summarizing the information contained in the data into a reduced feature set, addressing that way the small sample size problem. After parceling the striatum using a mask individually generated per each subject, several morphological measures such as volume, perimeter or eccentricity were computed.²⁴ Then, these measures were used along with a support vector machine classifier to distinguish between parkinsonian patients and control subjects. This approach was evaluated using a dataset with 189 DaTSCAN neuroimages, obtaining an accuracy rate over 94%, which outperforms classical approaches.

2. Materials and Methods

2.1. Dataset

A dataset consisting of 189 DaTSCAN neuroimages was used in order to evaluate the proposed methodology (demographic details are given in Table 1). The data were acquired during a longitudinal study carried out in the “Virgen de la Victoria” hospital (Málaga, Spain). All the patients included in this study had been referred for DaTSCAN examination by the neurology department because of the suspicion of involvement of the nigrostriatal

Table 1. Group composition and demographic details of the data used in this work. μ and σ stand for the average and the standard deviation, respectively.

	Sex			Age		
	#	M	F	μ	σ	Range
Controls	94	49	45	69.26	10.16	33–89
Patients	95	54	41	68.29	9.62	30–87

pathway according to the clinical history and neurological examination. Subjects on treatment with drugs which have an effect, known or suspected, by a direct competitive mechanism at the level of dopaminergic transporters were excluded. The neuroimaging data were acquired 3 h after the radiopharmaceutical injection (185 MBq of ^{123}I -ioflupane) using a SPECT gamma camera (Millennium model from General Electric) equipped with a dual head and general-purpose collimator. The SPECT examination was performed after the first visit and at that time the patients were not undergoing treatment for movement disorders. The neuroimages were reconstructed by means of filtered back-projection algorithms without attenuation correction. A Hanning filter of frequency 0.7 was also applied.

After the reconstruction, the neuroimages were preprocessed in order to make them comparable. This procedure consisted of two steps: spatial registration and intensity normalization. The former was carried out using the template matching algorithm implemented in the SPM toolbox (version 8). To this end, a specific template was computed²⁵ as follows: First, the control neuroimages were spatially registered to a randomly selected one. Then, these neuroimages were averaged and the resulting image was made symmetrical. The latter step required a flipped image that was generated by flipping each coronal slice of the averaged image over the sagittal plane. Both, averaged and flipped images, were averaged and resulting image was used as template. The intensity normalization^{26–28} was performed by dividing the intensity level of each voxel by a value computed as the average of the 1% of the voxels of highest intensity (per neuroimage).

The data were labeled through visual inspection of the neuroimages by three experienced clinicians from the Nuclear Medicine service of the hospital. Two groups were defined:

- **Control subjects.** Bilateral, symmetrical uptake appeared in caudate and putamen nuclei.
- **Parkinsonian patients.** There were areas of significant reduced uptake in any of the striatal structures.

2.2. Morphological features

The approach we are proposing characterizes the potential striatal dopamine deficiency by means of

morphological measurements of the striatal region in DaTSCAN neuroimages. In order to deal with the non-bilateral disease variants,²⁹ left and right brain hemispheres were independently analyzed. Thus, after preprocessing the data as described in previous section, each neuroimage was divided into two, resulting in two volumes per subject, each one containing a brain hemisphere. Then, the striatal region of each hemisphere was parceled using an *ad-hoc* mask computed through an adaptative threshold. This procedure started by selecting the voxels whose intensity was higher than a value $I_{th} = 0.80 \times I_{max}$, being I_{max} the mean intensity of the 100 voxels with largest intensity in the hemisphere. I_{th} was iteratively increased with increments $\Delta = 0.01$ until the selected voxels are in the neighborhood of each other, i.e. they form a unique region. Algorithm 1 shows the pseudocode corresponding to this process.

Both, 3-dimensional and 2-dimensional measures were calculated for the two striatal regions of each subject. 3-dimensional measures are described as follows:

- **Volume.** Number of voxels.
- **Centroid.** Location in space of the central voxel of the region (3 coordinates). This measure matches with the center of mass assuming homogeneous (binaryzed) regions.

In order to compute 2-dimensional measures, the volumes were projected over the three possible planes, resulting in six 2-dimensional images (3 per hemisphere) per subject. The measures we computed are as follows:

- **Area.** Number of voxels in the plane.
- **Centroid.** Location of the central voxel of the region (2 coordinates).
- **Eccentricity.** Ratio of the distance between the foci of a conic section and its major axis length.
- **Roundness.** Indicate how closely the shape of the region approaches that of a circle. It was calculated as $4 \times \pi \times A \times P$, where A and P are the region's area and perimeter respectively.
- **Length of radial vectors.** Length of the 8 radii that extend from the centroid to the perimeter with angles 0° , 45° , 90° , 135° , 180° , 225° , 270° and 315° respect to the horizontal axis (8 values).

Algorithm 1: Parcellation of the striatal region

Input: Brain volume containing one brain hemisphere, V
Output: Striatal region, V_{sel}

$V_{\text{max}} =$ Select the 100 voxels with maximum intensity from V
 $I_{\text{max}} =$ Compute mean intensity in V_{max}
 $T = 0.8 \times I_{\text{max}}$

repeat

$V_{\text{sel}} =$ Select voxels with intensity $> T$ from V
$N_{\text{regs}} =$ Number of regions defined by voxels in V_{sel}
$T = T - 0.01$

until $N_{\text{regs}} > 1$

- **Orientation.** Angle (measured in degrees and given in the range $[-90, 90]$) between the major axis of the ellipse and the x axis.
- **Perimeter slope.** Slope of the 8 perimeter strokes that connect the end of the radial vectors (8 values).
- **Length of the major axis.** Length (in voxels) of the major axis of the ellipse that has the same normalized second central moments as the region.
- **Length of the minor axis.** Length (in voxels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

This procedure resulted in 76 features per hemisphere and, therefore, 152 features per subject.

2.3. Feature selection

As shown in previous section, the striatal region of each hemisphere was modeled by 76 features. This procedure intended to completely define the morphology of that region and produced a relatively high number of features. In order to reduce this number, we carried out a feature selection step that selected the most discriminant features. Three class-separability measures were analyzed to rank our feature set: Bhattacharyya distance, Fisher’s discriminant ratio and relative entropy.

These measures were independently computed per each feature, and the features with highest distance/ratio were selected. Specifically, we used the minimum number of features so that the sum of the distance/ratio of the selected features is 80% of the total (the sum of the distance/ratio of all features). That way, we selected the smallest possible feature

set with as much discriminative information as possible.

2.3.1. Bhattacharyya distance

The Bhattacharyya distance, B , is defined as

$$B = \frac{1}{8}(\mu_1 - \mu_2)^T \left(\frac{\sigma_1 + \sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \frac{|\frac{\sigma_1 + \sigma_2}{2}|}{\sqrt{|\sigma_1| |\sigma_2|}}, \quad (1)$$

where μ_i and σ_i , respectively are the mean and standard deviation of the i th class and $|\cdot|$ denotes the determinant of the respective matrix.³⁰ Note that, in our case (features are individually ranked), σ_i is a scalar and $|\sigma_i|$ is simply σ_i . The value B can be considered a class separability measure and it is commonly used in the literature to calculate the Chernoff bound, a minimum bound to estimate the minimum attainable classification error of a classification procedure.

2.3.2. Fisher’s discriminant ratio

The Fisher’s discriminant ratio (FDR),³⁰ J , is defined as

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}, \quad (2)$$

where \mathbf{w} represents a direction in the data space and S_B and S_W are respectively the “between classes” and the “within classes” scatter matrices. Note that, scatter matrices are proportional to covariance matrices and, when only 2 classes are considered, S_B

can be expressed as

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T, \quad (3)$$

where μ_i denotes the mean of the samples belonging to the i th class. In this case, the Fisher's discriminant ratio for a single feature can be estimated as

$$J = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}, \quad (4)$$

where μ_i and σ_i are, respectively, the mean and the standard deviation of the selected feature for subjects belonging to the i th class.

2.3.3. Relative entropy

The cross or relative entropy³⁰ is a measure of the Kullback–Leibler divergence, D , which can be expressed as

$$D = \int_{-\infty}^{\infty} (p(\mathbf{x}|\omega_1) - p(\mathbf{x}|\omega_2)) \ln \frac{p(\mathbf{x}|\omega_2)}{p(\mathbf{x}|\omega_1)} d\mathbf{x}, \quad (5)$$

where $p(\mathbf{x}|\omega_i)$ is the probability density function for class ω_i and \mathbf{x} is a feature vector (describing a specific feature selection). Assuming Gaussian density functions and 2 classes, the one-dimensional case (feature vectors with only a feature) can be calculated as

$$D = \frac{1}{2} \left(\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} - 2 \right) \frac{1}{2} (\mu_1 - \mu_2)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right), \quad (6)$$

where μ_i and σ_i are, respectively, the mean and the standard deviation of the selected feature for subjects belonging to the i th class.

2.4. Background on support vector machine

Given a set of training data composed by N -dimensional patterns, $\mathbf{x}_i \in \mathbb{R}^N$, and binary labels, $y_i \in \{-1, 1\}$ defining the patterns' class

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l) \quad (7)$$

a Support Vector Machine classifier calculates an hyperplane, $g(x)$, that has the largest distance to the closest training data point of any class

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = 0, \quad (8)$$

where \mathbf{w} and w_0 are, respectively the weight vector, orthogonal to the decision hyperplane, and the threshold. Using this hyperplane, the classifier is able to estimate the class label of new unseen patterns

according to the side of the hyperplane, where the patterns are

$$\hat{y} = \text{sign}(\mathbf{w}^T \mathbf{x} + w_0). \quad (9)$$

The computation of the hyperplane can be solved using as a maximization problem and using Lagrangian functions, resulting in

$$\begin{aligned} & \text{maximize} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\ & \text{subject to} \quad \sum_{i=1}^N \alpha_i y_i = 0 \\ & \quad \quad \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N, \end{aligned} \quad (10)$$

where α is the vector of dual variables corresponding to each separation constraint, and $k(\mathbf{x}_i, \mathbf{x}_j)$ is a function $\mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ known as “kernel”.³¹ For linear SVM, $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j$.

2.5. Univariate analyses

A voxel-wise statistical analysis was carried out in order to examine effects contained in the data.³² Specifically, we performed a 2×2 factor analysis on SPM to evaluate the presence of significant diagnosis \times hemisphere interactions. To this end, we generated brain maps containing only data from left or right hemisphere and then a general linear model at each voxel was fit with diagnosis and hemisphere as fixed factors and age as nuisance covariate.³³ Statistical outcome were corrected for multiple comparisons using the widely used family-wise error (FWE) rate of 0.05.^{34,35} In addition, effects due to gender and age were discarded using ANOVA. F -statistic and corresponding p -value were computed to determine whether controls and patients have different ages. Effects due to gender were assessed by a voxel-wise ANOVA analysis performed on SPM.

2.6. A computer aided diagnosis system for Parkinsonism

A CAD system for Parkinsonism was build using the feature extraction method proposed in Secs. 2.2 and 2.3, and a Support Vector Machine (SVM) classifier. This system was compared with the classical approach consisting on using the intensity of voxels located in the striatum as feature. Two methodologies were considered to parcel the striatal voxels

for the classical approach: (i) Using an atlas, and (ii) using an intensity threshold. Classification measures of all these systems were estimated using the database described in Sec. 2.1 along with a 10-fold CV scheme. Moreover, the reported measures were computed as mean (\pm standard deviation) across 10 repetitions of the CV procedure (with different training/test sets). In all cases a stratified sampling strategy was followed to calculate training and test sets. This sampling procedure keeps in each sample the same (or approximately same) proportion of examples of each class as there is in the whole population. In our case, this ensured approximately same number of controls and patients in all training and test sets. In order to avoid biased results, the feature selection procedure was run in an inner CV loop so that only training data were used to perform the feature selection. Thus, during the training step, we saved the position of most discriminative features in the feature vectors (obtained by means of the feature selection procedure) and then, during the test step, we selected the features located at the saved positions (which are supposed to be the most discriminative ones). The SVM cost parameter, C , was set to the commonly used value of $C = 1$.

The significance of the accuracy rate obtained by the proposed method was estimated using

a permutation test.³⁶ In this experiment the classification procedure was repeated 1000 times using different label sets generated as random permutations of the original one. The p -value was estimated as the percentage of these classification procedures whose accuracy was equal to or greater than the accuracy obtained with the true labels.

3. Results

Figure 1 shows the brain regions with significant effects of diagnosis determined by the factor analysis carried out in SPM. As expected, big clusters covering the striatal region were found.^{37,38} No significant clusters were found on hemisphere. Similarly, no significant differences were found by the ANOVA tests for age (F -statistic = 0.45; $p = 0.51$) and gender ($p < 0.05$, FWE).

Once the morphological features were extracted, the difference between patients and controls can be roughly estimated from the mean value of those morphological measures. Table 2 shows the absolute value of volume and area, eccentricity and length of major and minor axes of each possible 2-dimensional projection. Note that, differences between patients and controls are clear for some features, as volume and eccentricity of the axial projection.

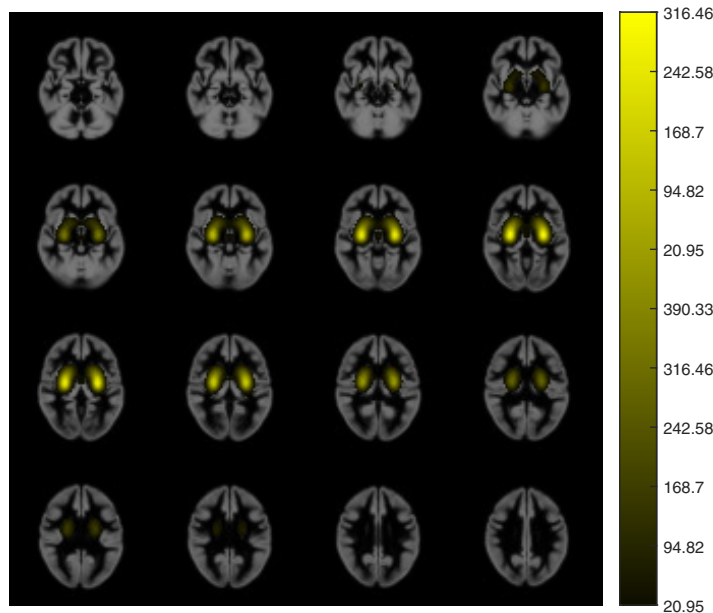


Fig. 1. Brain regions showing significant effects ($p < 0.05$, FWE) of diagnosis in the 2×2 factorial design. Clusters were represented over the gray matter segment of a study-specific template.

Table 2. Average value of most representative morphological features for control subjects and parkinsonian patients.

	Left hem.		Right hem.	
	CTL	PKS	CTL	PKS
Volume	176	142	173	157
Axial projection:				
Area	41	32	41	33
Eccentricity	0.84	0.63	0.83	0.63
Length of the major axis	10.2	7.5	10.1	7.7
Length of the minor axis	5.5	5.4	5.4	5.5
Coronal projection:				
Area	37	33	36	35
Eccentricity	0.54	0.57	0.54	0.60
Length of the major axis	7.7	7.5	7.6	7.7
Length of the minor axis	6.3	5.8	6.3	5.9
Sagittal projection:				
Area	45	36	45	40
Eccentricity	0.69	0.59	0.68	0.60
Length of the major axis	9.2	7.8	9.2	8.2
Length of the minor axis	6.5	5.9	6.5	6.2

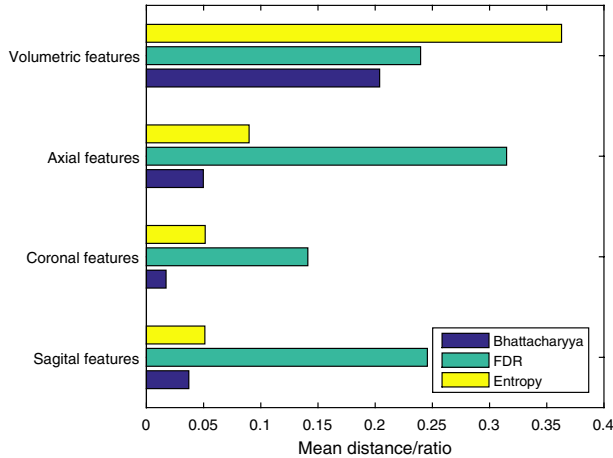


Fig. 2. Mean Bhattacharyya distance, Fisher's discriminant ratio and relative entropy of the extracted features. They were grouped into 4 groups according to the data they were extracted from: (1) Projection over the sagittal plane, (2) Projection over the coronal plane, (3) Projection over the axial plane and (4) Whole volume.

All morphological features can be categorized into four groups according to the data they were extracted from. Figure 2 shows the mean distance/ratio of 4 feature groups: features extracted from sagittal, coronal or axial projections and from

the 3-dimensional volume. This allows us to estimate the importance of these data in the separation problem.

The three class-separability measures described in Sec. 2.3 were evaluated in terms of the distance that they separate the groups defined in our dataset (Fig. 3) as well as in terms of the performance of a CAD system that uses them. Table 3 shows the classification metrics obtained by the implemented CAD systems. The accuracy rate can be further analyzed in Fig. 4, which shows the accuracies obtained in the 10 times that the 10-fold cross-validation procedure was repeated.

For the sake of completeness, we also evaluated the performance of a SVM classifier fed by striatal standardized uptake values (SUV),³⁹ which are considered by many as the standard biomarkers for PD diagnosis. Figure 5 shows the classification hyperplane for this approach, which yields an accuracy about 64% (estimated through 10-fold CV). To a certain extent, this is similar to use only the striatal volume as feature.

A permutation test was performed to estimate a p -value for the accuracy achieved by the CAD system that provided the best performance (the one based on the proposed feature extraction and Bhattacharyya distance). As a result, a p -value of $p < 0.001$ was obtained. Figure 6 shows the histogram corresponding to the accuracy rates

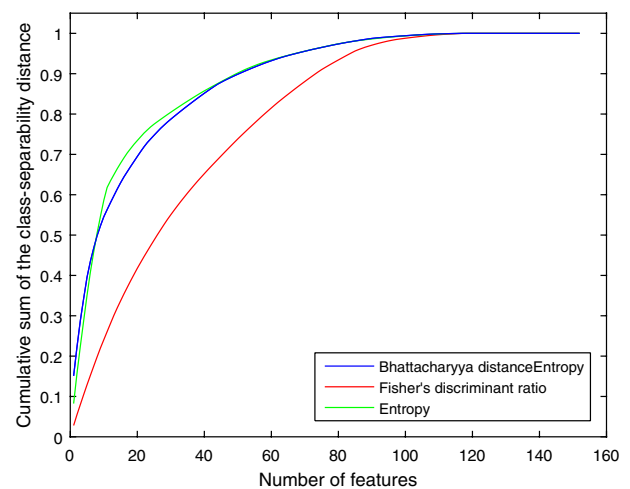


Fig. 3. Cumulative sum of the Bhattacharyya distance, Fisher's discriminant ratio and relative entropy corresponding to the 152 initial features sorted in descending order. They were normalized to fix in the interval $[0, 1]$.

Table 3. Classification measures obtained by the CAD systems implemented in this work. They were calculated as the mean (\pm SD) of those obtained from 10 repetitions of a 10-fold cross-validation procedure. TPR, FPR, FNR and TNR stand for true positive rate, false positive rate, false negative rate and true negative rate, respectively.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	TPR	FPR	FNR	TNR
Proposed system (Bhattacharyya)	94.25 (\pm 0.74)	91.26 (\pm 1.12)	96.17 (\pm 0.74)	86.70	3.60	8.30	90.40
Proposed system (FDR)	90.21 (\pm 1.10)	86.74 (\pm 2.49)	92.77 (\pm 2.06)	82.40	6.80	12.60	87.20
Proposed system (entropy)	90.58 (\pm 0.75)	81.58 (\pm 1.14)	98.62 (\pm 0.51)	77.50	1.30	17.50	92.70
Striatal intensity (atlas)	90.16 (\pm 0.94)	88.95 (\pm 1.43)	91.38 (\pm 0.78)	84.50	8.10	10.50	85.90
Striatal intensity (threshold)	88.99 (\pm 0.89)	87.68 (\pm 1.00)	90.32 (\pm 1.27)	83.30	9.10	11.70	84.90

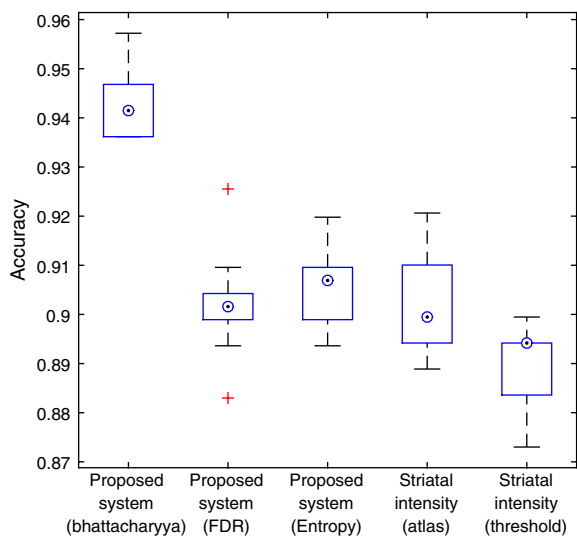


Fig. 4. (Color online) Accuracies obtained across 10 repetitions of the cross-validation procedure. Blue boxes and circled dots represent accuracies' range and median, respectively. Some values considered as outliers are plotted using the + symbol.

obtained in all classification procedures carried out in this test.

The influence of the most discriminative morphological features could be also analyzed by examining the weights assigned by the SVM classifier.⁴⁰ Figures 7 and 8 show the projection of the data from all our patients over some selected weight coordinates, w_i , computed by the SVM classifier. We selected the weights corresponding to the morphological features with larger Bhattacharyya distance between groups. Specifically, the volume and area and center of mass of each of the three 2-dimensional projections.

4. Discussion and Conclusion

The results shown in Table 3 suggest that the striatal size and shape contain enough information to

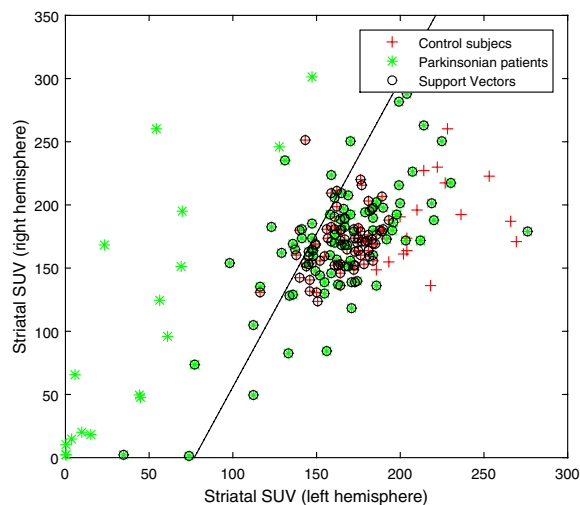


Fig. 5. Support vectors and classification hyperplane obtained with a system that uses only the striatal SUV as feature.

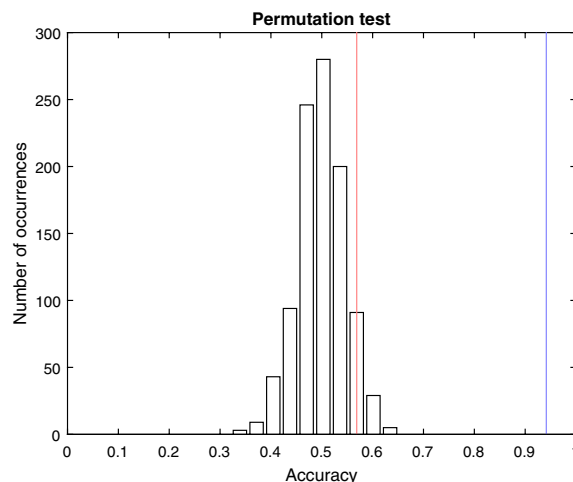


Fig. 6. (Color online) Histogram of the accuracy rates achieved by using randomly generated label sets (1000 repetitions) and the proposed method (with Bhattacharyya distance). Red and blue lines are, respectively the accuracy associated with a p -value of 0.05 and the accuracy obtained when using the true labels (94.25%).

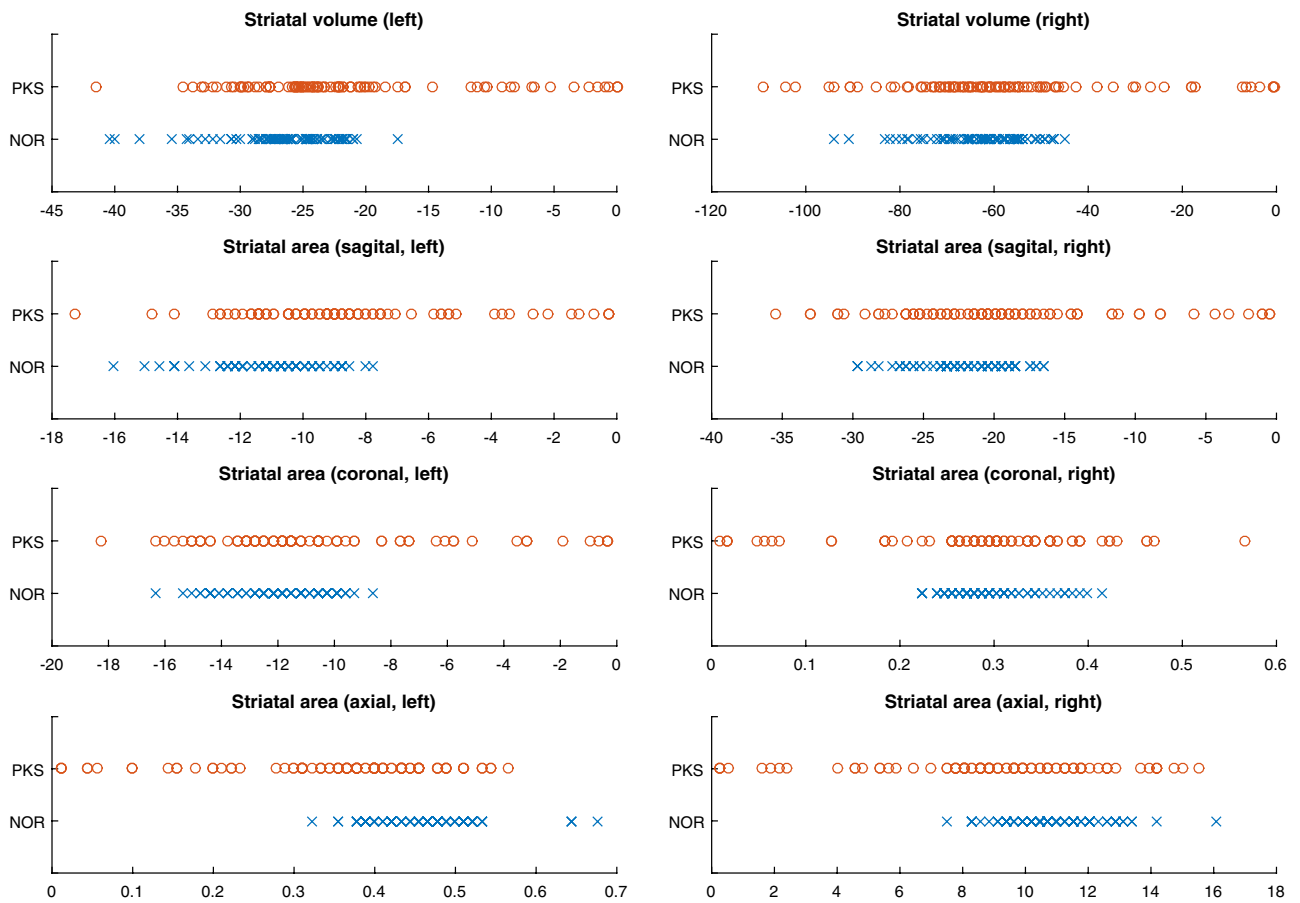


Fig. 7. (Color online) Projection of selected morphological measures of each patient's striatum over the weight coordinate corresponding to those features. Left and right columns show representations for features corresponding, respectively to left and right hemisphere. In all plots, the data are grouped by diagnosis and different markers were used for controls (blue crosses) and patients (red circles).

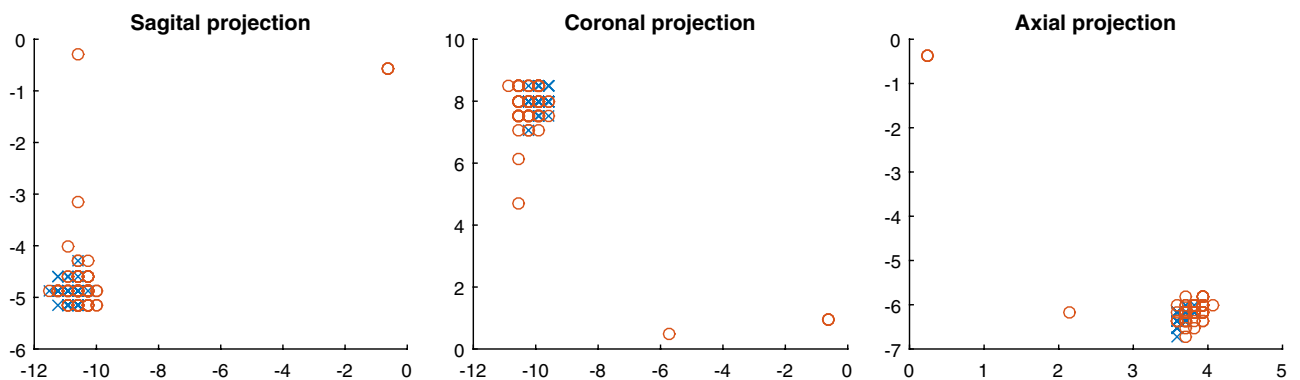


Fig. 8. (Color online) Projection of center of mass of each patient's striatum over the corresponding weight coordinate computed by the SVM classifier. As in Fig. 7, different markers were used for controls (blue crosses) and patients (red circles).

differentiate parkinsonian patients and control subjects. This is consistent with a previous work that reported differences between PD patients and controls on some morphological measures.⁴¹ In this work, it has been shown that using this information, a CAD system for Parkinsonism based on SVM classification reached an accuracy rate over 94% ($p < 0.001$). These results outperformed the ones obtained by classical approaches that used the intensity of the striatal voxels as feature and are in line with modern systems based on more sophisticated approaches. For the same neuroimaging modality, previous works have reported accuracy rates about 94% using principal component analysis⁹ or partial least squares to reduce the dimensionality of the data.¹⁶ Similar rates were reported by a recent work that used deep neural networks to separate PD patients and normal controls.⁴² Compared with previous works, the approach proposed here can be considered a simpler way to achieve an accurate and automatic separation of parkinsonian patients and controls. In addition, the reported results corroborate that the size and shape of the distribution of the striatal dopamine is an effective biomarker for the diagnosis of Parkinsonism. Nevertheless, we should be careful when comparing classification results obtained by using different datasets because the accuracy and remaining performance metrics highly depend on how they were estimated and the data used for it. An accuracy rate about 95% is probably close to highest precision achievable with a specific dataset. Note that, an accuracy of 100% is highly unlikely and maybe not desirable due to potential labeling errors.

According to Fig. 2, the features that contribute the most to separate the groups are the volumetric ones, i.e. the volume and centroid of the striatum. Their mean Bhattacharyya distance and relative entropy are considerably larger than those corresponding to any of the 2-dimensional projections. This is consistent with the results shown in Table 2 which shows that the striatal volume (along with the eccentricity of the axial projection) is the measure with the largest difference between parkinsonian patients and controls. This result also coincides with clinical findings, since it is known that Parkinsonism produces a decrease of the striatal dopamine that is reflected on DaTSCAN neuroimages.²

One of the key properties of the proposed method is the adaptative threshold described in

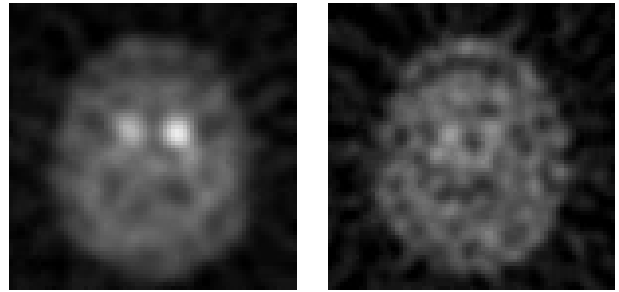


Fig. 9. Axial slices showing the striatum of two parkinsonian patients. Left figure: patient with non-homogeneous dopamine deficiency across hemispheres; the left one is more affected. Right figure: patient in advanced stage; it is difficult to parcel the striatum.

Algorithm 1. This method allows us to perform an accurate parcellation of the striatum in an automatic way, i.e. without human intervention, even in the non-bilateral disease (see Fig. 9). Theoretically the intensity normalization applied during the data preprocessing should make possible to use a common threshold for all the neuroimages. But, in practice, this results in two or more not connected regions for some brain hemispheres and no region for some others. The approach we propose ensures only one region is selected and, as expected, this region is very small in patients in advanced stages, when it is difficult to detect the striatal region in DaTSCAN neuroimages.

In this work, we decided to include a large number of morphological measures, assuming this decision could lead to some redundancy. For example, roundness and eccentricity are similar and, moreover, all 2-dimensional measures were repeated for the three possible projections. As a result, we obtained 152 features per patient, which are not many compared with the number of features used by classical approaches but enough to affect the classification accuracy because of the small sample size problem. For this reason, a feature selection algorithm was included before the classification to discard the least discriminative features. Specifically, three classical approaches for feature selection (Bhattacharyya distance, Fisher's discriminant ratio and relative entropy) were evaluated. Among them, the Fisher's discriminant ratio is probably the most widely used in neuroimaging studies. However, according to our experiments, this is the approach that produced the least dimensionality reduction (Fig. 3). These three techniques work similarly. They rank each feature

according to the mean of that feature for each class and take into account, somehow, the standard deviation. If they are analyzed in terms of classification performance, both, FDR and entropy, gave rise to accuracies similar to those obtained by classical approaches based on voxel intensity. Conversely, the Bhattacharyya distance provided a more effective feature selection that leads to better classification performance as it is clearly shown in Fig. 4. These differences can be explained by the lower weight given by the Bhattacharyya distance to the difference between the mean of each class, which increase the relative weight of the standard deviation in the calculation (note that the difference between the mean of each class is multiplied by $\frac{1}{8}$ (Eq. 1)). It is worth noting that the usefulness of a given feature set depends on the statistical classifier used in the subsequent classification procedure. Thus, our experiments determined that the Bhattacharyya distance provides a better feature selection than FDR or entropy when SVM is used to classify the selected features. The selection of a SVM classifier was supported by the successful results reported in previous neuroimaging studies that used SVM.^{43,44} Indeed, this classifier is considered by many as the standard classifier for neuroimaging data.

This work focused on the automatic separation of parkinsonian patients and control subjects. However, the major challenge in the diagnosis of Parkinsonism is commonly the differentiation among parkinsonian syndromes. This is therefore a limitation of this work. Analyzing whether the striatal morphology can be used to distinguish between different parkinsonian pathologies would require a dataset with neuroimaging data from patients suffering those disorders in a comparable number and accurately labeled (unfortunately some parkinsonian patients are misdiagnosed with a related parkinsonian syndrome⁴⁵). In addition, it would require using an appropriate neuroimaging modality. Several parkinsonian syndromes such as multiple system atrophy or progressive supranuclear palsy produce a similar reduction of dopamine transporters⁴⁶ and, therefore, the usefulness of DaTSCAN to differentiate among them is limited.

Another limitation is related to the *gold standard* used to label the images. In this work, we used the report issued by three experienced clinicians who visually analyzed the data. Thus, the developed

systems were intended to reproduce the knowledge of those clinicians, which is not exempt of errors.

Acknowledgments

This work was supported by the MINECO/FEDER under the TEC2015-64718-R project, the Ministry of Economy, Innovation, Science and Employment of the Junta de Andalucía under the P11-TIC-7103 Excellence Project and the Vicerectorate of Research and Knowledge Transfer of the University of Granada.

References

1. D. Greenberg, M. Aminoff and R. Simon, *Clinical Neurology, 8th Edition*, 8th edn. (McGraw-Hill Professional, New York, 2012).
2. J. Booij, G. Tissingh, G. J. Boer, J. D. Speelman, J. C. Stoof, A. G. Janssen, E. C. Wolters and E. A. van Royen, [123I]FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labeling in early and advanced Parkinson's disease, *J. Neurol. Neurosurg. Psychia* **62** (1997) 133–140.
3. M. D. Martínez-Valle Torres, S. J. Ortega Lozano, M. J. Gómez Heredia, T. Amrani Raissouni, E. Ramos Moreno, P. Moya Espinosa and J. M. Jiménez-Hoyuela, Longitudinal evaluation using FP-CIT in patients with parkinsonism, *Neurología (English Edition)* **29** (2014) 327–333.
4. C. Nanni, S. Fanti and D. Rubello, 18F-DOPA PET and PET/CT, *J. Nucl. Med.* **48** (2007) 1577–1579.
5. C. Scherfler, S. W. Scholz, E. Donnemiller, C. Decristoforo, M. Oberladstätter, N. Stefanova, E. Diederer, I. Virgolini, W. Poewe and G. K. Wenning, Evaluation of [123I]IBZM pinhole SPECT for the detection of striatal dopamine D2 receptor availability in rats, *NeuroImage* **24** (2005) 822–831.
6. G. Garraux, C. Phillips, J. Schrouff, A. Kreisler, C. Lemaire, C. Degueldre, C. Delcour, R. Hustinx, A. Luxen, A. Destée and E. Salmon, Multiclass classification of FDG PET scans for the distinction between Parkinson's disease and atypical parkinsonian syndromes, *NeuroImage: Clin.* **2** (2013) 883–893.
7. H. Adeli and S. Ghosh-Dastidar, *Automated EEG-Based Diagnosis of Neurological Disorders: Inventing the Future of Neurology*, 1st edn. (CRC Press, Boca Raton, FL, 2010).
8. G. Mirzaei, A. Adeli and H. Adeli, Imaging and machine learning techniques for diagnosis of Alzheimer's disease, *Rev. Neurosci.* **27** (2016) 857–870.
9. D. J. Towey, P. G. Bain and K. S. Nijran, Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images, *Nucl. Med. Commun. August 2011* **32**(8) (2011) 699–707.

10. I. A. Illán, J. M. Górriz, J. Ramírez, F. Segovia, J. M. Jiménez-Hoyuela and S. J. O. Lozano, Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging, *Med. Phys.* **39** (2012) 5971–5980.
11. F. C. Morabito, M. Campolo, N. Mammone, M. Versaci, S. Franceschetti, F. Tagliavini, V. Sofia, D. Fatuzzo, A. Gambardella, A. Labate, L. Mumoli, G. G. Tripodi, S. Gasparini, V. Cianci, C. Sueri, E. Ferlazzo and U. Aguglia, Deep learning representation from electroencephalography of early-stage creutzfeldt-jakob disease and features for differentiation from rapidly progressive dementia, *Int. J. Neur. Syst.* **27** (2017) 1650039.
12. L. Guo, Z. Wang, M. Cabrerizo and M. Adjouadi, A cross-correlated delay shift supervised learning method for spiking neurons with application to interictal spike detection in epilepsy, *Int. J. Neur. Syst.* **27** (2017) 1750002.
13. S. Ghosh-Dastidar and H. Adeli, A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection, *Neur. Network.* **22** (2009) 1419–1431.
14. Y. Zhang, Y. Wang, J. Jin and X. Wang, Sparse bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification, *Int. J. Neur. Syst.* **27** (2017) 1650032.
15. N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti and F. C. Morabito, Permutation Jaccard distance-based hierarchical clustering to estimate EEG network density modifications in MCI subjects, *IEEE Trans. Neur. Networks Learn. Syst.* **29**(10) (2018) 1–14.
16. F. Segovia, J. M. Górriz, J. Ramírez, I. Álvarez, J. M. Jiménez-Hoyuela and S. J. Ortega, Improved parkinsonism diagnosis using a partial least squares based approach, *Med. Phys.* **39** (2012) 4395–4403.
17. A. Ortiz, J. Munilla, J. M. Górriz and J. Ramírez, Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease, *Int. J. Neur. Syst.* **26** (2016) 1650025.
18. D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. Álvarez, M. López, F. Segovia and C. G. Puntonet, Two approaches to selecting set of voxels for the diagnosis of Alzheimer's disease using brain SPECT images, *Digit. Sign. Process.* **21** (2011) 746–755.
19. F. Segovia, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, I. Álvarez, M. López and R. Chaves, A comparative study of feature extraction methods for the diagnosis of Alzheimer's disease using the ADNI database, *Neurocomput.* **75** (2012) 64–71.
20. J. M. Górriz, F. Segovia, J. Ramírez, A. Lassel and D. Salas-Gonzalez, GMM based SPECT image classification for the diagnosis of Alzheimer's disease, *Appl. Soft Comput.* **11** (2011) 2313–2325.
21. F. Segovia, J. Górriz, J. Ramírez, D. Salas-González and I. Álvarez, Early diagnosis of Alzheimer's disease based on partial least squares and support vector machine, *Expert Syst. Appl.* **40** (2013) 677–683.
22. R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proc. 14th International Joint Conf. Artificial Intelligence — Volume 2, IJCAI'95* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995), pp. 1137–1143.
23. S. Varma and R. Simon, Bias in error estimation when using cross-validation for model selection, *BMC Bioinformatics* **7** (2006) 91.
24. F. Segovia, J. M. Górriz, J. Ramírez, F. J. Martínez-Murcia, D. Castillo-Barnes, I. A. Illán, A. Ortiz and D. Salas-Gonzalez, Automatic separation of parkinsonian patients and control subjects based on the striatal morphology, *Natural and Artificial Computation for Biomedicine and Neuroscience* (Springer, Cham, La Coruña Spain Ponencia, 2017), pp. 345–352.
25. D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, P. Padilla, F. J. Martínez-Murcia and E. W. Lang, Building a FP-CIT SPECT brain template using a posterization approach, *Neuroinformat.* **13** (2015) 391–402.
26. D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán and E. W. Lang, Linear intensity normalization of FP-CIT SPECT brain images using the α -stable distribution, *NeuroImage* **65** (2013) 449–455.
27. A. Brahim, J. M. Górriz, J. Ramírez and L. Khedher, Intensity normalization of DaTSCAN SPECT imaging using a model-based clustering approach, *Appl. Soft Comput.* **37** (2015) 234–244.
28. A. Brahim, J. Ramírez, J. M. Górriz, L. Khedher and D. Salas-Gonzalez, Comparison between different intensity normalization methods in 123I-Ioflupane imaging for the automatic detection of parkinsonism, *Plos One* **10**(6) (2015) e0130274.
29. R. Djaldetti, I. Ziv and E. Melamed, The mystery of motor asymmetry in Parkinson's disease, *Lancet Neurol.* **5** (2006) 796–802.
30. S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th edn. (Academic Press, Amsterdam, 2008).
31. K. Müller, S. Mika, G. Rätsch, K. Tsuda and B. Schölkopf, An introduction to kernel-based learning algorithms, *IEEE Trans. Neural Networks* **12** (2001) 181–201.
32. K. Friston, Chapter 02 — Statistical parametric mapping, *Statistical Parametric Mapping*, eds. K. Friston, J. Ashburner, S. Kiebel, T. Nichols and W. Penny (Academic Press, London, 2007), pp. 10–31.
33. D. Glaser and K. Friston, Chapter 10 — Covariance components, *Statistical Parametric Mapping*, eds. K. Friston, J. Ashburner, S. Kiebel, T. Nichols and W. Penny (Academic Press, London, 2007), pp. 140–147.

34. Y. Hochberg and A. C. Tamhane, *Multiple Comparison Procedures*, 1st edn. (Wiley, New York, 2009).
35. J. R. Chumbley and K. J. Friston, False discovery rate revisited: FDR and topological inference using Gaussian random fields, *NeuroImage* **44** (2009) 62–70.
36. F. Pereira, T. Mitchell and M. Botvinick, Machine learning classifiers and fMRI: A tutorial overview, *NeuroImage* **45** (2009) S199–S209.
37. A. Antonini, K. L. Leenders, P. Vontobel, R. P. Maguire, J. Missimer, M. Psylla and I. Günther, Complementary PET studies of striatal neuronal function in the differential diagnosis between multiple system atrophy and Parkinson's disease, *Brain* **120** (1997) 2187–2195.
38. B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. Senthil Kumaran, R. K. Agrawal and M. Behari, Regions-of-interest based automated diagnosis of Parkinson's disease using T1-weighted MRI, *Expert Syst. Appl.* **42** (2015) 4506–4516.
39. P. E. Kinahan and J. W. Fletcher, PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy, *Seminars in Ultrasound, CT, and MR* **31** (2010) 496–505.
40. D. Caragea, D. Cook and V. G. Honavar, Gaining insights into support vector machine pattern classifiers using projection-based tour methods, *Proc. Seventh ACM SIGKDD International Conf. Knowledge Discovery and Data Mining, KDD '01* (ACM, New York, NY, USA, 2001), pp. 251–256.
41. R. Prashanth, S. D. Roy, S. Ghosh and P. K. Mandal, Shape features as biomarkers in early Parkinson's disease, *2013 6th International IEEE/EMBS Conf. Neural Engineering (NER)*, 2013, San Diego, CA, USA, pp. 517–520.
42. H. Choi, S. Ha, H. J. Im, S. H. Paek and D. S. Lee, Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging, *NeuroImage: Clinical* **16** (2017) 586–594.
43. R. Yuvaraj, M. Murugappan, U. R. Acharya, H. Adeli, N. M. Ibrahim and E. Mesquita, Brain functional connectivity patterns for emotional state classification in Parkinson's disease patients without dementia, *Behavioral Brain Res.* **298** (2016) 248–260.
44. L. Khedher, I. A. Illán, J. M. Górriz, J. Ramírez, A. Brahim and A. Meyer-Baese, Independent component analysis-support vector machine-based computer-aided diagnosis system for Alzheimer's with visual support, *Int. J. Neural Syst.* **27** (2017) 1650050.
45. N. J. Weerkamp, G. Tissingh, P. J. E. Poels, S. U. Zuidema, M. Munneke, R. T. C. M. Koopmans and B. R. Bloem, Diagnostic accuracy of Parkinson's disease and atypical parkinsonism in nursing homes, *Parkinsonism Related Disorders* **20** (2014) 1157–1160.
46. C. La Fougère, G. Pöpperl, J. Levin, B. Wängler, G. Böning, C. Uebles, P. Cumming, P. Bartenstein, K. Bötzel and K. Tatsch, The value of the dopamine D2/3 receptor ligand 18F-desmethoxyfallypride for the differentiation of idiopathic and nonidiopathic parkinsonian syndromes, *J. Nucl. Med.* **51** (2010) 581–587.
47. G. Sáez, J. Manuel, J. R. P. de Inestrosa, J. Suckling, I. A. Illán, A. Ortiz, F. J. Martínez-Murcia, F. S. Román, D. S. Gonzalez and Y. S. Wang, Case-based statistical learning: A non-parametric implementation with a conditional-error rate SVM, *IEEE Access* **5** (9 de junio de 2017): 11468–78. <https://doi.org/10.1109/ACCESS.2017.2714579>.